Mechanism of transcription modulation by the transcription-repair coupling factor

Bishnu P Paudel[1,2], Zhi-Qiang Xu[1,2], Slobodan Jergic[1,2], Aaron J Oakley[1,2], Nischal Sharma[1,2], Simon HJ Brown[1,2], James C Bouwer[1,2], Peter J Lewis[3], Nicholas E Dixon[1,2], Antoine M van Oijen[1,2], Harshad Ghodke[1,2]*

[1]Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, Australia

[2] Illawarra Health and Medical Research Institute, Wollongong, Australia

[3]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia

Correspondence:

Harshad Ghodke, Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, New South Wales, 2522, Australia

Abstract

Elongation by RNA polymerase is extensively modulated by accessory factors. The transcription-repair coupling factor (TRCF) recognizes distressed RNAPs and either rescues transcription or initiates transcription termination. Precisely how TRCFs choose to execute either outcome remains unclear. With *Escherichia coli* as a model, we used single-molecule assays to study dynamic modulation of elongation by Mfd, the bacterial TRCF. We found that nucleotide-bound Mfd chaperones the elongation complex (EC) into a catalytically active state, presenting the EC with an opportunity to restart transcription. After long-lived residence in this catalytically poised state, ATP hydrolysis by Mfd remodels the EC through an irreversible process leading to loss of the RNA transcript. Further, electron cryo–microscopy revealed that the motor domain of Mfd binds and partially melts DNA containing a template strand overhang. The results explain pathway choice determining the fate of the elongation complex and provide a molecular mechanism for transcription modulation by TRCF.

Introduction

Gene expression involves synthesis of RNA transcripts by RNA polymerases across all domains of life. During elongation, RNA polymerase (RNAP) maintains a 9 – 10 nucleotide RNA:DNA hybrid within the transcription bubble through the templated addition of ribonucleotides (reviewed by (Belogurov and Artsimovitch, 2019)). Processive nucleotide addition cycles result in the growth of the RNA transcript that is then extruded from the polymerase through the exit channel, accompanied by translocation of the polymerase along the template. Finally, during termination, nucleotide addition ceases and RNAP is displaced from the template DNA and recycled for subsequent rounds of transcription.

Elongation is punctuated by sequence-dependent pausing, stalling at sites of damage on the DNA template or at sites of roadblocks in the form of DNA-bound proteins. Prolonged residence in paused states can lead to escape into backtracked states in which the catalytic site of RNAP loses register with the 3' end of the RNA transcript, allowing the newly synthesized transcript to invade the secondary channel of RNAP (Abdelkareem et al., 2019; Gabizon et al., 2018; Komissarova and Kashlev, 1997). Rescue from such backtracked states is catalyzed by processivity factors such as GreA and GreB (Abdelkareem et al., 2019; Borukhov et al., 1993). Elongation is modulated by the universally conserved elongation/pause factor, the bacterial NusG, homologous with Spt5 in eukaryotes (reviewed by (Wang and Artsimovitch, 2020)). In *Escherichia coli*, NusG maintains processive transcription (Burova et al., 1995) by suppressing backtracking through contacts on the upstream DNA (Herbert et al., 2010; Kang et al., 2018).

Elongation is also modulated by the transcription-repair coupling factor (TRCF) resulting in either transcription termination or rescue. The bacterial TRCF – Mfd is a multi-domain adaptor protein consisting of 7 domains connected by flexible linkers (Deaconescu et al., 2006). Domains 1a/2/1b form the UvrB–homology module that bears structural homology to the nucleotide excision repair protein UvrB and enables contacts with the repair protein UvrA. Domain 3 serves as a structural element playing a role in orienting the UvrB-homology module during the activation of the protein (Brugger et

52 al., 2020), following interactions of domain 4, the RNAP interaction domain, with distressed RNAP
53 (Kang et al., 2021). Domains 5 and 6 together form the motor domain, representing a superfamily II
54 ATPase that exhibits structural homology to the RecG translocase (Mahdi et al., 2003). Domain 7 is an
55 auto-inhibitory domain that contacts the UvrB-homology module to maintain it in a closed
56 conformation, suppressing promiscuous DNA binding of Mfd in the absence of interactions with RNAP.

57 Through molecular biology, biochemical (Chambers et al., 2003; Haines et al., 2014; Manelyte
58 et al., 2010; Park et al., 2002; Selby and Sancar, 1993a, b; Smith et al., 2012; Smith et al., 2007), and
59 structural methods (Brugger et al., 2020; Deaconescu et al., 2006; Kang et al., 2021; Shi et al., 2020),
60 as well as single-molecule (Fan et al., 2016; Graves et al., 2015; Ho Wan et al., 2012; Le et al., 2017)
61 and live-cell imaging techniques (Ghodke et al., 2020; Ho et al., 2018, 2020), a picture has been formed
62 of the bacterial TRCF Mfd acting in a complex role in modulating transcription activity, including at
63 hard-to-transcribe regions (Ragheb et al., 2021). At sites of DNA damage, Mfd orchestrates premature
64 transcription termination leading to an enhancement in repair of DNA damage in the template strand
65 – a phenomenon known as transcription-coupled DNA repair (Fan et al., 2016; Ghodke et al., 2020;
66 Haines et al., 2014; Ho et al., 2020; Mellon and Hanawalt, 1989; Selby and Sancar, 1993a). In bacterial
67 cells, the $UvrA_2B$ nucleotide excision repair machinery is recruited to the site of distressed RNAPs by
68 Mfd, resulting in the site-specific loading of the damage verification enzyme UvrB (Ho et al., 2020).
69 This ability to recruit DNA-repair machinery at sites of lesion-stalled RNAP to promote ATP-dependent
70 premature transcription termination is a property that is shared by eukaryotic functional homologs of
71 Mfd – the yeast Rad26 (Duan et al., 2020; Xu et al., 2017) and human CSB (van der Weegen et al.,
72 2020). Mutations in CSB protein result in deficient repair machinery recruitment causing Cockayne
73 syndrome in humans (Mayne and Lehmann, 1982; Troelstra et al., 1992; Venema et al., 1990).

74 In addition to initiating transcription termination, the three best-studied TRCFs also serve as
75 elongation factors. Like Rad26 (Xu et al., 2017) and CSB (Selby and Sancar, 1997), Mfd has also been
76 shown to rescue transcription at pause sites *in vitro* through a 'release-and-catch up' mechanism (Le
77 et al., 2017; Park et al., 2002). Forward translocation of the elongation complex (EC) is central to TRCF

activities including rescue from backtracked states, and transcription termination (Le et al., 2017; Park et al., 2002; Selby and Sancar, 1997; Xu et al., 2017). Considering their membership in the class of superfamily 2 helicases (Byrd and Raney, 2012), the structural homology of the TRCF motor domains and a similar mode of engagement with the elongation complex (EC) between Mfd (Brugger et al., 2020; Kang et al., 2021; Shi et al., 2020) and Rad26 (Xu et al., 2017), it is plausible that the molecular mechanisms underlying these outcomes are conserved.

However, precisely how TRCF activity results in transcription versus rescue is unknown. Two models have been proposed to explain forward translocation of RNAP by TRCFs: first, the TRCF 'pushes' RNAP forward while translocating on dsDNA as proposed for Mfd (Park et al., 2002), and second, the TRCF pulls the ssDNA template to effectively translocate the polymerase forward (Xu et al., 2017). Recent structures of *Ec*Mfd approaching an *in vitro* reconstituted EC assembled on a non-complementary bubble have demonstrated that following initial recognition of the EC, Mfd translocates forward until it bumps into the RNAP, pushing the polymerase and resulting in global remodeling of the EC (Kang et al., 2021). The structure of yeast Rad26 bound to upstream DNA in the RNA Pol II EC (Xu et al., 2017) reveals direct contacts between dsDNA-bound Rad26 and ssDNA template strand. Based on the homology to the Swi2/Snf2-family core ATPase domains, and the mechanism of Snf2 in chromatin remodeling, it has been proposed that Rad26 pulls template DNA in the Pol II EC (Xu et al., 2017). Structures of DNA-bound Mfd have established the homology of its motor domains to ATPases involved in chromatin remodeling (most notably Snf2) (Brugger et al., 2020). Since available structures depict initial global remodeling of the EC but not transcription termination leading to disassembly of the EC accompanied by loss of the transcript (Kang et al., 2021), it remains to be seen whether a similar Mfd-ssDNA complex forms during transcription termination. Further, while it is clear that displacement of the polymerase in the absence of downstream binding partners of Mfd is a slow event requiring several minutes *in vitro* (Graves et al., 2015; Ho Wan et al., 2012), the sequence of events leading to loss of the transcript, and its residence time in the EC during this remodeling remains unclear.

104 Here, we set out to understand how Mfd dynamically remodels the EC by monitoring the fate
105 of the RNA transcript under various scenarios. We developed a single-molecule Förster resonance
106 energy transfer (smFRET)-based assay that directly visualizes dynamic changes in the distance
107 between donor-labeled upstream DNA and acceptor-labelled RNA transcript forming the transcription
108 bubble. We reconstituted catalytically active ECs *in situ* and monitored the FRET signal from surface-
109 immobilized molecules under conditions of partial or complete nucleotide starvation, and probed
110 global remodeling of the EC by NusG and Mfd on the timescale of minutes. NusG efficiently rescued
111 ECs from catalytically inactive states consistent with its role as an elongation factor. On the other hand,
112 Mfd was able to rescue ECs from catalytically inactive states into a long-lived catalytically poised state,
113 prior to continued ATP-hydrolysis dependent forward translocation consistent with transcription
114 termination. This catalytically poised state acts as a 'molecular timer', presenting the EC with an
115 opportunity to restart transcription. Further biophysical and structural biology studies collectively
116 reveal that in addition to the previously described dsDNA binding mode, Mfd makes contacts with
117 ssDNA on DNA substrates mimicking the transcription bubble.

118 **Results**

119 **Rationale for a smFRET assay to study transcription elongation**

120 We monitored transcription elongation by assembling functional ECs on dye-labeled fully-
121 complementary transcription bubbles *in situ* by adopting a previously described strategy to study RNA
122 Pol II dynamics (Andrecka et al., 2008) in combination with well-established approaches to
123 reconstitute functional *E. coli* ECs (Komissarova et al., 2003). Specifically, we introduced a Cy3 donor
124 dye on the upstream DNA and a Cy5 acceptor on the RNA (Figure 1A) at sites that are predicted to
125 exhibit high smFRET signals within the dynamic range of the technique. Three features of this strategy
126 enable precise dissection of the dynamics of molecular mechanisms of Mfd-dependent remodeling of
127 the EC: (i) the use of synthetic RNA and DNA nucleotides enables site-specific introduction of dyes in
128 the phosphodiester backbone that serve as FRET pairs, and along with it, the ability to probe
129 conformational changes between various parts of this complex (ii) smFRET affords sub-second

130 temporal resolution (200 ms) enabling reliable measurements of dynamic interactions on the
131 timescale of seconds to minutes, and (iii) the ability to specify RNA sequences that cannot hybridize
132 with upstream DNA limits extensive backtracking of ECs. RNAP in the elongation phase can exist as a
133 backtracked complex (where the 3' end of the transcript is extruded into the secondary channel
134 (Abdelkareem et al., 2019)), a pre-translocated conformation prior to nucleic acid addition, and a post-
135 translocated conformation following nucleic acid incorporation. Since backtracking requires
136 hybridization of the RNA transcript with the upstream DNA, the choice of a non-complementary RNA
137 sequence is expected to minimize the heterogeneity in the observed population of ECs in this assay.

138 **Monitoring transcription elongation by RNAP**

139 Functional RNAP ECs were assembled *in situ* by flowing a solution containing RNAP that had
140 been pre-incubated with the template:RNA hybrid into a microfluidic device with surface immobilized
141 biotinylated non-template DNA. After complex assembly, the flow cell was washed with a high-ionic
142 strength solution (1M KCl) followed by rapid exchange with a buffer solution supporting transcription
143 but without ribonucleotides (NTPs). Subsequent measurements of the smFRET signal between the Cy3
144 donor and the Cy5 acceptor (see Methods) reveals complexes assembled resided in stable high FRET
145 states centered at 0.9 (Figure 1B upper trace). In addition, a small fraction (15% of 318 trajectories)
146 also exhibited mid-FRET states centered at 0.6 (Figure 1B lower trace). These states were recapitulated
147 in temporal heatmaps (Figure 1C) that are scatterplots describing how the smFRET signal evolves in
148 time for the observed population. These temporal heat maps are a convenient representation
149 presenting an overview of the progress of FRET trajectories over time. Here, the overlaid black lines
150 reflect contours parsing the data into deciles.

151 Since the majority of the complexes thus assembled exhibited time-FRET trajectories that
152 fluctuate around 0.9 FRET, and a minority exhibited transitions to the 0.6 FRET state, we pursued an
153 unbiased approach to visualize transitions between various FRET states. FRET trajectories were
154 idealized using vbFRET (http://vbfret.sourceforge.net/) and the detected states were plotted in
155 transition density plots (TDPs; Supplemental Figure 1A, 'RNAP alone') to visualize interconversion

156 among different states and the directionality of those changes (Bronson et al., 2009). In this
157 representation, features that are only present in one half of the plot on either side of the diagonal
158 manifest as asymmetric features and denote irreversible transitions between states on the timescale
159 of the observation. As demonstrated by the TDP, the interconversions between the states were
160 reversible, and were restricted to the high FRET regime. Together, the observed symmetry of these
161 peaks and their proximity to the diagonal, and to each other indicate that the complex thus
162 reconstituted transitions between states reversibly. Within the paradigm that FRET changes reflect
163 conformational changes, a simple interpretation of these data is that the complex undergoes relatively
164 minor, but reversible, conformational changes. For convenience, here we assign smFRET signals
165 between 0.7 – 1 as being 'high', 0.4 – 0.7 as 'mid' and 0 – 0.4 as being 'low'.

166 To assign states, and to test that the complexes are catalytically functional, we challenged
167 complexes with deoxyNTPs (dNTPs), 'wrong' (or non-complementary) NTPs and the complete set of
168 NTPs. In this assay, transcript elongation is expected to increase the distance between the FRET pair
169 dictated by the geometry of the EC, leading to a drop in FRET. Since dNTPs are selected against
170 incorporation into the transcript (Svetlov et al., 2004), we expected that smFRET signal from the
171 reconstituted ECs would not drop. Consistent with this expectation, complexes exposed to dNTPs
172 revealed a high-FRET population (Figure 1D, 1E). In contrast, non-complementary NTPs converted the
173 population of complexes in the high-FRET states to the mid-FRET states (Figure 1F, G). In both cases,
174 the interconverting populations were found to be symmetric across the diagonal indicating that in the
175 presence of either dNTPs or the wrong NTPs, the ternary complex visits the observed states reversibly
176 (Supplemental Figure 1A).

177 'Walking' the complex by supplying a subset of the 'correct' complementary NTPs revealed
178 gradual conversion from the high-FRET to a mid-FRET population (Supplemental Figure 1B–D). When
179 presented with the complete set of NTPs, the ternary complex exhibited unidirectional conversion of
180 the FRET signal from high- to low-FRET values (Figure 1H, 1I and Supplemental Figure 1A). Notably,
181 the transition density plots, and indeed the individual traces, revealed a conversion of the high-FRET

182 states to the low-FRET states *via* the mid-FRET states.

183 Collectively, the data indicate that the complexes assembled *in situ* represent functional
184 ternary ECs that exist in a catalytically 'inactive' state and a catalytically 'poised' state. The
185 observations that complexes first transition to the mid-FRET state (0.6) prior upon addition of NTPs
186 suggests that this represents a catalytically poised state. We tentatively assigned the high-FRET states
187 centered at 0.9 to a catalytically inactive state. This catalytically inactive state may represent a pre-
188 translocated or backtracked conformation of the EC. Considering that the sequence of the 5' end of
189 the transcript is non-complementary to the upstream DNA, we expect that extensive backtracking is
190 disfavored on this substrate. We propose that the likeliest explanation is that the high- and mid- FRET
191 states detected here represent pre- and catalytically poised post-translocated states accessed by the
192 EC during the nucleotide addition cycle.

193 **NusG promotes residence of ECs in the catalytically poised state**

194 Next, we tested this assignment of states by assessing the response of the complex to the elongation
195 factor – *Ec*NusG (Figure 2A), which is predicted to retain ECs in the catalytically poised state. Addition
196 of NusG indeed resulted in an increased occupancy of mid-FRET states accompanied by a reduction of
197 the high-FRET (Figures 1B, 1C *cf*. 2B, 2C). Trajectories revealed a conversion from the high-FRET state
198 into the mid-FRET state and fluctuation around the mid-FRET state (Figure 2B). This behavior confirms
199 our assignment of the mid-FRET state as the catalytically active state. The reverse transition from mid-
200 FRET to high-FRET states that would reflect unbinding of NusG appears to be disfavored in this
201 experimental condition as the detected asymmetric features in the transition density plot are
202 unidirectional (Supplemental Figure 2A). These trajectories are readily explained if binding of NusG
203 restricts conformational rearrangements of the EC.

204 Incubation with incorrect NTPs in the presence of NusG revealed enrichment of molecules
205 exhibiting the 0.6 FRET state (Figure 2E, 2F), similar to that observed in the case of incorrect NTPs
206 alone (Figure 1F, 1G). Reversible short-lived transitions to low-FRET states were also evident in this
207 data set (Figure 2D, 2E and Supplemental Figure 2A). These transitions may reflect changes in the path

208 of the upstream DNA that have been detected previously (Kang et al., 2018). Importantly, the
209 complete set of NTPs revealed efficient, irreversible conversion to the low-FRET states (0 – 0.2) from
210 the mid-FRET state (Figure 2F, 2G and Supplemental Figure 2A) in the presence of NusG suggesting
211 processive extension of the transcript through catalysis and associated forward translocation by RNAP.

212 Together, these observations are consistent with the role of *Ec*NusG as a suppressor of
213 backtracking of RNAP (Burova et al., 1995; Herbert et al., 2010; Kang et al., 2018). Overall, these results
214 indicate that *Ec*NusG promotes residence of the complex in a catalytically poised conformation. The
215 loss of high-FRET states upon binding of NusG supports the interpretation that the high-FRET states
216 represent inactive states.

217 **Remodeling of the elongation complex by Mfd**

218 With the assignment of the high-FRET state as a catalytically inactive state, and the mid-FRET state as
219 a catalytically poised state, we next assessed the interaction of Mfd with ECs. To characterize the
220 binding of Mfd to RNAP, we first measured the affinity of Mfd for pre-assembled ECs, or the dsDNA
221 alone using surface plasmon resonance (SPR). Since Mfd interacts poorly with DNA in bulk assays in
222 the absence or presence of ATP, measurements of binding of purified Mfd to either DNA or the EC
223 were carried out in the presence of slowly hydrolyzable ATPγS. SPR measurements revealed a
224 dissociation constant (K_D) of 170 ± 20 nM (Supplemental Figure 2B) for 49-mer dsDNA alone,
225 consistent with recent measurements by other groups (Brugger et al., 2020; Le et al., 2017). Binding
226 to reconstituted RNAP ECs on a 49-mer exhibited similar affinity with a K_D of 127 ± 6 nM (Supplemental
227 Figure 2C). These data are consistent with the model that dsDNA serves as a major binding partner for
228 Mfd-ATPγS in the context of the EC as previously demonstrated (Kang et al., 2021; Le et al., 2017; Park
229 et al., 2002; Shi et al., 2020).

230 We then investigated how interactions of Mfd with RNAP influence the conformational
231 dynamics of the transcription bubble using our smFRET assay (Figure 3A). Consistent with the lack of
232 stable binding of Mfd with DNA in the absence of ATP, Mfd engaged the EC poorly in the smFRET
233 measurements with 58% of trajectories (n_{obs} = 75) presenting no observable interactions with Mfd

234 (Figure 3B lower panel). The remaining trajectories exhibited reversible transitions from the high- to
235 mid-FRET states even in the absence of any nucleotide co-factor (Figure 3B upper panel, 3C and
236 Supplemental Figure 3A 'RNAP, Mfd') reflecting attempts at stable engagement of Mfd with the EC.
237 Notably, differently from the EC alone (Supplemental Figure 1A), dynamic trajectories exhibited
238 transitions mainly in the mid-FRET region (Figure 3B upper panel, Supplemental Figure 3A 'RNAP,
239 Mfd').

240 Mfd-ATPγS resulted in dynamic ECs visiting additional mid FRET states (Figure 3D, 3E and
241 Supplemental Figure 3A 'RNAP, Mfd, ATPγS'; 59% of 69 trajectories presented dynamic behavior).
242 ATPγS alone did not account for these observations (Supplemental Figure 3B). Curiously, Mfd-ATPγS
243 exhibits quantized transitions in the transition density plot compared to Mfd alone, indicating that
244 binding of Mfd-ATPγS displaces RNAP forward (Supplemental Figure 3A 'RNAP, Mfd' *vs.* 'RNAP, Mfd,
245 ATPγS'). Together the data reveal that Mfd can isomerize RNAP merely upon binding to the nucleotide
246 starved EC. The efficiency of this isomerization is appreciably greater in the presence of ATPγS.

247 Treatment of *in situ* assembled ECs with Mfd and ATP revealed four distinct types of
248 trajectories in the data set: (i) static trajectories that exhibited high FRET values throughout the
249 observation reflecting molecules that are presumably unaffected by Mfd-ATP (54/176 observations;
250 31%), (ii) dynamic trajectories exhibiting high FRET values at the start of the observation resulting in a
251 direct transition to low-FRET states (24/176 trajectories, 14%; Figure 3F lower panel), (iii) dynamic
252 trajectories starting at a high-FRET value but transiting to a low-FRET state through a single mid-FRET
253 intermediate (56/176 trajectories; 32%, Figure 3F upper panel) and (iv) dynamic trajectories that are
254 at the mid-FRET state at the start of the observation (42/176 corresponding to 24%; compared to 14%
255 of observations in the absence of Mfd). Since the dataset presented here represents a pooled data set
256 from several videos collected after Mfd-ATP was introduced to the flow chamber, this last category of
257 trajectories likely reflects ECs that are already being remodeled by Mfd at the start of the observation.

258 We characterized the trajectories describing efficient and stepwise conversion of the
259 population of ECs from the high- to the mid-FRET state, with further conversion to low-FRET states for

260 individual ECs (Figure 3F, upper panel). The intermediate step was randomly distributed in the range
261 of FRET from 0.35 to 0.5 (Supplemental Figure 3C). Strikingly, the lifetime of this mid-FRET state did
262 not exhibit a dependence on the concentration of ATP in the range from 10 μM to 10 mM
263 (Supplemental Figure 3D). To further explore the nature of these transitions, we visualized them in a
264 transition density plot (Supplemental Figure 3A, 'RNAP, Mfd, ATP'; entire data set is plotted here).
265 Notably, irreversible transitions manifesting as asymmetric features in these plots appeared mainly as
266 transitions from the mid-FRET to low-FRET states (Supplemental Figure 3A, 'RNAP, Mfd, ATP', red
267 ellipse). As a control, titration of ATP in the absence of Mfd resulted in RNAP complexes exhibiting
268 primarily a high-FRET population, with a gradual increase in the mid-FRET state at millimolar
269 concentrations (Supplemental Figure 3E). Compared to the wild-type, the Mfd(R953A) mutant
270 (Chambers et al., 2003), known to be proficient in ATP hydrolysis while strikingly deficient in RNAP
271 displacement and transcription termination exhibited inefficient conversion to the low-FRET states
272 (Figure 3H, 3I and Supplemental Figure 3A 'RNAP, Mfd(R953A), ATP'). Notably, unlike wild-type Mfd,
273 irreversible transitions were observed mainly in the high- to mid- FRET regime (Supplemental Figure
274 3A 'RNAP, Mfd(R953A), ATP', red ellipse). Intriguingly, Mfd(R953A)-ATP exhibited a pattern of
275 quantized transitions that is similar to that of Mfd-ATPγS in the transition density plots (Supplemental
276 Figure 3A, compare 'RNAP, Mfd, ATPγS' and 'RNAP, Mfd(R953A), ATP'), suggesting that the mutation
277 in this residue serves to reveal distinct aspects of Mfd's binding to the EC.

278 Together, the data indicate that both wild-type and translocase deficient Mfd rescue RNAP
279 from the catalytically inactive (potentially pre-translocated or weakly backtracked) states accessible
280 in our substrate to catalytically-poised states, albeit with varying efficiencies. However, only the wild-
281 type Mfd can further process the EC in the presence of ATP. The low-FRET states observed in our
282 system represent circumstances in which the distance between the donor-acceptor pair is outside the
283 dynamic range of efficient FRET as might be expected when the RNAP is forward translocated well
284 past the stall site and/or the Cy5-labelled RNA transcript is lost. The observation that the EC fails to
285 return to the mid- or high-FRET states rules out a scenario where this low-FRET state simply reflects

286 an over-wound transcription bubble, or a bent upstream DNA conformation as is seen in the case of
287 NusG (Kang et al., 2018). Hence, this loss of FRET is most consistent with an irreversible loss of RNA.
288 In this model, Mfd(R953A) fails to displace the transcript from the EC, consistent with prior findings
289 (Chambers et al., 2003).

290 **Mfd contacts template ssDNA**

291 Structures of Mfd bound to dsDNA and ECs containing non-complementary transcription bubbles have
292 revealed a series of translocation intermediates that are formed prior to and leading up to stable
293 engagement of Mfd and initial disruption of the EC (Kang et al., 2021; Shi et al., 2020). Currently, little
294 is known about the intermediate states that follow this initial contact. We set out to identify whether
295 Mfd makes contacts with ssDNA in substrates mimicking the upstream edge of the transcription
296 bubble by employing a combination of approaches: First, we employed smFRET experiments to detect
297 whether and how Mfd contacts template ssDNA. Second, we visualized Mfd bound to a primed DNA
298 template using cryo-EM.

299 To detect ATP-dependent contacts of Mfd with ssDNA we employed a previously described
300 smFRET assay to study translocase activity (Park et al., 2010). We used a DNA substrate comprised of
301 an 18-mer dsDNA duplex (mimicking upstream DNA) and a 5' dT_{40} overhang mimicking the template
302 DNA (Figure 4A) with a FRET pair positioned on either end of the ssDNA overhang (Table 1:
303 bio_AS18_Cy5: Cy3_dT40_S18, see Materials and Methods). In smFRET experiments, this substrate
304 exhibited an average FRET of 0.33 ± 0.15 (Figure 4B) in the absence of Mfd reflecting rapid sampling
305 of various conformations accessed by the entropic coil formed by ssDNA (Figure 4C and Supplemental
306 Figure 4A 'template overhang DNA'). Mfd binds this substrate with a K_D of 390 ± 10 nM (see
307 Supplemental Information and Supplemental Figure 4B–D).

308 In the presence of ATPγS, the smFRET signal from the Mfd-bound substrate revealed
309 additional distinct populations at higher FRET values of 0.6 and 0.8 (Figure 4D, 4E and Supplemental
310 Figure 4A 'Mfd, ATPγS'). Transitions among these states were further enhanced in the presence of ATP
311 as the substrate exhibited a continuous spectrum of mid- and high-FRET states (Figure 4F, 4G and

312 Supplemental Figure 4A 'Mfd, ATP') in addition to the distinct states detected in the presence of
313 ATPγS. In this case, the higher FRET states were only transiently accessed indicating that ATP hydrolysis
314 plays a role in exiting from these high–FRET states to low–FRET states. This observation is consistent
315 with a picture in which Mfd-ATP contacts the ssDNA to bring the dyes close to each other resulting in
316 a high FRET signal, and subsequent ATP hydrolysis induces a lower–affinity conformation in which
317 Mfd-ADP binds the ssDNA very weakly resulting in a return to the baseline FRET value of 0.33.

318 In the presence of the ATPase-deficient but DNA-binding competent E730Q mutant of Mfd
319 (Deaconescu et al., 2012; Ho et al., 2018), this template exhibited high-FRET states extremely rarely
320 compared to the wild-type protein (Supplemental Figure 4F, 4G). Therefore, the observation that
321 Mfd(E730Q)-ATP is unable to stably engage with ssDNA, unlike Mfd-ATPγS suggests that the
322 Mfd(E730Q)-ATP and Mfd-ATPγS pre-hydrolytic states are sufficiently different that only the wild-type
323 Mfd retains the ability to engage ssDNA efficiently.

324 The RNAP-displacement deficient R953A mutant (Chambers et al., 2003) did not exhibit high
325 FRET states in the presence of ATP, indicating that the ability to hydrolyze ATP alone is insufficient to
326 engage ssDNA on primed DNA templates (Figure 4H, 4I). Therefore, we conclude that both the ability
327 to hydrolyze bound ATP and the presence of an intact R953 are essential for Mfd to engage the ssDNA
328 template DNA overhang. We propose that R953 serves to 'hook' ssDNA, and ATP binding serves to
329 'pull' ssDNA closer to the junction resulting in the high FRET signal detected in this assay with wild-
330 type Mfd. We note that the recent structures of Mfd translocating on dsDNA indicate that R953 along
331 with key residues in the translocation motif (TRG: translocation in RecG (Mahdi et al., 2003)) of Mfd,
332 specifically R929 and Q963, contacts the non-template strand in dsDNA. From these structures it is
333 not immediately obvious how the template strand overhang is contacted by Mfd in the smFRET assays.
334 The results presented here uncover a new role for R953 in making critical contacts with the template
335 ssDNA distinct from its previously described role in translocation on dsDNA.

336 Intriguingly, each of the three Mfd variants also exhibited a distinct sub-population of
337 molecules centered at ~0.1 FRET. Here, whereas ATPγS promoted the formation of this state almost

338 immediately, Mfd (or mutants) in the presence of ATP entered this state in a time window ~5-20
339 seconds after introduction into the flow cell (Figure 4E, 4G, 4I and Supplemental Figure 4H). The
340 presence of the low-FRET peak and the excursions to high-FRET states suggested that multiple
341 independent events may underlie the FRET signal observed here. Whereas all conditions involving Mfd
342 (or mutant) exhibited the low-FRET peak, the excursions to the high FRET states were only observed
343 for wild type Mfd in the presence of ATP or non-hydrolyzable analog.

344 **Mfd melts dsDNA upon contacting template ssDNA**

345 To further understand how Mfd binds to 5' overhang–containing DNA, we visualized Mfd
346 bound to the dye labelled smFRET substrate by cryo-EM (Table 1: bio_AS18_Cy5: Cy3_dT40_S18, see
347 Materials and Methods). Full-length Mfd bound to the primed-template substrate in the presence of
348 transition-state analog ADP-AlF_x was subjected to single particle cryo-EM analysis (Figure 5A) and a
349 ~5.2 Å resolution map was obtained (real resolution ~8 Å, Supplemental Figure 5A–C). This
350 reconstruction yielded electron density that could at best accommodate the motor domains of Mfd
351 (residues 575–956 representing domains 5 and 6); other domains are presumably flexible and could
352 not be seen in the electron density maps.

353 The cryo-EM map reconstructed here suggests that the dsDNA on the upstream side of the
354 motor domains and the placement of the motor domains with respect to dsDNA is consistent with
355 previously reported complexes (Brugger et al., 2020; Kang et al., 2021) (Figure 5B). However, density
356 for residues 917 onward, containing the TRG motif, appeared to be absent from the side of domains
357 5 and 6, and the density of the DNA as it passes from the upstream side to the downstream side
358 becomes inconsistent with dsDNA (Figure 5B). Using molecular dynamics flexible fitting (Singharoy et
359 al., 2016) to equilibrate the structure in density suggests that the TRG motif has dropped into the path
360 of the dsDNA (yellow ribbon, Figure 5C), resulting in the partial melting of the duplex. The motif forms
361 an arch that interacts with the template DNA (Figure 5C, 5D). There are precedents for conformational
362 plasticity in the TRG motif: the pair of helices forming the bulk of the TRG motif are almost orthogonal
363 in the apo-structure of Mfd (PDB ID 2EYQ) with the sulfate group from 4-(2-hydroxyethyl)-1-piperazine

364 ethanesulfonic acid bound between the helices, and are parallel in the complex with ADP-AlF_x and
365 dsDNA (PDB 6XEO). These movements are attributable to both to shifts in domains 5 and 6 relative to
366 each other and to movements of the helices relative to each other. The conformation of domains 5
367 and 6 suggested here may represent a mode of engagement whereby Mfd grabs and stabilizes the
368 template DNA in the upstream edge of the transcription bubble.

369 The model suggests that Mfd partially melts about five nucleotides of the dsDNA near the
370 junction forming a 'TRCF bubble' (Figure 5E). Residues R953 and R929 line the path of the template
371 ssDNA three and six nucleotides away from the junction respectively (Figure 5C, 5F). ADP-AlF_x–bound
372 Mfd scrunches the template ssDNA (Figure 5D–F), explaining the basis of the high–FRET signal
373 observed in the smFRET experiments (Figure 4). The proximity of R953 to the ssDNA in the model also
374 explains how R953 can hook the template ssDNA in an ATP-dependent manner (Figure 5C, 5F) and in
375 case of the distressed EC, transform the energy of ATP binding and hydrolysis into forward
376 translocation of RNAP.

377 Several highly conserved basic residues (H842, R846, R848, H890, H897, R900; Figure 5E,
378 residues indicated as blue spheres) line the path of the template ssDNA along with absolutely
379 conserved polar (N817, T868) (Figure 5G, residues indicated as yellow spheres) and aromatic residues
380 (Y816, F891) (Figure 5G, residues indicated as green spheres) capable of base stacking interactions or
381 H-bonding interactions with ssDNA (Figure 5G, residues indicated as spheres). A cluster of highly
382 conserved basic residues in domain 5 (R685, K690, H711, K712, H732, R733, R737, K739, R761) (Figure
383 5G, blue spheres) along with polar (T710) (Figure 5G, yellow spheres) and aromatic (F734) (Figure 5G,
384 green spheres) residues comprise the surface in contact with the non-template DNA strand (Figure
385 5E). The binding mode proposed here is partially validated by prior observations implicating several
386 key conserved residues in domain 6 (Y816, N817, H842, T868 and R887) as well as domain 5 (R685,
387 K690, T710, H711 and K712) implicated in DNA binding (Brugger et al., 2020; Le et al., 2017). Consistent
388 with this scenario, Mfd-ATPγS binds a 49-mer dsDNA substrate containing a non-complementary
389 bubble of 9 nucleotides with an affinity of 395 ± 7 nM as assessed by SPR (Supplemental Figure 5D).

390 Together, the findings describe how Mfd contacts DNA substrates mimicking the upstream
391 edge of the transcription bubble: the cryo-EM data suggest that Mfd binds and partially melts the
392 dsDNA near the junction, contacting the unpaired template ssDNA via the conserved R929 and R953
393 residues in the TRG motif, and the smFRET experiments reveal that Mfd dynamically cycles between
394 high- and low-affinity states in an ATP-hydrolysis dependent manner on the template ssDNA overhang,
395 pulling and scrunching the template strand in this process.

396 **Discussion**

397 In this work, we established that a 'molecular timer' intermediate formed by Mfd and the transcription
398 elongation complex determines the fate of the elongation complex. Our attempts to understand this
399 intermediate resulted in the discovery of a novel DNA-binding mode in which Mfd binds ssDNA in a
400 transcription bubble mimic. Together, the results provide unprecedented insight into the transcription
401 termination reaction and enable us to propose a comprehensive model describing how events unfold
402 after initial engagement of the elongation complex with the transcription-repair coupling factor.

403 Interactions of Mfd with RNAP are complex. ATP-dependent remodeling of the EC by Mfd
404 proceeds via multiple stages: In the first stage, Mfd recognizes a distressed polymerase and is
405 recruited to the EC (Ho et al., 2018; Ho Wan et al., 2012; Le et al., 2017; Park et al., 2002; Proshkin and
406 Mironov, 2016) through a conserved RNAP-interacting domain (Park et al., 2002) (Figure 6A). Mfd can
407 recognize extensively backtracked (Le et al., 2017), reversibly paused as well as stalled complexes
408 (Proshkin and Mironov, 2016). In the second stage, EC-tethered Mfd undergoes a series of concerted
409 conformational changes (Smith et al., 2012) that enable it to translocate on upstream dsDNA and
410 'load' onto the EC (Kang et al., 2021; Shi et al., 2020) (Figure 6B). In the third stage, Mfd either
411 reactivates paused transcription (Le et al., 2017; Park et al., 2002), or terminates transcription (Ho
412 Wan et al., 2012; Selby and Sancar, 1993a). Alongside these activities, Mfd recruits its binding partners
413 $UvrA_2B$ to the site during transcription-coupled DNA repair (Fan et al., 2016; Ghodke et al., 2020; Ho
414 et al., 2018, 2020; Selby and Sancar, 1993a).

415 Our studies shed light on hitherto mysterious aspects of Mfd-RNAP interactions central to this

416 reaction and present a new model for resolution of a critical intermediate. The smFRET-based
417 elongation modulation assay revealed that Mfd can engage the EC in an ATP independent manner and
418 manifest transient, reversible conformational changes in the EC. Stable association requires ATP or at
419 minimum non-hydrolysable ATP, likely required to impart a dsDNA binding conformation in the motor
420 domain. Indeed, a recent study reported the structural basis of ATP-dependent translocation of Mfd
421 following initial recognition of distressed elongation complexes leading to docking of Mfd with the
422 upstream edge of the transcription bubble (Kang et al., 2021). Our findings agree well with these
423 findings (Figure 6 A – C).

424 Following a period of dynamic remodeling in which the EC transits from high- to mid-FRET
425 states, Mfd-ATP, Mfd-ATPγS and Mfd(R953A)-ATP each chaperone the EC into a single long-lived
426 gating intermediate (Figures 3 and 6D) hovering in the mid-FRET range controlling the transition to
427 low-FRET states. Efficient entry into this state is granted to Mfd in the presence of ATP or ATPγS,
428 whereas Mfd(R953A)-ATP enters this state only inefficiently. The observation that Mfd-ATPγS is
429 sufficient to promote entry into this state indicates that binding of ATP alone is a sufficient
430 requirement for the formation of the gating intermediate.

431 The suggestion that Mfd-ADP-AlF_4 might bind and melt primed DNA templates to unwind
432 dsDNA (Figure 5), provides insight into this transition. Only 13 nucleotides of the 18-mer DNA template
433 are consistent with dsDNA, with the remainder being partially melted. The lack of observations of
434 dsDNA melting in a recently reported structure (Brugger et al., 2020) suggests that Mfd requires access
435 to unpaired ssDNA *via* critical residues in the TRG motif to melt the dsDNA.

436 We propose that in the context of the upstream dsDNA, ATP- or non-hydrolyzable analog
437 bound - Mfd partially melts about five nucleotides of dsDNA to produce a bubble, henceforth referred
438 to as the 'TRCF bubble' (Figure 5). Considering the three nucleotide proximity of the binding site of
439 Mfd to the upstream edge of the transcription bubble (Kang et al., 2021), the upstream DNA may
440 further melt allowing the TRCF bubble to fuse with the transcription bubble to produce the isomeric
441 TRCF-transcription ('TRCF-T') bubble (Figure 6D). Since Mfd has not been reported to melt dsDNA

442 lacking an overhang, a more likely possibility is that Mfd merely extends the transcription bubble by
443 further melting the upstream edge, resulting in essentially a TRCF-T bubble.

444 Binding of Mfd to bubble substrates has hitherto escaped extensive characterization, although
445 we documented that Mfd binds a nine nucleotide bubble with a dissociation constant that is only two-
446 fold weaker compared to dsDNA, consistent with this cryo EM–derived model. Several independent
447 pieces of evidence in the literature support our finding: (i) A previous study investigating specificity of
448 binding of Mfd to a 12 nucleotide bubble reported a comparable but not enhanced binding affinity to
449 a 12–nucleotide bubble substrate (Selby and Sancar, 1995). This observation is consistent with our
450 finding that binding to bubble substrates is only marginally weaker than to dsDNA. (ii) Although the
451 authors do not interpret the observations as a TRCF-T bubble, Mfd-dependent exposed ssDNA in the
452 20 nucleotides coinciding with Mfd's footprint in the upstream DNA is evident in potassium
453 permanganate sensitivity assays aimed at monitoring Mfd-dependent rewinding of the transcription
454 bubble (Ho Wan et al., 2012), and (iii) ATPγS–mediated a 12–base pair movement in the footprint of
455 bound Mfd toward the EC in unzipping mapper experiments (Le et al., 2017). This effect was attributed
456 by the authors to a nucleotide dependent conformational change in Mfd, consistent with the
457 formation of the TRCF-T bubble described here.

458 Our cryo-EM derived model suggests that R953 and R929 line the path of the template
459 overhang (Figure 5). The observation that Mfd(R953A)-ATP fails to stably engage the ssDNA overhang
460 in our smFRET study supports a model in which this residue (potentially along with other residues in
461 the translocation motif) serves to stabilize the fork at the downstream edge of the TRCF bubble, or
462 template ssDNA in the TRCF-T bubble. We propose that binding of R953 to the ssDNA template occurs
463 simultaneously with the melting of the upstream dsDNA. Evidently along with a role in DNA
464 translocation, R953 has a newly discovered role in hooking and stabilizing ssDNA. Our structure also
465 predicts a critical role for R929 in this activity.

466 A clear prediction from our model is that ATP– (or non-hydrolyzable analog)-bound Mfd
467 contacts upstream DNA in the EC and this results in the formation of an extended TRCF-T bubble and

468 a contact through R953. Although available structures of Mfd-[ATP or non-hydrolyzable analog]-EC do
469 not report melting of the upstream DNA or a TRCF-T bubble, and a structure of Mfd-EC bound to the
470 TRCF-T bubble is not as yet forthcoming, precisely such an intermediate has been observed in the
471 Rad26-Pol II structure where an extended TRCF-T bubble of 17 nucleotides is observed (Xu et al.,
472 2017). The conserved motor domains of Rad26 contact three nucleotides each from partially unwound
473 upstream DNA in this structure. Analogous to R953 in Mfd, R613 in Rad26 makes a contact with ssDNA
474 in the TRCF-T bubble. These similarities between the prototypical members of this class of
475 transcription-coupled DNA repair factors hint to a more profound conservation of mechanism of
476 action.

477 We propose that the formation of the TRCF-T bubble results in forward translocation of the
478 EC (or at minimum restricts backtracking of the EC) leading to the formation of the catalytically poised
479 state of the RNAP. Here, in the presence of the correct NTP and an intact R953, resumption of
480 transcription would lead to faster forward translocation of RNAP leaving behind Mfd-ATP bound to
481 the TRCF bubble (Figure 6E). In effect this finding explains how Mfd can 'release' distressed
482 polymerases in the so-called 'release and catch-up' model for transcription rescue (Le et al., 2017). A
483 prediction of our model is that Mfd bound to a non-hydrolyzable ATP analog should be sufficient to
484 induce transcription rescue.

485 To understand what happens once this gating intermediate (Figure 6D) forms, we must first
486 consider that exit from this intermediate is signified by a loss of FRET. To what might this irreversible
487 loss of FRET correspond? This catalytically poised intermediate formed prior to irreversible remodeling
488 of the EC is directly comparable to the Mfd-concentration independent intermediate detected in
489 nanomanipulation assays (Ho Wan et al., 2012). However, two key differences are immediately
490 apparent: first, whereas the lifetimes of the intermediate detected in our FRET studies follows a single
491 exponential distribution with a lifetime of 6 s (reflecting a process with a single dominant rate-limiting
492 step), the lifetimes of the intermediate observed in the previous study exhibit a Gaussian distribution
493 (reflecting a stepwise process with several comparable reaction rates). Second, the lifetime of the

494 intermediate in the nanomanipulation assays (335 ± 3s) is much longer than the lifetime of the
495 intermediate detected here (6 s). To reconcile these differences, we note that the nanomanipulation
496 assays reflect the extension of the supercoiled DNA as a proxy for rewinding of the transcription
497 bubble, while our smFRET measurements reflect the distance between the dye pair positioned on the
498 upstream DNA and the RNA transcript. Thus loss of FRET observed once the complex exits from the
499 gating intermediate can reflect either loss of the transcript, or forward translocation of the polymerase
500 to a point where the acceptor is no longer within the dynamic range of the technique. Since the
501 lifetime of the intermediate formed during complete rewinding of the transcription bubble (335 s) (Ho
502 Wan et al., 2012) is much longer than measured here, the irreversible process detected here likely
503 describes loss of the RNA transcript (Figure 6D, F). This interpretation is supported by previous
504 observations that loss of RNA from the EC occurs rapidly after Mfd binds the EC (Graves et al., 2015).

505 An important consequence of entry into this gating intermediate is that a finite time window
506 is available beyond which, transcription restart is no longer possible. The prolonged residence in the
507 catalytically poised state prior to irreversible remodeling of the EC by Mfd may serve to provide an
508 opportunity to restart transcription at sites of pausing. This observation explains how Mfd can
509 orchestrate transcription restart (Le et al., 2017) or termination (Ho Wan et al., 2012) via a kinetic
510 proofreading mechanism (Hopfield, 1974) whereby paused ECs can be substrates for transcription
511 restart, and irreversibly or DNA damage-stalled ECs are substrates for transcription termination. In
512 effect, this gating intermediate serves as a molecular timer determining the fate of the EC.

513 It can be deduced that resolution of this gating intermediate during transcription termination
514 proceeds via at least partial, and potentially complete, shrinking of the TRCF-T bubble leaving behind
515 Mfd bound to the TRCF bubble. How might these events play out? Our results demonstrate that the
516 lifetime of this state is only 6 s at 20 °C. While entry into this state is dependent on the presence of
517 ATP, the observation that this lifetime is independent of ATP concentration in the range from 10 μM
518 to 10 mM, suggests that hydrolysis of stoichiometric amounts of ATP, may be sufficient to promote
519 exit from this state. Further, an intact R953 is required for exit from this state. Since dissociation of

520 the 9-nt RNA:DNA hybrid (possessing a melting temperature of 15 °C) is expected to be spontaneous
521 at this temperature in the absence of stabilizing interactions with the EC, the loss of the transcript
522 signals the resolution of the molecular timer, accompanied by the formation of an intermediate where
523 RNAP is extensively remodeled, and the TRCF-T bubble is partially or completely rewound. The cryo-
524 EM derived model presented here sheds light on how Mfd may scrunch template DNA in this process.

525 A model for transcription termination that accounts for these observations is as follows: Entry
526 into the gating intermediate reflects a configuration in which Mfd-ATP binds the EC on an extended
527 TRCF-T bubble intermediate in which the 3'-OH of the RNA transcript is aligned with the active site of
528 the polymerase, and R953 hooks the template ssDNA. Based on the structures of translocation
529 intermediates of Mfd (Kang et al., 2021), we propose that ATP hydrolysis leads to unhooking of ssDNA
530 (Figure 4), completing the pull stroke. This in turn leads to re-zipping of the upstream edge of the TRCF
531 bubble and advancement of the TRCF bubble toward the EC. This ATP hydrolysis event triggers
532 catastrophic remodeling of the EC (Figure 3 and Supplemental Figure 3), exposing the template-RNA
533 (potentially through a rearrangement of the β' lid of RNAP) and enabling its rapid dissociation from
534 the former EC. While this model awaits further validation, a prediction is that Mfd must possess an
535 annealase activity as has been documented for the human homolog CSB (Muftuoglu et al., 2006).

536 What happens to Mfd after this ATP hydrolysis event? Single-molecule nanomanipulation
537 assays monitoring the extension of the DNA during the resolution of the Mfd-RNAP intermediate
538 suggest that the transcription bubble shrinks by two-thirds of its size (Graves et al., 2015; Ho Wan et
539 al., 2012). The observation that the transcription bubble was rapidly rewound upon interaction with
540 Mfd led the authors to suggest that this residual DNA extension reflects "bending/wrapping
541 interactions". Alternately, this residual DNA extension might reflect Mfd binding to a TRCF bubble and
542 translocating forward via continued ATP hydrolysis.

543 Mfd recruits the repair machinery to the site of distressed polymerases during transcription-
544 coupled nucleotide excision repair. During global genomic nucleotide excision repair, the damage
545 recognition enzyme $UvrA_2$ partially melts putative damage containing dsDNA and loads two copies of

546 the damage verification enzyme UvrB at the site (Kisker et al., 2013). Two UvrB helicases then examine
547 the unpaired ssDNA with single-nucleotide resolution to precisely identify the damaged strand (Jaciuk
548 et al., 2020; Pakotiprapha et al., 2012). In contrast, during transcription coupled DNA repair, Mfd
549 recruits $UvrA_2B$ to the site of the distressed polymerase (Ho et al., 2020). Compared to global genomic
550 nucleotide excision repair, only a single copy of UvrB is recruited to a site of damage by Mfd through
551 interactions with $UvrA_2$. Thus, sequestration of the damage–containing template by Mfd, as suggested
552 here, raises the intriguing possibility that Mfd possesses the intrinsic ability to faithfully convey the
553 identity of the damaged strand to complete the handoff to a single UvrB via the adaptor $UvrA_2$. Further
554 elucidation of transcription-coupled DNA repair intermediates will shed light on this model.

555 Exposed ssDNA in the TRCF bubble may be more susceptible to DNA damage, and consistent
556 with this expectation, overexpression of Mfd was found to sensitize cells to 4-nitroquinoline,
557 ultraviolet light and nitrofurazone (Kamarthapu et al., 2016). Based on the findings of our study, we
558 can propose a model in which high cellular concentrations of Mfd lead to indiscriminate binding of
559 Mfd to ECs, leading to deposition of TRCF bubble-bound translocating Mfd molecules in the wake of
560 transcription elongation. Treatment with DNA damaging agents may serve to damage the exposed
561 nucleobases in single-stranded DNA. Thus, the TRCF bubble may explain a hitherto mysterious
562 property of Mfd: the ability to promote mutagenesis (Ragheb et al., 2019).

563 **Acknowledgments**

564 We thank Rachel Ann Mooney and Robert Landick (UW Madison) for generously sharing expression
565 vectors for RNA polymerase and purified NusG used in these studies. HG is supported by NIH Grant
566 1RM1GM130450 and acknowledges UOW Small Project Grant (2020/SPGA-S/04). ND, AJO, PJL, HG
567 acknowledge support by the Australian Research Council (DP210100365). AMVO acknowledges
568 support by the NIH (1RM1GM130450) and Australian Research Council (DP210100167).

569 **Author Contributions**

570 Data curation: BP, ZQ, SJ, SB, HG; Data analysis: BP, ZQ, SJ, HG, AJO, PJL; Writing—Original draft: HG;
571 Writing—review and editing: all authors; Resources: ND and AvO; Conceptualization: BP and HG;

Supervision: HG.

Declaration of Interests

No competing interests.

576 **Materials and Methods**

577 **Proteins used in this study**

578 Purification of RNA polymerase

579 RNA polymerase was purified essentially as described previously (Svetlov and Artsimovitch, 2015).
580 pIA900 was a generous gift from Irina Artsimovitch (Addgene #104401)(Svetlov and Artsimovitch,
581 2015). pIA900 was transformed into BL21(λDE3)*recA*$^-$ cells. A single colony was isolated and streaked
582 on an ampicillin plate (100 mg L^{-1}). Four clones were then tested for small scale protein expression (25
583 mL) and were found to produce similar levels of overexpressed RNAP. Colony 1 was then used for
584 overproduction. 6 L of culture were inoculated with an overnight culture of colony 1 and shaken for
585 2.5 h at 37 °C to reach A_{595} = 0.6, followed by induction with a final concentration of 1 mM IPTG.
586 Following induction, cultures were grown for an additional 3 h followed by pelleting. Overproduced
587 protein was purified from cells following the exact protocols described in (Svetlov and Artsimovitch,
588 2015). Following chromatographic separation over the Mono-Q column, two fractions were isolated,
589 one corresponding to the RNAP core complex, and a second smaller fraction corresponding to the
590 holoenzyme. For this work, the RNAP core complex was isolated at a concentration of 35.7 mg mL^{-1}
591 corresponding to 92 μM in storage buffer (50 mM Tris/HCl pH 7.6, 100 mM NaCl, 1 mM EDTA, 1 mM
592 dithiothreitol and 50% *v/v* glycerol). Purified protein was stored at -80 °C following snap freezing with
593 liquid N_2. Aliquots in use were stored at -20 °C.

594 Overproduction and purification of Mfd variants

595 Sequence verified *mfd-ypet, mfd(E730Q)-ypet* and *mfd(R953A)-ypet* genes were obtained from
596 Aldevron. Plasmids for overexpression of Mfd, Mfd(E730Q) and Mfd(R953A) were created as follows:
597 first, *mfd* or mutant gene inserts (lacking the *ypet* gene) were amplified using PCR from the custom
598 synthesized constructs using the primers (pETMCSII_Mfd_F and pETMCSII_Mfd_R; Table 1). PCR
599 products were digested with with *Bam*HI and *Eco*RI and purified over a Qiagen miniprep kit. The *mfd*
600 (or mutant) fragments were inserted between the same restriction sites in vector pSH1017 (Hamdan
601 et al., 2002), previously isolated using electrophoretic separation and an extraction from a 1% (w/v)

602 agarose gel. Ligations were performed with T4 DNA ligase at 16 °C overnight. Reactions were then
603 transformed into *E. coli* DH5α cells, colonies were screened for the presence of insert by colony PCR
604 and intact *mfd* sequences in plasmids pHG4137 (coding for wildtype Mfd), pHG42305 (Mfd(R953A))
605 and pHG42306 (Mfd(E730Q)) were confirmed by sequencing.

606 A generalized procedure for the expression and purification of Mfd variants used in this study
607 is best exemplified by the methods used in the production of Mfd(E730Q). *E. coli* strain
608 BL21(λDE3)*recA*/pHG42306 (E730Q) was grown at 37 °C in LB medium supplemented with thymine
609 (50 mg L^{-1}) and ampicillin (200 mg L^{-1}) to A_{600} = 0.6. To induce Mfd(E730Q) overproduction, 1 mM IPTG
610 was added to the shaking culture. Cultures were grown for a further 3 h, and then chilled in ice. Cells
611 were harvested by centrifugation (11,000 x *g*; 6 min), frozen in liquid N_2 and stored at –80 °C.

612 After thawing, cells (17.1 g, from 6 L of culture) were resuspended in 120 mL lysis buffer (50
613 mM Tris/HCl, pH 7.6, 2 mM dithiothreitol, 0.5 mM EDTA, 20 mM spermidine) and lysed by two
614 passages through a French press operated at 12000 psi. Cell debris was removed from the lysate by
615 centrifugation (35,000 x *g*; 30 min) to yield the soluble Fraction I. Proteins in Fraction I that were then
616 precipitated by addition of solid ammonium sulfate (0.38 g mL^{-1}), followed by stirring for 30 min, were
617 collected by centrifugation (38,000 x *g*; 30 min) and dissolved in 45 mL buffer A (50 mM Tris/HCl, pH
618 7.6, 2 mM dithiothreitol, 0.5 mM EDTA, 10% *v/v* glycerol) containing 180 mM NaCl. The solution was
619 dialyzed against 2 L of the same buffer to yield Fraction II.

620 Fraction II was divided into two sub-fractions (23 mL each), which were successively applied
621 at 1 mL min^{-1} onto a column (2.5 x 15 cm) of Toyopearl DEAE-650M resin that had been equilibrated
622 in buffer A + 190 mM NaCl. Fractions containing proteins that did not bind to the resin (50 mL per sub-
623 fraction; 100 mL in total) were pooled and dialyzed against two changes of 2 L of buffer A to yield
624 Fraction III.

625 The dialysate (Fraction III) was loaded at 1 mL min^{-1} onto the same DEAE column that had been
626 equilibrated with buffer A. After the column was washed with 80 mL of the same buffer, Mfd(E730Q)
627 was eluted at 2.5 mL min^{-1} using a linear gradient (700 mL) of 0 – 500 mM NaCl in buffer A. The

628 Mfd(E730Q) eluted in a single, relatively narrow, peak centered around 80 mM NaCl. Fractions
629 containing Mfd(E730Q) (40 mL) were pooled and dialyzed against four changes of 1.2 L of buffer B (50
630 mM Na/MES, pH 6.5, 2 mM dithiothreitol, 0.5 mM EDTA, 10% *v/v* glycerol) to yield Fraction IV.

631 Fraction IV (40 mL) was loaded at 1.5 mL/min onto a 2.5 x 13 cm SP Sepaharose FastFlow
632 column (GE Healthcare) that had been equilibrated with buffer B. After washing the column with 80
633 mL of the same buffer B, Mfd(E730Q) was eluted at 2.5 mL min^{-1} using a linear gradient (700 mL) of
634 0–500 mM NaCl in buffer B. The Mfd(E730Q) eluted in a single peak centered around 100 mM NaCl.
635 Fractions containing highly purified Mfd(E730Q) were pooled and dialyzed against four changes of 1.2
636 L of storage buffer (40 mM Tris/HCl, pH 7.6, 100 mM NaCl, 4 mM dithiothreitol, 0.5 mM EDTA, 30%
637 *v/v* glycerol) to yield final product containing 130 mg of protein at a final concentration of 46.7 μM.
638 The same protocol yielded 51.5 mg of Mfd(R953A) at a final concentration of 18 μM. Final protein
639 preparations were snap frozen in liquid N_2 and stored at –80 °C.

640 In case of wildtype Mfd, a modified purification procedure was used whereby the fractions
641 containing Mfd from the second DEAE (40 mL) column were pooled and directly applied at 1 mL min^{-}
642 ^{1} onto a 2.5 x 10 cm heparin Sepharose 4B that had been equilibrated with buffer A + 50 mM NaCl.
643 Mfd was eluted at 1 mL min^{-1}, in a broad peak centered around ~120 mM NaCl, using a linear gradient
644 (450 mL) of 50–400 mM NaCl in buffer A. Fractions containing highly purified Mfd were pooled and
645 dialyzed against 2 L of storage buffer (40 mM Tris/HCl, pH 7.6, 100 mM NaCl, 4 mM dithiothreitol, 0.5
646 mM EDTA, 25% *v/v* glycerol) to yield 115 mg of protein at a final concentration of 40 μM. Purified Mfd
647 was snap frozen in liquid N_2 and aliquots were stored at –80 °C.

648 NusG was a generous gift from Prof Robert Landick (University of Wisconsin-Madison) (Kang
649 et al., 2018).

650

651 **Table 1: Nucleic Acids used in this study**

Experiment	DNA	Sequence
pETMCSII_Mfd_F	Cloning primer	GTT TAA TCG GAT CCT AAG GAG GTT AAT TCC CGC TAT GCC TGA ACA ATA TCG TTA TAC G
pETMCSII_Mfd_R	Cloning primer	GGGAGCTCGAATTCTTAAGCGATCGCGTTCTCT
smFRET (Figures 1–3)	Non-template (NT70_bio)	ATC GAG CAA CTA CTC AGA CAG CAC TAC TGC GAC TTA CAG ACA TCG AGA GGG TAA TGG CGA ATA GCA CTG A /3 bioTEG/
smFRET (Figures 1–3)	Template (T70_31Cy3)	TCA GTG CTA TTC GCC ATT ACC CTC TCG ATG T/iCy3/CT GTA AGT CGC AGT AGT GCT GTC TGA GTA GTT GCT CGAT
smFRET (Figures 1–3)	RNA(R15_4Cy5)	rArUrArU/iCy5/rArU rArUrC rGrArG rArGrG
smFRET (Figure 4)	bio_AS18_Cy5	/5Biosg/ TGG CGA CGG CAG CGA GGC/3Cy5Sp/
smFRET (Figure 4)	Cy3_dT40_S18	/5Cy3/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTG CCT CGC TGC CGT CGC CA
SPR (Supplemental Figure 4)	bio_AS18	/5Biosg/ TGG CGA CGG CAG CGA GGC
SPR	dT40_S18	TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTG CCT

(Supplemental Figure 4)		CGC TGC CGT CGC CA
SPR (Supplemental Figure 4)	bio_dT40_S18	/Biosg/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTG CCT CGC TGC CGT CGC CA
SPR (Supplemental Figure 4)	AS18	TGG CGA CGG CAG CGA GGC
SPR (Supplemental Figure 2)	TS_49	TCA GTG CTA TTC GCC ATT ACC CTC TCG ATG T/iCy3/CT GTA AGT CGC AGT AGT G
SPR (Supplemental Figure 2)	NT_49_bio	CAC TAC TGC GAC TTA CAG ACA TCG AGA GGG TAA TGG CGA ATA GCA CTG A /3 bioTEG/
SPR (Supplemental Figure 5)	NT_49_bio_bubble	CAC TAC TGC GAC TTA CAG ACT TTT TTT TTG TAA TGG CGA ATA GCA CTG A /3BioTEG/
SPR (Supplemental Figure 2)	RNA_15	rArUrAr UrArU rArUrC rGrArG rArGrG

652

653 All oligonucleotides were purchased from Integrated DNA Technologies (IDT, Singapore). Primed DNA
654 template for smFRET and cryo-EM was first annealed by hybridizing bio_AS18_Cy5 and Cy3_dT40_S18
655 in 10 mM Tris/HCl 7.6, 50 mM NaCl, 5 mM $MgCl_2$ at a final concentration of 10 μM of Cy3_dT40_S18

656 and 1.05-fold excess of the bio_AS18_Cy5 complementary strand.

657 **Single molecule FRET measurements**

658 smFRET TIRF microscope setup:

659 A home built objective-type TIRF microscope based on a Nikon TiE2 model was used to record single
660 molecule movies. FRET was measured by excitation with a 532 nm laser and the emissions at 555 and
661 647 nm were collected using a band-pass filter at 555 nm and a long-pass filter at 650 nm. Scattered
662 light was removed by using a 560 nm long pass filter. Cy3 and Cy5 signals were separated by a 638 nm
663 dichroic using photometrics dual view (DV-2) and both signals were focused onto a CCD camera
664 (Hamamatsu C9 100–13), simultaneously. Data were collected at 5 frames s^{-1}.

665 Preparation of flow cells for imaging:

666 Coverslips were prepared as described previously (Johnston et al., 2020). Quartz coverslips were
667 cleaned with 100% ethanol and 1 mM KOH. Aminosilanization of cleaned coverslips was carried out
668 using a 1% (*v/v*) (3-aminopropyl)triethoxy silane (Alfa Aesar, A10668, UK) solution in acetone.
669 PEGylation was carried out by incubating a mixture of biotinPEG-SVA MW 5000 and mPEG-SVA MW
670 5000 (Laysan Bio, AL) at a ratio of 1:10 prepared in 50 mM MOPS pH 7.5 solution on the top of the
671 silanized coverslip for 3–4 h. PEGylated coverslips were stored under dry nitrogen gas at −20 °C.

672 Neutravidin solution was prepared in transcription buffer (1 × TB; 20 mM Tris/HCl pH 8.0, 40
673 mM KCl, 5 mM $MgCl_2$ and 1 mM β-mercaptoethanol) and spread on the top of dry a PEGylated
674 coverslip followed by a 10 min incubation. Sample flow chambers were created by sandwiching
675 polydimethylsiloxane (PDMS) on the top of the streptavidin coated coverslip. Then, blocking buffer (1
676 × TB + 0.25% (*v/v*) Tween 20) was injected into the channel to reduce non-specific binding of proteins
677 on the surface followed by 10–15 min incubation. This was then followed by washing with 1 × TB to
678 remove unbound Tween 20.

679 *In situ* reconstitution of ECs for smFRET measurements:

680 Protocols described by (Komissarova et al., 2003) were adapted for *in situ* reconstitution of ECs in
681 single-molecule FRET studies. 1 μM template DNA was annealed to 3 μM RNA primer in 1 × TB by

682 heating to 65 °C for 5 min followed by 3 °C drops in temperature every 2 min until the reaction
683 temperature reached 27 °C, followed by a hold step at 25 °C. Typically, 200 μL of reaction were
684 prepared at a time. To form the RNAP:template (T70_31Cy3):RNA(R15_4Cy5) complex, 0.22 μL of 92
685 μM RNAP was incubated with 2 μL of template:RNA hybrid in 1 × TB at room temperature for 30 min.
686 Meanwhile, 200 μL of 100 pM of NT70_bio (Table 1) were immobilized on the neutravidin-coated
687 surface. Following this, the RNAP:template:RNA mixture was diluted serially 6000 × in a final
688 suspension of 200 μL of 1 × TB and introduced into the flow cell and incubated for at least 5 min to
689 promote annealing of the template DNA to the biotinylated non-template DNA resulting in formation
690 of the EC. At the end of this incubation step, the flow cell was washed with 200 μL of 1 × TB + 1 M KCl
691 and subsequently with 1 × TB supplemented with an oxygen-scavenging system (OSS) consisting of
692 protocatechuic acid (2.5 mM) and protocatechuate-3,4-dioxigenase (50 nM) to reduce photo-
693 bleaching of the fluorophores and 2 mM Trolox to reduce photo-blinking of dyes, followed by imaging.
694 All elongation experiments in Figures 1 and 2 were performed at 100 μM of each NTP or dNTP (where
695 indicated). smFRET experiments involving NusG were performed at 500 nM of NusG in 1 × TB (Figure
696 2). smFRET experiments involving wildtype Mfd (Figures 3 and 4) were performed at 500 nM of Mfd
697 in 1 × TB . Experiments involving Mfd(R953A) (Figure 3 and 4) and Mfd(E730Q) (Figure 4) were
698 performed at 1 μM of Mfd mutant in 1 × TB . Concentrations of ATP and ATPγS were at 1 mM and 0.5
699 mM respectively unless otherwise specified.

700 Primed DNA template (Figure 4):

701 Primed DNA substrate (bio_AS18_Cy5: Cy3_dT40_S18) used in smFRET experiments in Figure 4 was
702 diluted to 50 pM in 200 μL of 1 × TB prior to immobilization in the flow cell.

703 Data Analysis

704 Single-molecule data acquisition was carried out using NIS element software. Then, single-molecule
705 intensity time trajectories and FRET-time trajectories were generated in MASH-FRET software
706 (https://rna-fretools.github.io/MASH-FRET). Further data analysis and figures are plotted using
707 custom written Python scripts.

708

709 **Surface plasmon resonance**

710 Activation of SPR chip:

711 SPR experiments were performed using a BIAcore T200 instrument (GE Healthcare) using a
712 streptavidin (SA) coated sensor chip to study the binding kinetics of Mfd to reconstituted ECs, or
713 various DNA substrates. All experiments were carried out at 20 °C with a flow rate of 5 μL min^{-1}, unless
714 specified otherwise. The SA chip was activated with three sequential 1 min injections of 1 M NaCl, 50
715 mM NaOH, then stabilized by 1 min treatment with 1 M $MgCl_2$.

716 Immobilization of 49-mer dsDNA substrates (Supplemental Figure 2):

717 49-mer dsDNA substrate was prepared by hybridizing NT_49_bio (final concentration of 1 μM) with
718 two-fold excess of TS_49 in 1 × TB in a final volume of 120 μL by heating in a heat block at 90 °C for 5
719 min followed by a slow cooling step until the room temperature was reached. NT_49_bio:TS_49
720 dsDNA substrate was resuspended in 1 × SPR buffer (30 mM Tris/HCl, pH 7.6, 40 mM KCl, 5 mM $MgCl_2$,
721 0.005% (*v/v*) surfactant P20, 0.25 mM EDTA, 0.5 mM dithiothreitol) to a final concentration of 2.5 nM
722 (of the biotinylated oligo) and introduced onto the SPR chip for immobilization, followed by a wash
723 step with 1M KCl. The signal from the dsDNA template corresponded to 106 RUs. Measurement of the
724 Mfd binding affinity for the immobilized dsDNA ligand was performed as described below.

725 Immobilization of reconstituted EC on biotinylated 49-mer dsDNA (Supplemental Figure 2):

726 The RNAP:template (TS_49):RNA(RNA_15) complex was formed using the procedure described for *in*
727 *situ* reconstitution of ECs for smFRET measurements. Following this, biotinylated non-template strand
728 (NT49_bio; 10 μL of 10 μM working stock in water) was added to the mixture and the reaction was
729 incubated at room temperature for 15 min. Next, 50 μL of nickel resin suspension were washed and
730 equilibrated with 1 × TB. This resin resuspended in 50 μL of 1 × TB was then added to the mixture
731 containing the RNAP, template:RNA hybrid and non-template strand followed by incubation at 37 °C
732 for 15 min to allow binding of His-tagged RNAP complexes to the resin. The remaining steps were then
733 performed at room temperature. First, supernatant from this suspension was isolated by

734 centrifugation on a mini-benchtop centrifuge for 10 s. The pelleted resin was washed with 150 μL of
735 1 × TB supplemented with 1 M KCl for 2 min. This suspension was then subjected to centrifugation
736 and the supernatant was discarded. Following this, 150 μL of elution buffer (1 × TB + 500 mM
737 imidazole) were then applied to the resin followed by gentle mixing by pipetting. Finally, the
738 suspension was centrifuged and the eluate containing reconstituted elongation complexes was
739 collected. All reactions containing RNAP were performed in LoBind microcentrifuge tubes (Eppendorf).
740 The eluate was then applied to a size exclusion column (Wyatt SEC analytical column 030S5) previously
741 equilibrated with 30 mM Tris/HCl, pH 7.6, 150 mM KCl, 15% (*v/v*) glycerol, 0.5 mM EDTA, 1 mM
742 dithiothreitol to separate biotinylated elongation complexes from RNAP alone and collected in 200
743 μL fractions. Fraction 7 (F7) was found to contain the elongation complex as verified by smFRET
744 studies. Immobilization of these reconstituted ECs was performed by resuspending 40 μL of F7 in 200
745 μL of SPR buffer and introduction into the flow cell. Immobilization was stopped when the signal
746 reached 1380 RU over the background. This was followed by treatment with 1M KCl (10 μL min^{-1}) that
747 served to displace incorrectly assembled or non-specifically bound RNAP from the chip surface. The
748 resulting signal corresponded to 550 RU over the background. ECs thus immobilized were found to be
749 stable for several days. Measurement of Mfd binding affinity for the immobilized EC ligand was
750 performed as described below.

751 <u>Immobilization of primed DNA (Supplemental Figure 4)</u>:

752 Primed DNA templates were assembled *in situ* (on the chip surface) in two stages. First, the
753 biotinylated strand was immobilized in three of the four channels in the chip. Single-stranded
754 biotinylated DNA substrates (bio_dT40_S18, bio_AS18, and bio_AS18_Cy5) were diluted to 1 nM in
755 SPR buffer and immobilized in channels 2, 3, and 4 respectively. This corresponded to 150 RUs of
756 bio_dT40_S18 (channel 2), 75 RUs of bio_AS18 (channel 3), and 80 RUs of bio_AS18_Cy5 (channel 4).
757 Next, the complementary non-biotinylated strand (2 μM in 1 × SPR buffer) was introduced into the
758 flow cell to hybridize the template *in situ* at 5 μL min^{-1}. This led to a further increase in signal
759 corresponding to 39 RUs (channel 2, complementary strand: AS18), 138 RUs (channel 3,

760 complementary strand: dT40_S18) and 115 RUs (channel 4, complementary strand: Cy3_dT40_S18)
761 upon complete hybridization. Following this step, the flow cell was washed with 1 M $MgCl_2$.
762 Measurement of Mfd binding affinity to immobilized primed DNA template was performed as
763 described below.

764 Titration of Mfd (Supplemental Figures 2 and 4):

765 Following the immobilization of the DNA substrates or the EC, in each case, binding studies were
766 performed by injecting increasing amounts of Mfd in the indicated concentration range in the SPR
767 buffer supplemented with 0.5 mM ATPγS. For measurements of Mfd binding to dsDNA (Supplemental
768 Figure 2: NT_49_bio: TS_49), reconstituted EC (Supplemental Figure 2) and primed DNA substrates,
769 an optimized concentration range of [0, 10, 20, 40, 80, 160, 320, 640] μM of Mfd was used. For
770 measurements of Mfd binding to ssDNA substrate (bio_dT40_S18), an optimized concentration range
771 of [0, 0.25, 0.5, 1, 2, 4, 8, 16] μM was used.

772 Two conditioning steps of flushing the flow cell with 300 mM KCl for 30 s at 10 μL min^{-1} were carried
773 out prior to injection of Mfd. Mfd-ATPγS was flowed in for 150 s in the association phase, and
774 dissociation was observed for 200 s in 1 × SPR buffer lacking any Mfd or nucleotide cofactor.

775 Analysis of sensorgrams:

776 Sensorgrams were subtracted from the signal from unmodified flow cell (Flow channel 1), then zero-
777 subtracted (injection at 0 μM Mfd or mutant) and R_{eq} values, generated by averaging response values
778 in the sensorgram's steady-state region from the appropriate Mfd concentration range, fit against
779 [Mfd] using a 1:1 steady state affinity (SSA) model incorporated in the BIAevaluation software 4.0.1:

780 $$R_{\mathrm{eq}} = R_{\mathrm{max}}\left(\frac{[A]}{[A]+K_{\mathrm{D}}}\right) \text{.....} (1)$$

781 where R_{max} corresponds to the response when all the immobilized DNA ligands on the surface are
782 saturated with the analyte A, K_D is the dissociation constant, and [A] is the concentration of analyte
783 (Mfd or mutant) in solution. In general, raw sensorgrams and fit curves (where applicable) were
784 exported using Excel, then plotted in Origin, and final Figures prepared in Adobe Illustrator CC.

785 **Sample preparation for electron microscopy**

786 Electron microcopy data acquisition:

787 30 μL of Mfd stock (40 μM) were dialyzed overnight in 1 L of 1 × TB at 4 °C in a Slida-A-lyzer mini
788 dialysis unit with a 3,000 MWCO (Thermo Fisher) to reduce the amount of glycerol in the sample.
789 Following this, protein was recovered and concentrated to a 1.67 mg mL^{-1}. For the sample with ADP-
790 AlF_x, ADP, NaF and $AlCl_3$ were added to the mixture of Mfd (7.8 μL at 1.6 mg mL^{-1}) and fluorescent
791 oligonucleotides (10 μL of 10 μM stock) to final concentrations of 2, 5 and 0.5 mM, respectively. This
792 reaction mixture was incubated at 37 °C for 5 min to allow binding of Mfd to the template. The samples
793 were centrifuged at 20,000 × g for 10 min at 4 °C to clear any precipitate before grid preparation.
794 Samples (3 μL) were applied to glow-discharged Quantifoil R 1.2/1.3, Au 400 grids. The grids were
795 blotted at 6 °C for 3.5 s at 100% humidity with no extra blot force on an FEI Vitrobot Mark IV and
796 plunged into liquid ethane. The grids were stored in liquid nitrogen until use.

797 Electron microcopy data acquisition:

798 Data were collected using FEI Talos Arctica microscope at 200 kV equipped with a Falcon III (FEI)
799 camera in electron counting mode. Micrographs were acquired at a nominal defocus range of −1.0 to
800 −1.9 μm, with a pixel size of 0.74 Å and a total dose of 50 e^{-} $Å^{-2}$ accumulated over 50 s and fractionated
801 across 50 frames.

802 Image processing of EM micrographs

803 Data were processed in cryoSPARC v2 (Punjani et al., 2017). For the sample with ADP-AlF_x, 472 images
804 were manually selected after motion correction and CTF estimation for final processing. 110562
805 particles were selected after 2D classification and subjected to 3D classification in *ab initio*
806 reconstruction followed by heterogeneous refinement using four *ab initio* classes. 31718 particles
807 from class 3 were used for 3D refinement, which resulted in a 3D reconstruction of 5.2 Å (0.143 FSC
808 gold standard).

809 Model fitting

810 Molecular dynamics with explicit solvent was used to fit the model of Mfd on primer DNA into the
811 cryo-EM density map (Singharoy et al., 2016). The starting model was comprised of the Mfd motor
812 domain (PDB ID 2eyq) with idealized B-DNA with the same sequence as used for the cryo-EM
813 experiments but with a dT(12) overhang instead of dT(40). The system was equillibrated for 10 ns. The
814 resulting model was then used as the basis for modelling the complex in COOT (Emsley et al., 2010)
815 with minimal distortions to protein secondary structure and DNA. The model was then subjected to
816 restrained positional refinement in phenix real_space_refine (Liebschner et al., 2019). The final model
817 is comprised of 382 residues from Mfd and 47 nucleotide residues. Model was visualized in
818 Chimera(Pettersen et al., 2004).

819

820

821 **Figure legends**

822 **Figure 1: Reconstitution of a functional EC.**

823 (A) ECs were assembled on dye-labelled nucleic acids by flowing in pre-incubated RNAP-template:RNA
824 complexes into a flow cell with immobilized biotinylated non-template strand (EC shown in grey).
825 Following initial immobilization, a 1 M KCl wash step was introduced to remove poorly hybridized ECs
826 from the surface. Here, only complexes containing both the Cy3-labelled template strand and Cy5-
827 labelled transcript will be observed when stably associated with the biotinylated non-template strand
828 (shown in orange). Two example FRET trajectories and temporal heat maps are provided for RNAP (B,
829 C) (n = 318 molecules), RNAP in the presence of dNTPs (D, E) (n = 80 molecules), RNAP with non-
830 complementary ('wrong') NTPs (F, G) (n = 179 molecules) and the full set of NTPs (H, I) (n = 81
831 molecules) respectively. Fits to detected states in the FRET trajectories are shown in red. The sequence
832 of the four next correct incorporations is G, U, A and A.

833 **Supplemental Figure 1: Walking the EC with subsets of correct rNTPs. *Related to Figure 1.***

834 (A) Transition density plots for RNAP alone (n = 318 molecules), in the presence of dNTPs (n = 80
835 molecules), with non-complementary ('wrong') NTPs (n = 179 molecules) and the full set of NTPs (n =
836 81 molecules) respectively.

837 (B) Schematic of *in situ* assembled FRET pair labelled ECs incubated with GTP (the next correct NTP) (n
838 = 63 molecules) or with GTP, ATP and UTP (n = 87 molecules). Temporal heat maps of the reaction for
839 experiments conducted in the presence of GTP (C) or GTP, ATP and UTP (D) are presented here.

840 **Figure 2: Interactions between NusG and the EC**

841 (A) Schematic of FRET pair labelled ECs incubated with NusG (purple). Example FRET trajectories and
842 temporal heat maps for RNAP and NusG alone (B, C) (n = 461 molecules), RNAP and NusG in the
843 presence of wrong NTPs (D, E) (n = 268 molecules), and RNAP and NusG with the full set of NTPs (F, G)

844 (n = 425 molecules). Here, NusG promotes the residence of the EC in a catalytically poised state (EC
845 shown in blue).

846 **Supplemental Figure 2: Measurement of binding affinity of Mfd for dsDNA and the EC. *Related to***
847 ***Figures 2 and 3.***

848 (A)Transition density plots for RNAP and NusG alone (n = 461 molecules), RNAP and NusG in the
849 presence of wrong NTPs (n = 268 molecules), and RNAP and NusG with the full set of NTPs (n = 425
850 molecules). Here, NusG promotes the residence of the EC in a catalytically poised state (EC shown in
851 blue). Red ellipse highlights unidirectional transitions.

852 Schematic and SPR sensorgrams show association (150 s) and dissociation phases of serially-diluted
853 10 – 640 μM Mfd-ATPγS, including a 0 μM control for (B) biotinylated 49-mer dsDNA, and (C) *in vitro*
854 reconstituted EC on biotinylated 49-mer DNA flown into the SPR chip. Responses at equilibrium
855 averaged over the steady region of sensorgrams (R_{eq}) were fit (insets) using a steady-state affinity
856 (SSA) model to derive values of K_D (as indicated). Errors are *S.E* of the fit.

857 **Figure 3: Interactions between Mfd and the EC.**

858 (A) Schematic of FRET pair labelled surface immobilized ECs incubated with Mfd in solution. Two
859 example FRET trajectories and temporal heat maps are presented for RNAP and Mfd alone (B, C) (n =
860 75 molecules), RNAP and Mfd in the presence of 1 mM ATPγS (D, E) (n = 68 molecules), RNAP and Mfd
861 in the presence of ATP (F, G) (n = 174 molecules), and RNAP and DNA-binding competent but
862 translocase-deficient Mfd(R953A) in the presence of 1 mM ATP (H, I) (n = 234 molecules).

863 **Supplemental Figure 3: Interactions of ATP with the EC alone or in the presence of Mfd. *Related to***
864 ***Figure 3.***

865 (A) Transition density plots for RNAP and Mfd alone (n = 75 molecules), RNAP and Mfd in the presence
866 of 1 mM ATPγS (n = 68 molecules), RNAP and Mfd in the presence of ATP (n = 174 molecules), and

867 RNAP and DNA-binding competent but translocase-deficient Mfd(R953A) in the presence of 1 mM ATP
868 (n = 234 molecules). Red ellipses highlight irreversible transitions.

869 (B) Temporal heat map (upper panel) and transition density plot (lower panel) of FRET pair labelled
870 RNAP EC in the presence of 0.5 mM ATPγS alone (n = 124 molecules).

871 (C) Distribution of the catalytically-poised mid-FRET state for the indicated concentrations of ATP and
872 (D) Dependence of the lifetime (τ_{off}) of mid-FRET state for Mfd-RNAP ECs as a function of ATP
873 concentration in the range from 10 μM to 1 mM ATP (n = 43, 24, 77, 32 molecules respectively). Plots
874 represent dwell-time distributions of the surviving fraction of the population in the mid-FRET state fit
875 to a single exponential decay (red curve) to the data. Errors represent S.E to the fit.

876 (E) Temporal heat maps of FRET pair labelled RNAP EC when titrated with ATP in the range of 0.1 μM
877 to 1 mM (n = 151, 82, 81, 76, 157 molecules respectively) as indicated.

878 **Figure 4: Interactions between Mfd and a primed-DNA template**

879 (A) Schematic of FRET pair labelled primed DNA template comprised of an 18-mer dsDNA and a 5' 40-
880 mer poly-dT overhang to assess interactions of Mfd with ssDNA. Two example FRET trajectories,
881 temporal heat maps and transition density plots for primed DNA alone (B, C) (n = 100 molecules), Mfd
882 in the presence of ATPγS (D, E) (n = 94 molecules), Mfd in the presence of ATP (F, G) (n = 203
883 molecules), and Mfd(R953A) in the presence of ATP (H, I) (n = 105 molecules) are presented here.

884 **Supplemental Figure 4: Interactions between Mfd and a primed-DNA template.** ***Related to Figure 4.***

885 (A)Transition density plots for FRET pair labelled primed DNA alone (n = 100 molecules), or bound to
886 Mfd in the presence of ATPγS (n = 94 molecules), or ATP (n = 203 molecules), and Mfd(R953A) in the
887 presence of ATP (n = 105 molecules) are presented here. DNA substrate is the same as described in
888 Figure 4.

889

890 Schematic and SPR sensorgrams show association (150 s) and dissociation phases of serially-diluted
891 10 – 640 μM Mfd-ATPγS, including a 0 μM control for (B) primed DNA substrate (biotinylated 18-mer
892 dsDNA with dT_{40} overhang) (C) "flipped" primed DNA substrate (18-mer dsDNA with biotinylated dT_{40}
893 overhang) (D) FRET pair labelled primed DNA substrate (biotinylated 18-mer dsDNA with dT_{40}
894 overhang) and (E) 58-mer ssDNA containing the dT_{40} sequence. Responses at equilibrium averaged
895 over the steady region of sensorgrams (R_{eq}) were fit (insets) using a steady-state affinity (SSA) model
896 to derive values of K_D (as indicated). Errors are *S.E* of the fit.

897 (F) Temporal heat map and (G) transition density plot for ATPase mutant Mfd(E730Q) binding to FRET
898 pair labelled primed DNA template used in Figure 4 (n = 103 molecules).

899 (H) For convenience, kernel density estimations of the ensemble FRET distributions read as projections
900 along the ordinate in the heat-maps presented in Figure 4 (panels C, E, G, I) are presented to enable
901 direct visual comparisons.

902 **Figure 5: Interactions between Mfd and primed DNA**

903 (A) Views of the cryo-EM reconstruction of Mfd bound to the FRET pair labelled DNA substrate (used
904 in Figure 4).

905 (B) Front view of the molecular model derived from the cryo-EM reconstruction showing domain 5
906 (tan), domain6 (olive), template (dark red) and non-template (pink) DNA. Mfd (residues 575-956) and
907 DNA (template: residues 28-58; non-template: 1-18) are shown in surface representation to illustrate
908 the path of the DNA.

909 (C) Top view of molecular model showing TRG motif straddling the DNA. The unpaired 5' poly dT
910 overhang is colored as follows: dT1–dT3 (pink), dT4–dT6 (salmon), dT7–dT9 (sandy brown) and
911 dT11–dT12 (brown). R953 and R929 are positioned to contact the ssDNA overhang at 3 and 6 residues
912 away from the junction respectively.

913 (D) Model suggests dsDNA near the junction is melted to form the TRCF bubble. (F) Path of the DNA is
914 lined by highly conserved residues in Mfd across bacteria: basic residues (Lys, Arg, His) are shown in
915 blue, polar residues (Thr, Asn) are shown in gold, and aromatic residues (Phe, Tyr) are shown in green.
916 Domain 5 corresponds to residues 549 – 780, and domain 6 corresponds to residues 781-990 in Mfd.
917 Labels are provided in (G).

918 **Supplemental Figure 5: EM reconstruction of Mfd.** ***Related to Figure 5.***

919 (A) Overview of image processing using CryoSPARC. 110562 particles were selected after 2D
920 classification for *ab initio* reconstruction and heterogeneous refinement leading to four classes of 3D
921 reconstructions. Of these, 31718 particles that yielded the most complete reconstruction were
922 subjected to homogenous refinement to generate the map presented here. The colormap applied to
923 the reconstruction represents local resolution. (B) Gold-standard Fourier shell correlation curves of
924 the reconstruction reported a resolution of 5.2 Å (C) angular distribution of particle projections (D)
925 Measurement of dissociation constant of Mfd for a 49-mer dsDNA template containing a 9 nucleotide
926 non-complementary bubble where the complementary residues were replaced by dT_9. Errors reflect
927 S.E of the fit.

928 **Figure 6: Model for Mfd-mediated transcription modulation**

929 Mfd interacts with the (A) elongation complex weakly in the absence of ATP, however, (B, C) ATP
930 binding and hydrolysis are required for stable association of Mfd with the EC. Following, a brief period
931 of ATP-hydrolysis dependent translocation of Mfd on upstream dsDNA, Mfd loads on to the EC. This
932 results in chaperoning of the EC from catalytically inactive state(s) into a catalytically poised molecular
933 timer intermediate (D) with a lifetime of ~6 s. In this ATP-bound state, we propose that Mfd melts the
934 upstream DNA to form an extended transcription-repair coupling factor – transcription ("TRCF-T")
935 bubble. In this state R953 contacts the template ssDNA in the TRCF-T bubble. Melting of this upstream
936 DNA is sufficient to promote residence of the EC in a catalytically poised state. (E) In the presence of

937 NTPs on an undamaged template, transcription can be resumed. (F) On damaged templates or under
938 conditions of nucleotide starvation or in the presence of impediments to the progress of the EC, the
939 bound ATP is hydrolyzed to promote transcription termination. In our experiments this is seen as the
940 loss of RNA. We propose that in either case continued translocation occurs in a manner in which the
941 ATPase activity of Mfd translocates the TRCF bubble forward until Mfd interacts with its interacting
942 partners in the repair machinery.

943

944 **References**

945 Abdelkareem, M., Saint-Andre, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J., and
946 Weixlbaumer, A. (2019). Structural Basis of Transcription: RNA Polymerase Backtracking and
947 Its Reactivation. Mol Cell *75*, 298-309 e294.
948 Andrecka, J., Lewis, R., Bruckner, F., Lehmann, E., Cramer, P., and Michaelis, J. (2008). Single-
949 molecule tracking of mRNA exiting from RNA polymerase II. Proc Natl Acad Sci U S A *105*, 135-
950 140.
951 Belogurov, G.A., and Artsimovitch, I. (2019). The Mechanisms of Substrate Selection,
952 Catalysis, and Translocation by the Elongating RNA Polymerase. J Mol Biol *431*, 3975-4006.
953 Borukhov, S., Sagitov, V., and Goldfarb, A. (1993). Transcript cleavage factors from *E. coli*. Cell
954 *72*, 459-466.
955 Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., Jr., and Wiggins, C.H. (2009). Learning Rates
956 and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-
957 Molecule FRET Data. Biophys J *97*, 3196-3205.
958 Brugger, C., Zhang, C., Suhanovsky, M.M., Kim, D.D., Sinclair, A.N., Lyumkis, D., and
959 Deaconescu, A.M. (2020). Molecular determinants for dsDNA translocation by the
960 transcription-repair coupling and evolvability factor Mfd. Nat Commun *11*, 3740.
961 Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L., and Gottesman, M.E. (1995). *Escherichia coli*
962 NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. Journal of
963 bacteriology *177*, 1388-1392.
964 Byrd, A.K., and Raney, K.D. (2012). Superfamily 2 helicases. Front Biosci (Landmark Ed) *17*,
965 2070-2088.
966 Chambers, A.L., Smith, A.J., and Savery, N.J. (2003). A DNA translocation motif in the bacterial
967 transcription-repair coupling factor, Mfd. Nucleic Acids Res *31*, 6409-6418.
968 Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J., and
969 Darst, S.A. (2006). Structural Basis for Bacterial Transcription-Coupled DNA Repair. Cell *124*,
970 507-520.
971 Deaconescu, A.M., Sevostyanova, A., Artsimovitch, I., and Grigorieff, N. (2012). Nucleotide
972 excision repair (NER) machinery recruitment by the transcription-repair coupling factor
973 involves unmasking of a conserved intramolecular interface. Proc Natl Acad Sci U S A *109*,
974 3353-3358.
975 Duan, M., Selvam, K., Wyrick, J.J., and Mao, P. (2020). Genome-wide role of Rad26 in
976 promoting transcription-coupled nucleotide excision repair in yeast chromatin. Proc Natl
977 Acad Sci U S A *117*, 18608-18616.
978 Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of
979 Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.
980 Fan, J., Leroux-Coyau, M., Savery, N.J., and Strick, T.R. (2016). Reconstruction of bacterial
981 transcription-coupled repair at single-molecule resolution. Nature *536*, 234-237.
982 Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R.H., and Bustamante, C.J. (2018). Pause
983 sequences facilitate entry into long-lived paused states by reducing RNA polymerase
984 transcription rates. Nat Commun *9*, 2930.
985 Ghodke, H., Ho, H.N., and van Oijen, A.M. (2020). Single-molecule live-cell imaging visualizes
986 parallel pathways of prokaryotic nucleotide excision repair. Nat Commun *11*, 1477.
987 Graves, E.T., Duboc, C., Fan, J., Stransky, F., Leroux-Coyau, M., and Strick, T.R. (2015). A
988 dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and

989 fluorescence. Nat Struct Mol Biol *22*, 452-457.
990 Haines, N.M., Kim, Y.I., Smith, A.J., and Savery, N.J. (2014). Stalled transcription complexes
991 promote DNA repair at a distance. Proc Natl Acad Sci U S A *111*, 4037-4042.
992 Hamdan, S., Bulloch, E.M., Thompson, P.R., Beck, J.L., Yang, J.Y., Crowther, J.A., Lilley, P.E.,
993 Carr, P.D., Ollis, D.L., Brown, S.E.*, et al.* (2002). Hydrolysis of the 5'-p-nitrophenyl ester of TMP
994 by the proofreading exonuclease (epsilon) subunit of Escherichia coli DNA polymerase III.
995 Biochemistry *41*, 5266-5275.
996 Herbert, K.M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., and Block, S.M. (2010). *E. coli*
997 NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward
998 Translocation of RNA Polymerase. J Mol Biol *399*, 17-30.
999 Ho, H.N., van Oijen, A.M., and Ghodke, H. (2018). The transcription-repair coupling factor Mfd
1000 associates with RNA polymerase in the absence of exogenous damage. Nat Commun *9*, 1570.
1001 Ho, H.N., van Oijen, A.M., and Ghodke, H. (2020). Single-molecule imaging reveals molecular
1002 coupling between transcription and DNA repair machinery in live cells. Nat Commun *11*, 1478.
1003 Ho Wan, K., Smith, A.J., Westblade, L.F., Joly, N., Grange, W., Zorman, S., Darst, S.A., Savery,
1004 N.J., and Strick, T.R. (2012). Initiation of transcription-coupled repair characterized at single-
1005 molecule resolution. Nature *490*, 431-434.
1006 Hopfield, J.J. (1974). Kinetic proofreading: a new mechanism for reducing errors in
1007 biosynthetic processes requiring high specificity. Proc Natl Acad Sci U S A *71*, 4135-4139.
1008 Jaciuk, M., Swuec, P., Gaur, V., Kasprzak, J.M., Renault, L., Dobrychlop, M., Nirwal, S., Bujnicki,
1009 J.M., Costa, A., and Nowotny, M. (2020). A combined structural and biochemical approach
1010 reveals translocation and stalling of UvrB on the DNA lesion as a mechanism of damage
1011 verification in bacterial nucleotide excision repair. DNA Repair (Amst) *85*, 102746.
1012 Johnston, C.L., Marzano, N.R., Paudel, B.P., Wright, G., Benesch, J.L.P., van Oijen, A.M., and
1013 Ecroyd, H. (2020). Single-molecule fluorescence-based approach reveals novel mechanistic
1014 insights into human small heat shock protein chaperone function. J Biol Chem.
1015 Kamarthapu, V., Epshtein, V., Benjamin, B., Proshkin, S., Mironov, A., Cashel, M., and Nudler,
1016 E. (2016). ppGpp couples transcription to DNA repair in *E. coli*. Science *352*, 993-996.
1017 Kang, J.Y., Llewellyn, E., Chen, J., Olinares, P.D.B., Brewer, J., Chait, B.T., Campbell, E.A., and
1018 Darst, S.A. (2021). Structural basis for transcription complex disruption by the Mfd
1019 translocase. Elife *10*.
1020 Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T.V., Artsimovitch, I., Landick, R.,
1021 and Darst, S.A. (2018). Structural Basis for Transcript Elongation Control by NusG Family
1022 Universal Regulators. Cell *173*, 1650-1662 e1614.
1023 Kisker, C., Kuper, J., and Van Houten, B. (2013). Prokaryotic Nucleotide Excision Repair. Cold
1024 Spring Harb Perspect Biol *5*, a012591.
1025 Komissarova, N., and Kashlev, M. (1997). Transcriptional arrest: Escherichia coli RNA
1026 polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc
1027 Natl Acad Sci U S A *94*, 1755-1760.
1028 Komissarova, N., Kireeva, M.L., Becker, J., Sidorenkov, I., and Kashlev, M. (2003). Engineering
1029 of elongation complexes of bacterial and yeast RNA polymerases. Methods Enzymol *371*, 233-
1030 251.
1031 Le, T.T., Yang, Y., Tan, C., Suhanovsky, M.M., Fulbright, R.M., Jr., Inman, J.T., Li, M., Lee, J.,
1032 Perelman, S., Roberts, J.W.*, et al.* (2017). Mfd Dynamically Regulates Transcription via a
1033 Release and Catch-Up Mechanism. Cell.
1034 Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung,
1035 L.W., Jain, S., McCoy, A.J.*, et al.* (2019). Macromolecular structure determination using X-rays,

1036 neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol *75*, 861-
1037 877.
1038 Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q., and Lloyd, R.G. (2003). A model for dsDNA
1039 translocation revealed by a structural motif common to RecG and Mfd proteins. EMBO J *22*,
1040 724-734.
1041 Manelyte, L., Kim, Y.I., Smith, A.J., Smith, R.M., and Savery, N.J. (2010). Regulation and Rate
1042 Enhancement during Transcription-Coupled DNA Repair. Mol Cell *40*, 714-724.
1043 Mayne, L.V., and Lehmann, A.R. (1982). Failure of RNA synthesis to recover after UV
1044 irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma
1045 pigmentosum. Cancer Res *42*, 1473-1478.
1046 Mellon, I., and Hanawalt, P.C. (1989). Induction of the *Escherichia coli* lactose operon
1047 selectively increases repair of its transcribed DNA strand. Nature *342*, 95-98.
1048 Muftuoglu, M., Sharma, S., Thorslund, T., Stevnsner, T., Soerensen, M.M., Brosh, R.M., Jr., and
1049 Bohr, V.A. (2006). Cockayne syndrome group B protein has novel strand annealing and
1050 exchange activities. Nucleic Acids Res *34*, 295-304.
1051 Pakotiprapha, D., Samuels, M., Shen, K., Hu, J.H., and Jeruzalmi, D. (2012). Structure and
1052 mechanism of the UvrA-UvrB DNA damage sensor. Nat Struct Mol Biol *19*, 291-298.
1053 Park, J., Myong, S., Niedziela-Majka, A., Lee, K.S., Yu, J., Lohman, T.M., and Ha, T. (2010). PcrA
1054 helicase dismantles RecA filaments by reeling in DNA in uniform steps. Cell *142*, 544-555.
1055 Park, J.S., Marr, M.T., and Roberts, J.W. (2002). *E. coli* Transcription Repair Coupling Factor
1056 (Mfd protein) Rescues Arrested Complexes by Promoting Forward Translocation. Cell *109*,
1057 757-767.
1058 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and
1059 Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and
1060 analysis. J Comput Chem *25*, 1605-1612.
1061 Proshkin, S.A., and Mironov, A.S. (2016). Stalled RNA Polymerase Is a Target of the Mfd Factor.
1062 Mol Biol (Mosk) *50*, 381-384.
1063 Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for
1064 rapid unsupervised cryo-EM structure determination. Nat Methods *14*, 290-296.
1065 Ragheb, M.N., Merrikh, C., Browning, K., and Merrikh, H. (2021). Mfd regulates RNA
1066 polymerase association with hard-to-transcribe regions in vivo, especially those with
1067 structured RNAs. Proc Natl Acad Sci U S A *118*.
1068 Ragheb, M.N., Thomason, M.K., Hsu, C., Nugent, P., Gage, J., Samadpour, A.N., Kariisa, A.,
1069 Merrikh, C.N., Miller, S.I., Sherman, D.R.*, et al.* (2019). Inhibiting the Evolution of Antibiotic
1070 Resistance. Mol Cell *73*, 157-165 e155.
1071 Selby, C.P., and Sancar, A. (1993a). Molecular Mechanism of Transcription-Repair Coupling.
1072 Science *260*, 53-58.
1073 Selby, C.P., and Sancar, A. (1993b). Transcription-Repair Coupling and Mutation Frequency
1074 Decline. Journal of bacteriology *175*, 7509-7514.
1075 Selby, C.P., and Sancar, A. (1995). Structure and Function of Transcription-Repair Coupling
1076 Factor. I. Structural Domains and Binding Properties. J Biol Chem *270*, 4882-4889.
1077 Selby, C.P., and Sancar, A. (1997). Cockayne syndrome group B protein enhances elongation
1078 by RNA polymerase II. Proc Natl Acad Sci U S A *94*, 11205-11209.
1079 Shi, J., Wen, A., Zhao, M., Jin, S., You, L., Shi, Y., Dong, S., Hua, X., Zhang, Y., and Feng, Y. (2020).
1080 Structural basis of Mfd-dependent transcription termination. Nucleic Acids Res.
1081 Singharoy, A., Teo, I., McGreevy, R., Stone, J.E., Zhao, J., and Schulten, K. (2016). Molecular
1082 dynamics-based refinement and validation for sub-5 A cryo-electron microscopy maps. eLife

1083 *5*.
1084 Smith, A.J., Pernstich, C., and Savery, N.J. (2012). Multipartite control of the DNA translocase,
1085 Mfd. Nucleic Acids Res *40*, 10408-10416.
1086 Smith, A.J., Szczelkun, M.D., and Savery, N.J. (2007). Controlling the motor activity of a
1087 transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. Nucleic
1088 Acids Res *35*, 1802-1811.
1089 Svetlov, V., and Artsimovitch, I. (2015). Purification of bacterial RNA polymerase: tools and
1090 protocols. Methods Mol Biol *1276*, 13-29.
1091 Svetlov, V., Vassylyev, D.G., and Artsimovitch, I. (2004). Discrimination against
1092 Deoxyribonucleotide Substrates by Bacterial RNA Polymerase. J Biol Chem *279*, 38087-38090.
1093 Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J.H.
1094 (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's
1095 syndrome and preferential repair of active genes. Cell *71*, 939-953.
1096 van der Weegen, Y., Golan-Berman, H., Mevissen, T.E.T., Apelt, K., Gonzalez-Prieto, R.,
1097 Goedhart, J., Heilbrun, E.E., Vertegaal, A.C.O., van den Heuvel, D., Walter, J.C.*, et al.* (2020).
1098 The cooperative action of CSB, CSA, and UVSSA target TFIIH to DNA damage-stalled RNA
1099 polymerase II. Nat Commun *11*, 2104.
1100 Venema, J., Mullenders, L.H., Natarajan, A.T., van Zeeland, A.A., and Mayne, L.V. (1990). The
1101 genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA
1102 damage in transcriptionally active DNA. Proc Natl Acad Sci U S A *87*, 4707-4711.
1103 Wang, B., and Artsimovitch, I. (2020). NusG, an Ancient Yet Rapidly Evolving Transcription
1104 Factor. Front Microbiol *11*, 619618.
1105 Xu, J., Lahiri, I., Wang, W., Wier, A., Cianfrocco, M.A., Chong, J., Hare, A.A., Dervan, P.B.,
1106 DiMaio, F., Leschziner, A.E.*, et al.* (2017). Structural basis for the initiation of eukaryotic
1107 transcription-coupled DNA repair. Nature *551*, 653-657.
1108

1109

1110 # Supplemental Information

1111

1112 **Mechanism of transcription modulation by the transcription-repair coupling factor**

1113 Bishnu Paudel[1,2], Zhi-Qiang Xu[1,2], Slobodan Jergic[1,2], Aaron J Oakley[1,2], Nischal Sharma[1,2],

1114 Simon HJ Brown[1,2], James C Bouwer[1,2], Peter J Lewis[3], Nicholas E Dixon[1,2], Antoine M van

1115 Oijen[1,2], Harshad Ghodke[1,2]*

1116 [1]Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong,

1117 Wollongong, Australia

1118 [2] Illawarra Health and Medical Research Institute, Wollongong, Australia

1119 [3]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia

1120 **Correspondence:**

1121 Harshad Ghodke, Molecular Horizons and School of Chemistry and Molecular Bioscience, University

1122 of Wollongong, Wollongong, New South Wales, 2522, Australia

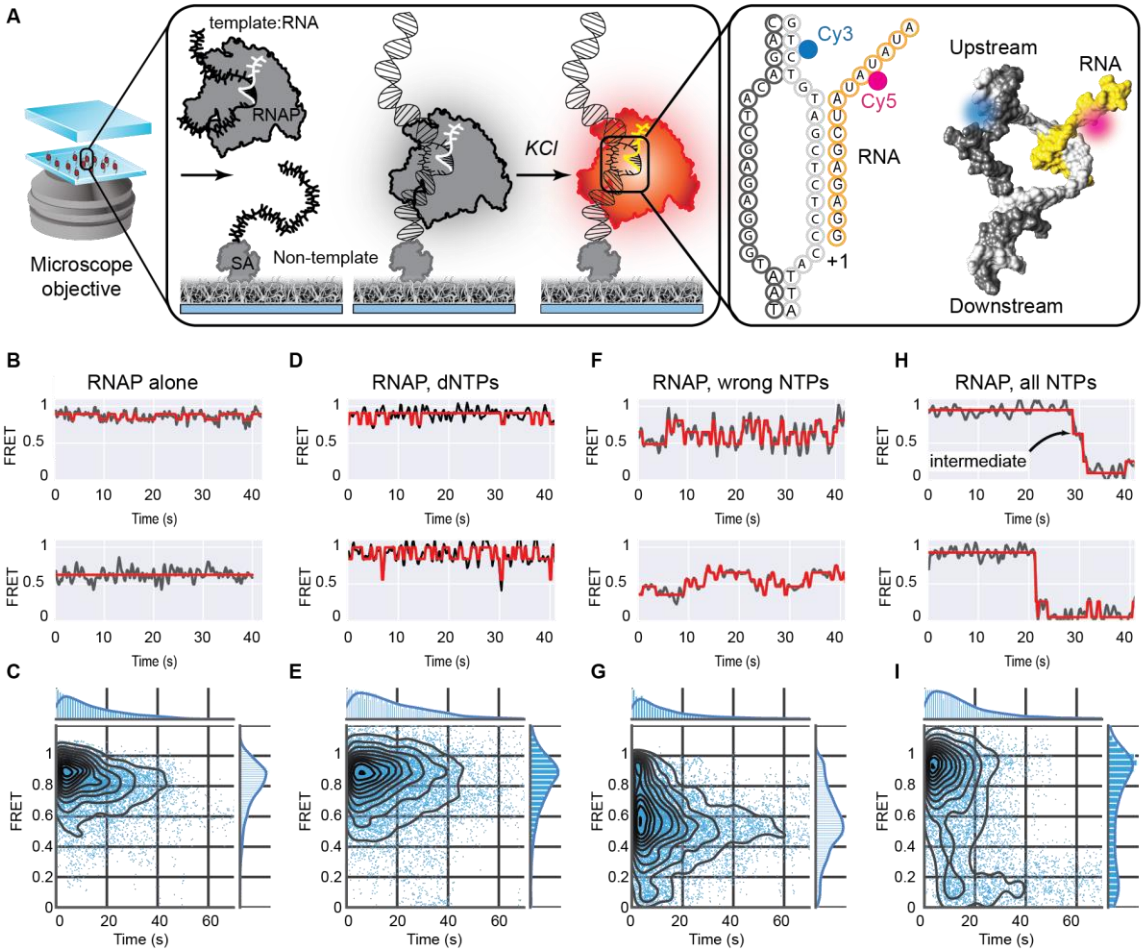
1123

1124 **Supplemental Note 1**

1125 SPR measurements of the strength of interaction between Mfd and this DNA substrate in the presence
1126 of ATPγS revealed a dissociation constant that is somewhat higher (K_D = 300 ± 10 nM) than that of
1127 dsDNA (170 ± 20 nM; Supplemental Figure 4B–D compared to Supplemental Figure 2B). To monitor
1128 changes in the distance between the junction and the 5' end, we introduced a FRET pair on either end
1129 of the ssDNA overhang (Figure 4A). SPR measurements revealed a binding affinity of 390 ± 10 nM for
1130 the binding of Mfd to the FRET pair labelled substrate (Supplemental Figure 4D).

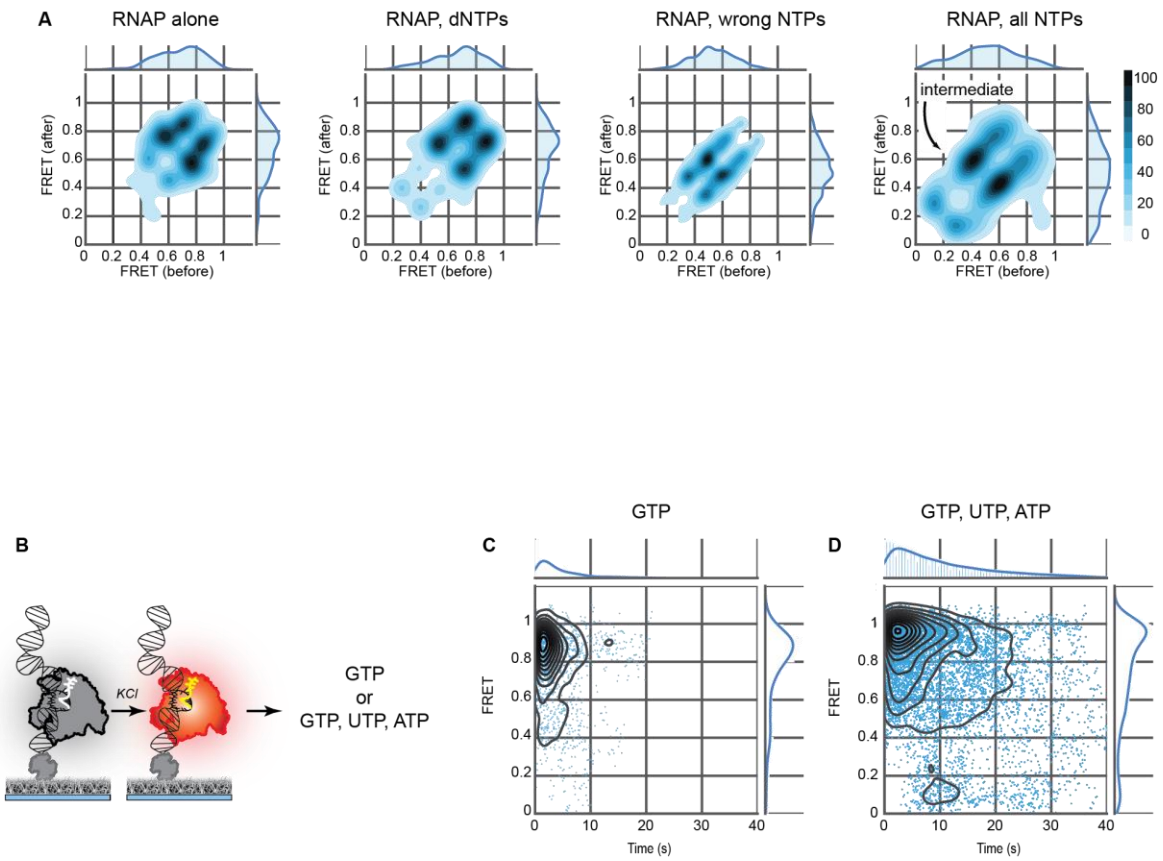
1131

Figure 1



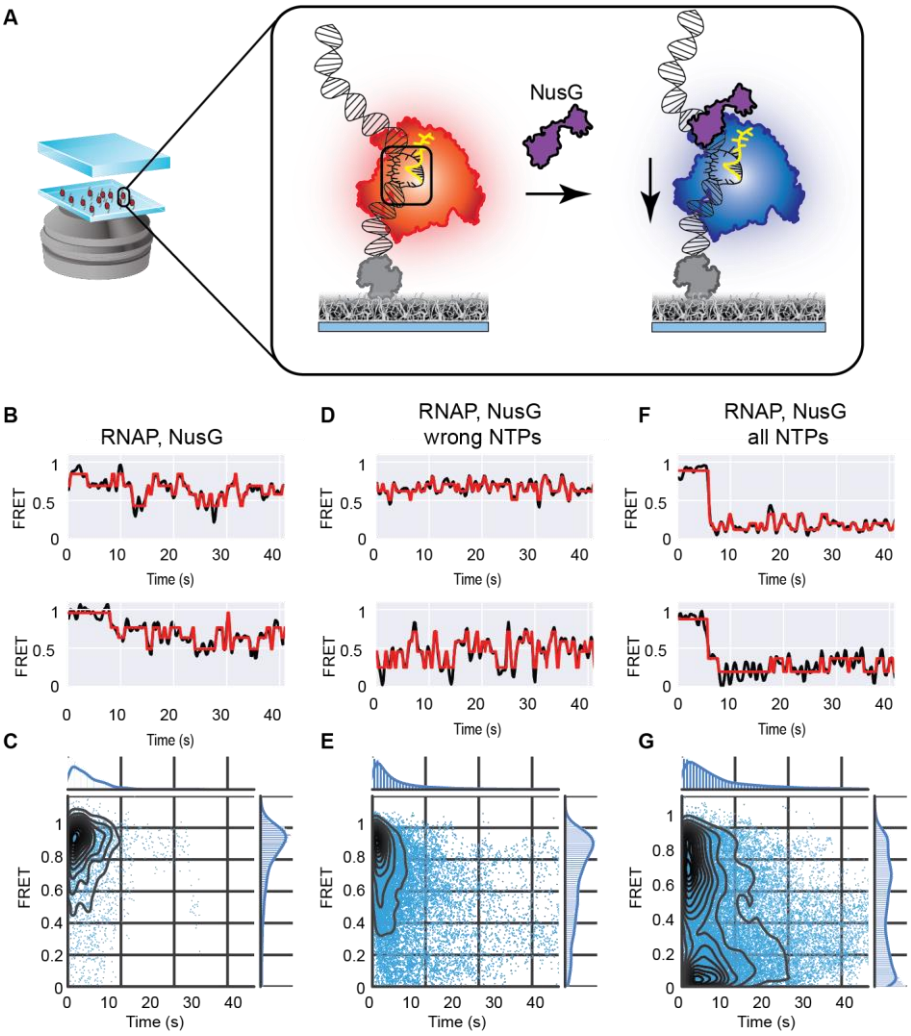
1132

Supplemental Figure 1



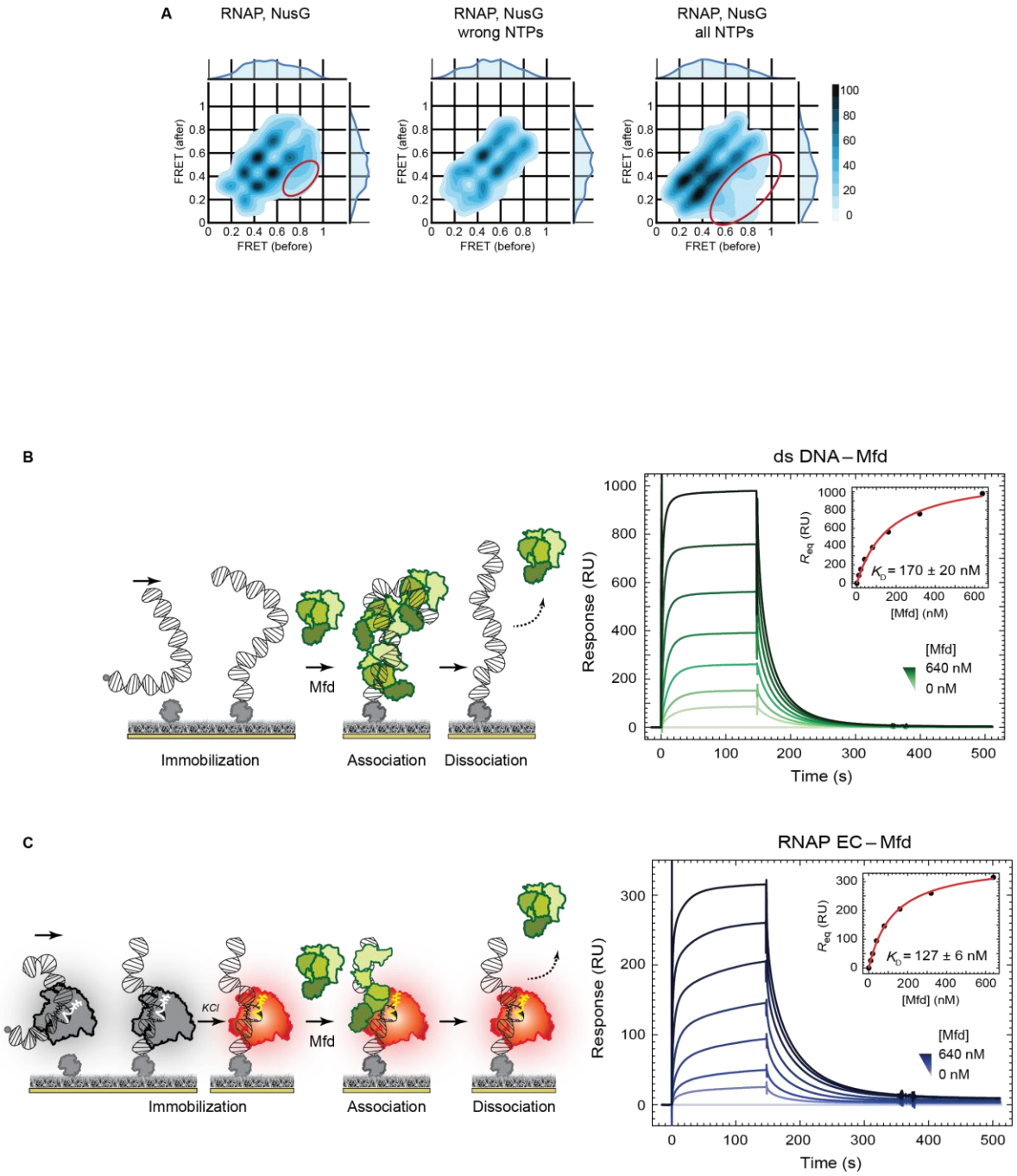
1133

Figure 2



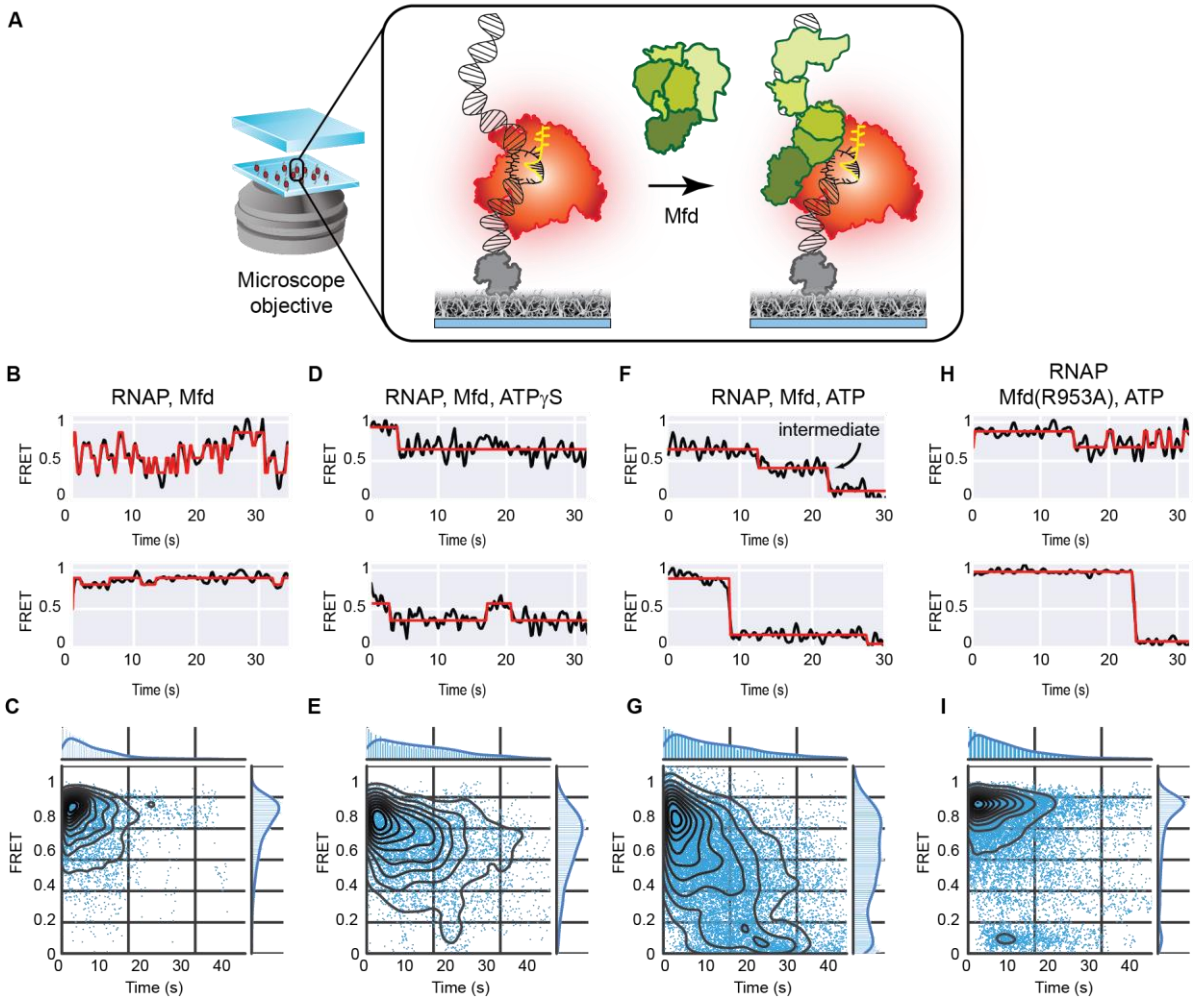
1134

Supplemental Figure 2



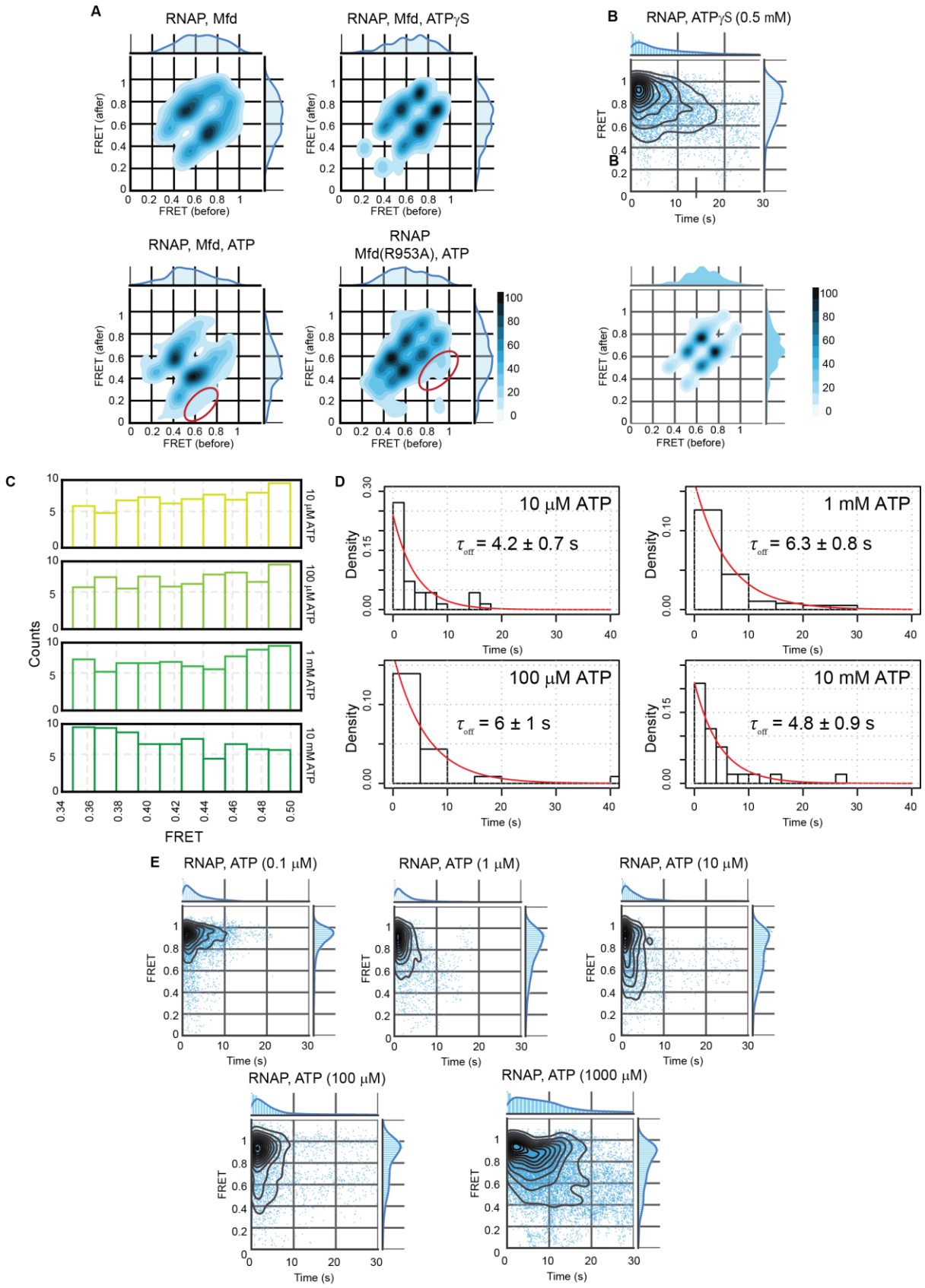
1135

Figure 3



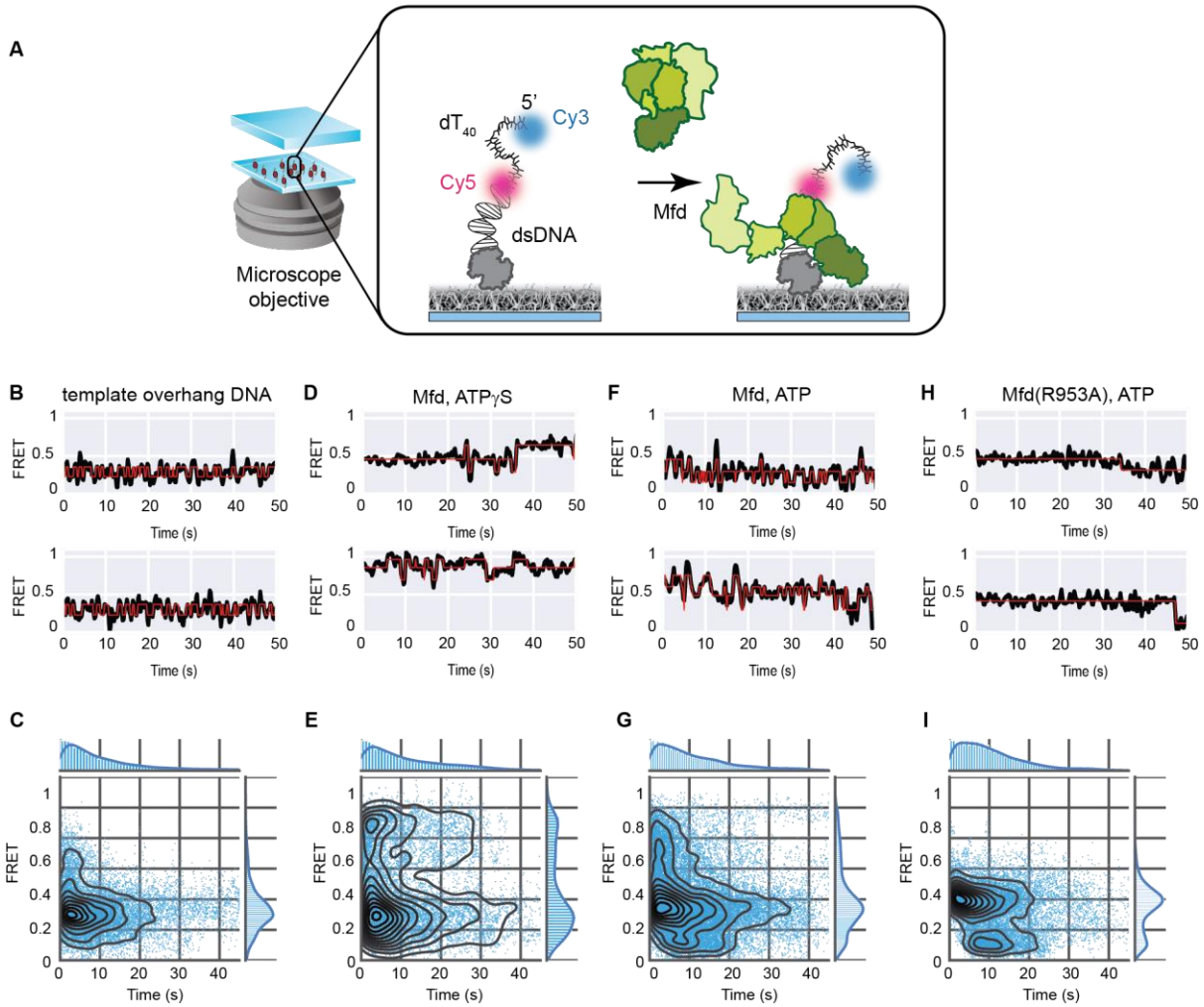
1136

Supplemental Figure 3



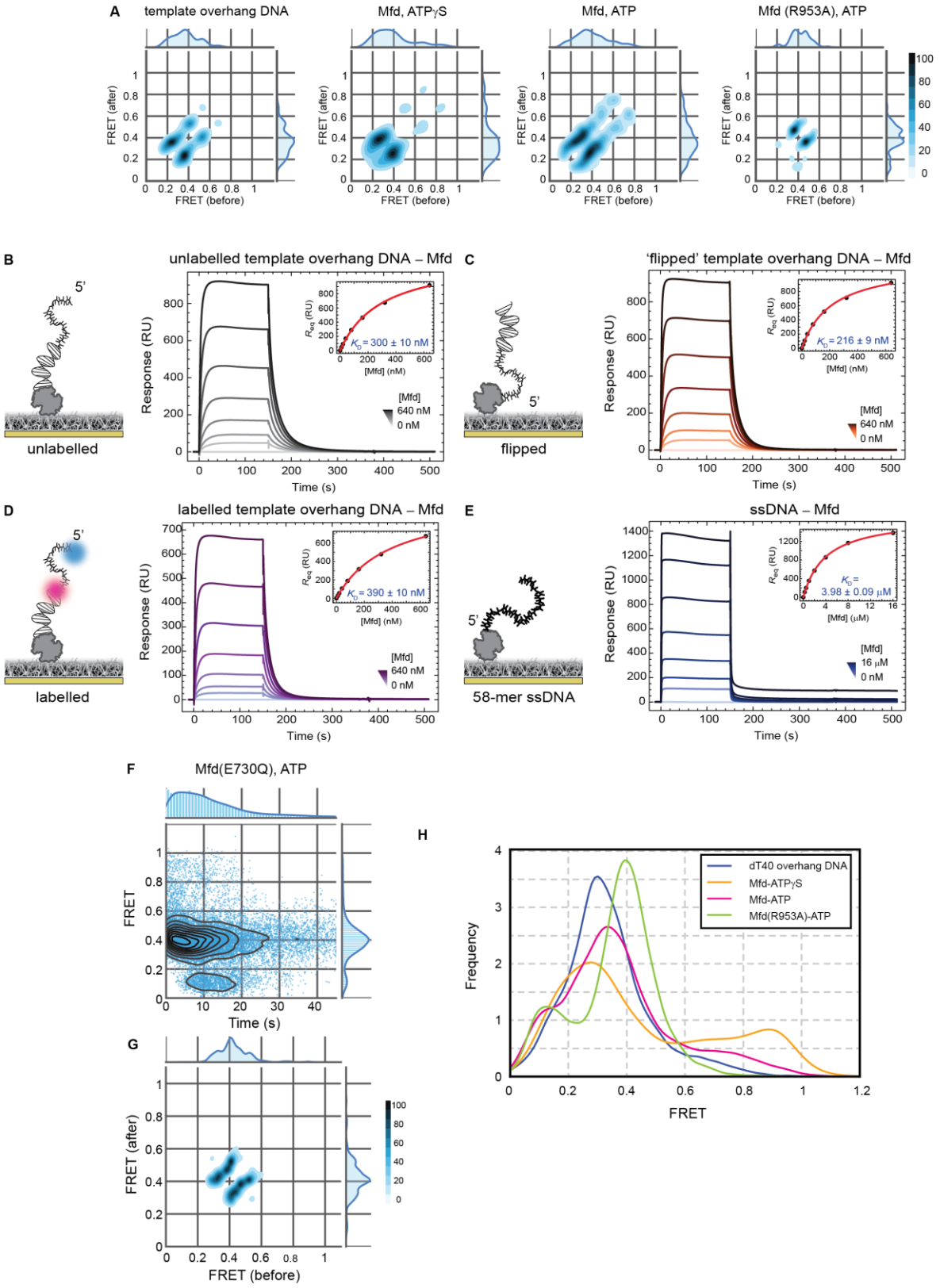
1137

Figure 4



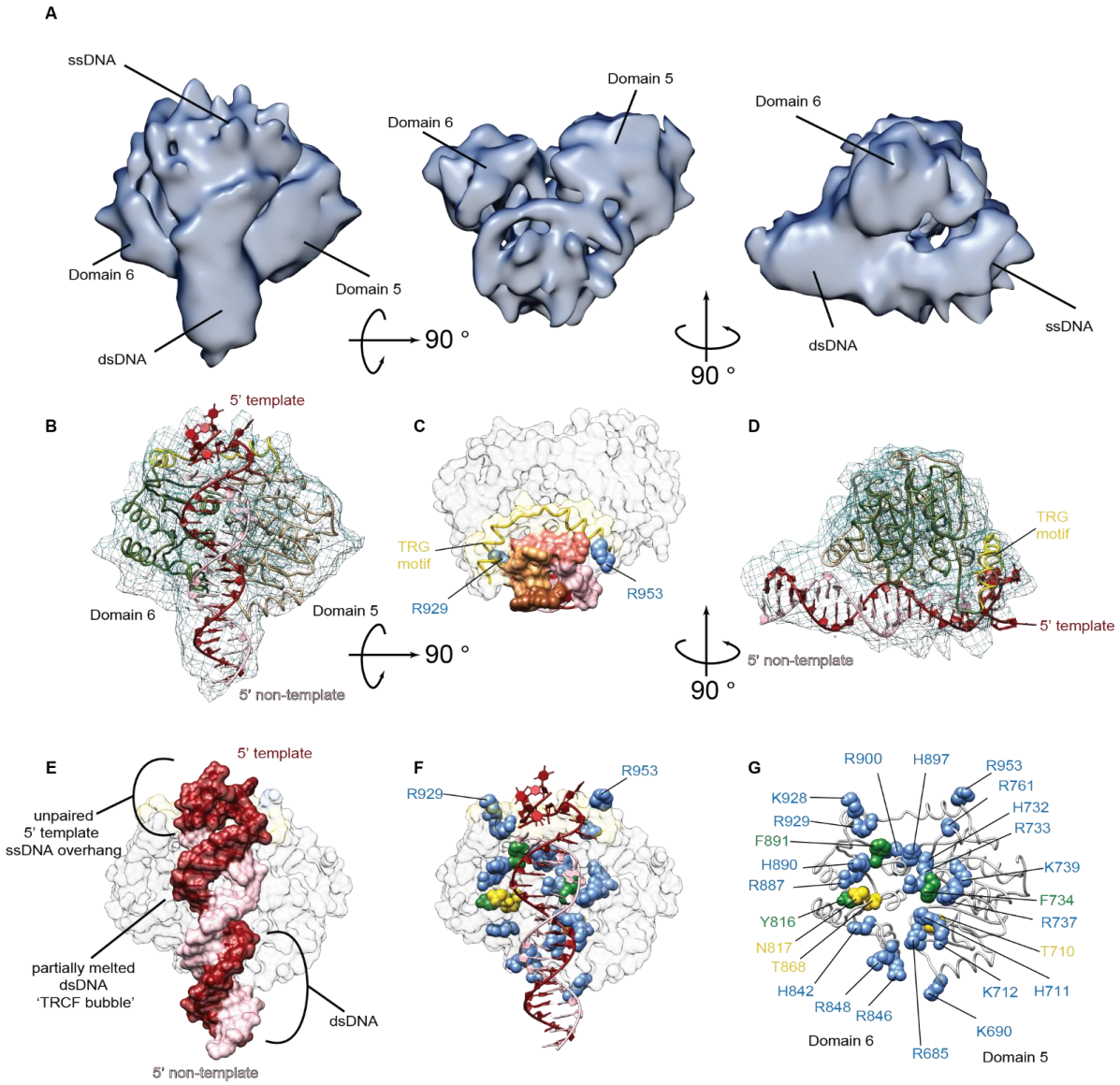
1138

Supplemental Figure 4



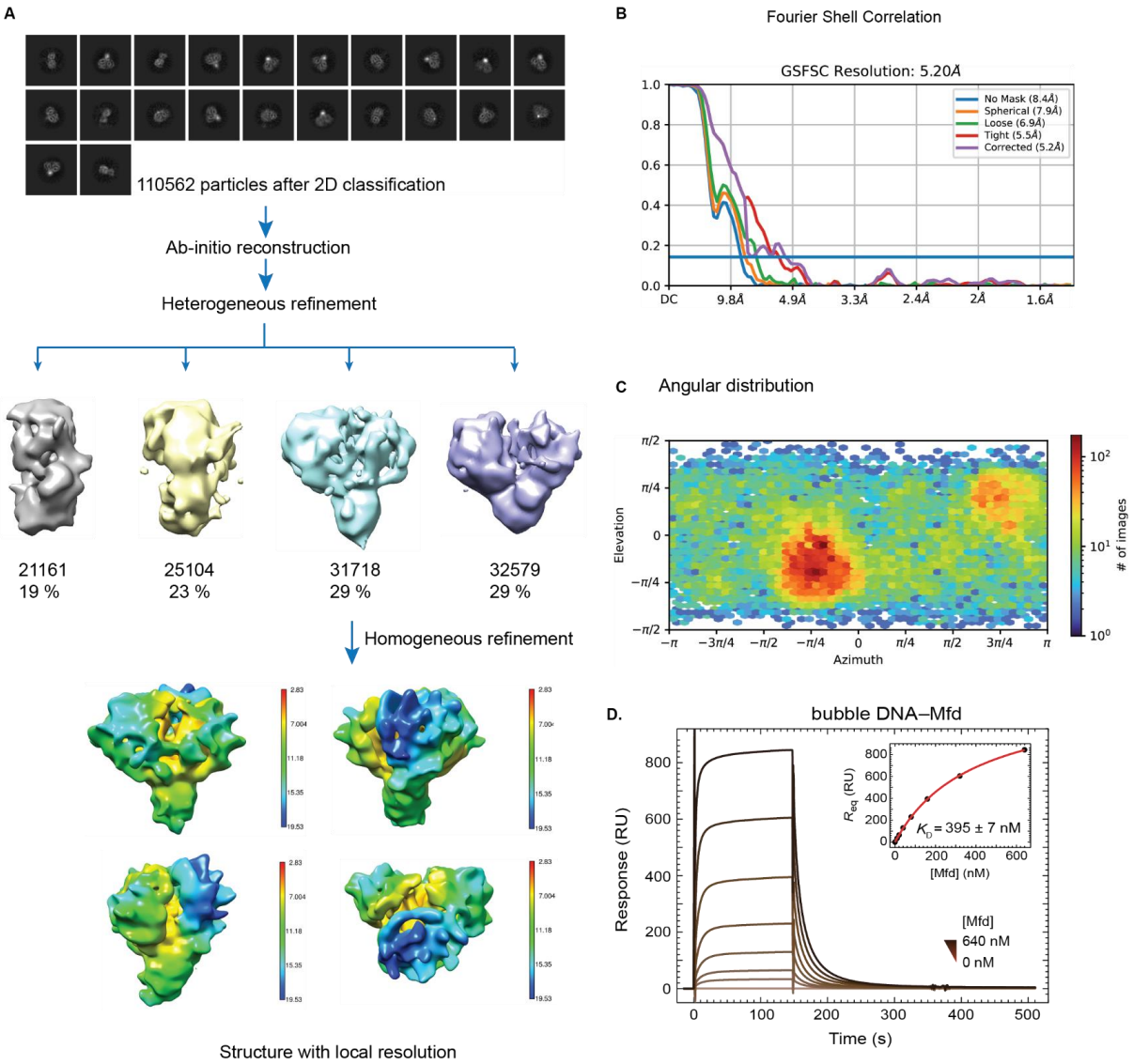
1139

Figure 5



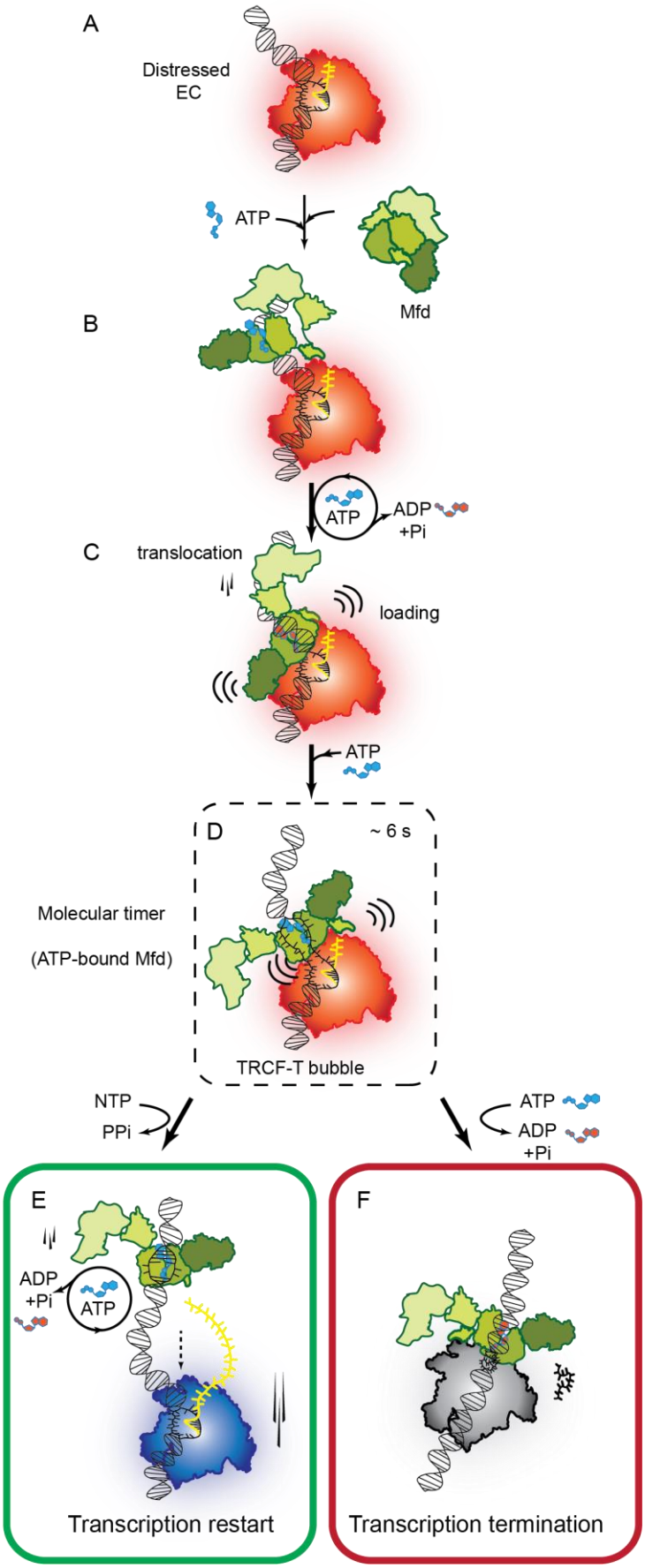
1140

Supplemental Figure 5



1141

Figure 6



1142