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[1,3], 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 dynamically 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 poised 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, biophysical studies 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.

27 **Introduction**

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

35 Elongation is punctuated by sequence-dependent pausing, stalling at sites of damage on the
36 DNA template or at sites of roadblocks in the form of DNA-bound proteins. Prolonged residence in
37 paused states can lead to escape into backtracked states in which the catalytic site of RNAP loses
38 register with the 3' end of the RNA transcript, allowing the newly synthesized transcript to invade the
39 secondary channel of RNAP (Abdelkareem et al., 2019; Gabizon et al., 2018; Komissarova and Kashlev,
40 1997). Rescue from such backtracked states is catalyzed by processivity factors such as GreA and GreB
41 (Abdelkareem et al., 2019; Borukhov et al., 1993). Elongation is modulated by the universally
42 conserved elongation/pause factor, the bacterial NusG, homologous with Spt5 in eukaryotes
43 (reviewed by (Wang and Artsimovitch, 2020)). In *Escherichia coli*, NusG maintains processive
44 transcription (Burova et al., 1995) by suppressing backtracking through contacts on the upstream DNA
45 (Herbert et al., 2010; Kang et al., 2018).

46 Elongation is also modulated by the transcription-repair coupling factor (TRCF) resulting in
47 either transcription termination or rescue. The bacterial TRCF – Mfd is a multi-domain adaptor protein
48 consisting of 7 domains connected by flexible linkers (Supplemental Figure 1A) (Deaconescu et al.,
49 2006). Domains 1a/2/1b form the UvrB–homology module that bears structural homology to the
50 nucleotide excision repair protein UvrB and enables contacts with the repair protein UvrA. Domain 3
51 serves as a structural element playing a role in orienting the UvrB-homology module during the

52 activation of the protein (Brugger et al., 2020), following interactions of domain 4, the RNAP
53 interaction domain, with distressed RNAP (Kang et al., 2021; Park et al., 2002). Domains 5 and 6
54 together form the motor domain, representing a (helicase) superfamily II ATPase that exhibits
55 structural homology to the RecG translocase (Deaconescu et al., 2006; Mahdi et al., 2003). Domain 7
56 is an auto-inhibitory domain that contacts the UvrB-homology module to maintain it in a closed
57 conformation, suppressing promiscuous DNA binding of Mfd in the absence of interactions with RNAP.

58 A detailed understanding of the activities of Mfd has been assembled through a combination
59 of molecular biology, biochemical (Chambers et al., 2003; Haines et al., 2014; Manelyte et al., 2010;
60 Park et al., 2002; Selby and Sancar, 1993a, b; Smith et al., 2012; Smith et al., 2007), structural methods
61 (Brugger et al., 2020; Deaconescu et al., 2006; Kang et al., 2021; Shi et al., 2020), as well as single-
62 molecule (Fan et al., 2016; Graves et al., 2015; Howan et al., 2012; Le et al., 2017) and live-cell imaging
63 techniques (Ghodke et al., 2020; Ho et al., 2018, 2020). Mfd acts in a complex role in modulating
64 transcription activity, including at hard-to-transcribe regions (Ragheb et al., 2021). At sites of DNA
65 damage, Mfd orchestrates premature transcription termination leading to an enhancement in repair
66 of DNA damage in the template strand – a phenomenon known as transcription-coupled DNA repair
67 (Fan et al., 2016; Ghodke et al., 2020; Haines et al., 2014; Ho et al., 2020; Mellon and Hanawalt, 1989;
68 Selby and Sancar, 1993a). In bacterial cells, the $UvrA_2B$ nucleotide excision repair machinery is
69 recruited to the site of distressed RNAPs by Mfd, resulting in the site-specific loading of the damage
70 verification enzyme UvrB (Ghodke et al., 2020; Ho et al., 2020). This ability to recruit DNA-repair
71 machinery at sites of lesion-stalled RNAP to promote ATP-dependent premature transcription
72 termination is a property that is shared by eukaryotic functional homologs of Mfd – the yeast Rad26
73 (Duan et al., 2020; van Gool et al., 1994; Xu et al., 2017) and human CSB (van der Weegen et al., 2020;
74 Venema et al., 1990). Mutations in CSB protein result in deficient repair machinery recruitment
75 causing Cockayne syndrome in humans (Mayne and Lehmann, 1982; Troelstra et al., 1992; Venema et
76 al., 1990).

77 In addition to initiating transcription termination, the three best-studied TRCFs also serve as

78 elongation factors. Like Rad26 (Xu et al., 2017) and CSB (Selby and Sancar, 1997), Mfd has also been
79 shown to rescue transcription at pause sites *in vitro* through a 'release-and-catch up' mechanism (Le
80 et al., 2017; Park et al., 2002). Forward translocation of the elongation complex (EC) is central to TRCF
81 activities including rescue from backtracked states and transcription termination (Le et al., 2017; Park
82 et al., 2002; Selby and Sancar, 1997; Xu et al., 2017). Considering their membership in the class of
83 superfamily 2 helicases (Byrd and Raney, 2012), the structural homology of the TRCF motor domains
84 and a similar mode of engagement with the EC between Mfd (Brugger et al., 2020; Kang et al., 2021;
85 Shi et al., 2020) and Rad26 (Xu et al., 2017), it is plausible that the molecular mechanisms underlying
86 these outcomes are conserved.

87 However, precisely how TRCF activity results in transcription elongation versus rescue is
88 unknown. Two models have been proposed to explain forward translocation of RNAP by TRCFs: first,
89 the TRCF 'pushes' RNAP forward while translocating on dsDNA as proposed for Mfd (Park et al., 2002),
90 and second, the TRCF pulls the ssDNA template to effectively translocate the polymerase forward (Xu
91 et al., 2017). Recent structures of *Ec*Mfd approaching an *in vitro* reconstituted EC assembled on a non-
92 complementary bubble have demonstrated that following initial recognition of the EC, Mfd
93 translocates forward until it bumps into the RNAP, pushing the polymerase and resulting in global
94 remodeling of the EC (Kang et al., 2021). The structure of yeast Rad26 bound to upstream DNA in the
95 RNA Pol II EC (Xu et al., 2017) reveals direct contacts between dsDNA-bound Rad26 and the ssDNA
96 template strand. Based on the homology to the Swi2/Snf2-family core ATPase domains, and the
97 mechanism of Snf2 in chromatin remodeling, it has been proposed that Rad26 pulls template DNA in
98 the Pol II EC (Xu et al., 2017). Structures of DNA-bound Mfd have established the homology of its
99 motor domains to ATPases involved in chromatin remodeling (most notably Snf2); however it remains
100 unclear whether the proposed mechanism for Rad26 is conserved (Brugger et al., 2020). Since
101 available structures depict initial global remodeling of the EC but not transcription termination leading
102 to disassembly of the EC accompanied by loss of the transcript (Kang et al., 2021), it remains to be
103 seen whether a similar Mfd-ssDNA interaction underlies transcription termination. Further, while it is

104 clear that displacement of the polymerase in the absence of downstream binding partners of Mfd is a
105 slow event requiring several minutes *in vitro* (Graves et al., 2015; Howan et al., 2012), the sequence
106 of events leading to loss of the transcript and its residence time in the EC during this remodeling
107 remains unclear.

108 Here, we set out to understand how Mfd dynamically remodels the EC by monitoring the fate
109 of the RNA transcript under various scenarios. We developed a single-molecule Förster resonance
110 energy transfer (smFRET)-based assay that directly visualizes dynamic changes in the distance
111 between donor-labeled upstream DNA and acceptor-labeled RNA transcript forming the transcription
112 bubble. We reconstituted catalytically-functional ECs *in situ* and monitored the FRET signal from
113 surface-immobilized molecules under conditions of partial or complete nucleotide starvation, and
114 probed global remodeling of the EC by NusG and Mfd on the timescale of minutes. NusG efficiently
115 rescued ECs from catalytically inactive states, consistent with its role as an elongation factor. On the
116 other hand, Mfd was able to rescue ECs from catalytically inactive states into a long-lived catalytically
117 poised state, prior to continued ATP-hydrolysis-dependent forward translocation consistent with
118 transcription termination. This catalytically poised state acts as a 'molecular timer', presenting the EC
119 with an opportunity to restart transcription. Further biophysical and structural biology studies reveal
120 a novel role for a conserved arginine residue as a "hook" in mediating contacts with ssDNA while the
121 nucleotide-bound motor domain melts the dsDNA on DNA substrates mimicking the transcription
122 bubble. Collectively, the studies enable us to propose a 'melt, hook and pull' model to rationalize how
123 ATP binding and hydrolysis by Mfd modulate transcription elongation or termination.

124 **Results**

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

126 We monitored transcription elongation by assembling functional ECs on dye-labeled, fully-
127 complementary transcription bubbles *in situ* by adopting a previously described strategy to study RNA
128 Pol II dynamics (Andrecka et al., 2008) in combination with well-established approaches to
129 reconstitute functional *E. coli* ECs (Komissarova et al., 2003). Specifically, we introduced a Cy3 donor

130 dye on the upstream DNA and a Cy5 acceptor on the RNA (Figure 1A; Supplemental Figure 1B) at sites
131 that are predicted to exhibit high smFRET signals within the dynamic range of the technique
132 (Supplemental Note 1). Three features of this strategy enable precise dissection of the dynamics of
133 molecular mechanisms of Mfd-dependent remodeling of the EC: (i) the use of synthetic RNA and DNA
134 oligonucleotides enables site-specific introduction of dyes in the phosphodiester backbone that serve
135 as FRET pairs, and consequently, the ability to probe conformational changes between various parts
136 of this complex; (ii) smFRET affords sub-second temporal resolution (200 ms in our experiments)
137 enabling reliable measurements of dynamic interactions on the timescale of seconds to minutes, and
138 (iii) the ability to specify RNA sequences that cannot hybridize with upstream DNA limits extensive
139 backtracking of ECs. RNAP in the elongation phase can exist as a backtracked complex (where the 3'
140 end of the transcript is extruded into the secondary channel (Abdelkareem et al., 2019)), a pre-
141 translocated conformation prior to nucleotide addition, and a post-translocated conformation
142 following nucleotide incorporation. Since backtracking requires hybridization of the RNA transcript
143 with the upstream DNA, the choice of a non-complementary RNA sequence is expected to minimize
144 the heterogeneity in the observed population of ECs in this assay.

145 **Monitoring transcription elongation by RNAP**

146 Functional RNAP ECs were assembled *in situ* on the microscope stage at room temperature
147 (24°C) by flowing a solution containing RNAP that had been pre-incubated with the template:RNA
148 hybrid into a microfluidic device with surface immobilized biotinylated non-template DNA. After
149 complex assembly, the flow cell was washed with a high-ionic strength solution (1M KCl) followed by
150 rapid exchange with a buffer solution supporting transcription but without ribonucleotides (NTPs).
151 These steps enabled the retention of only completely reconstituted elongation complexes that each
152 contain a Cy3-labeled template and a Cy5-labeled transcript (see Table 1 for sequences). Subsequent
153 measurements of the smFRET signal between the Cy3 donor and the Cy5 acceptor (see Methods)
154 reveal that complexes assembled resided in stable high FRET states centered at 0.9 (Figure 1B top
155 panel, upper trace; Supplemental Figure 1C for example intensity traces). In addition, a small fraction

156 (14% of 318 trajectories) also exhibited mid-FRET states centered at 0.6 (Figure 1B top panel, lower
157 trace).

158 These states were recapitulated in temporal heat maps (Figure 1B, middle panel) that are
159 scatterplots describing how the smFRET signal evolves in time for the ensemble population. The heat
160 maps represent the raw data, and the projections on the Y-axis represent the ensemble FRET
161 measurements (of all FRET trajectories), while the projection on the X-axis displays the evolution of
162 all FRET signals in time measured with a time resolution of 0.2 s. The loss of signal over time is
163 attributable to the photobleaching of individual dyes during the course of the acquisition due to
164 continuous laser illumination (Figure 1B, middle panel). Here, the overlaid black lines reflect contours
165 parsing the data into deciles. These temporal heat maps are thus a convenient representation
166 presenting an overview of the progress of FRET trajectories and population density over time.

167 Since the majority of the complexes thus assembled exhibited time-FRET trajectories that
168 fluctuate around 0.9 FRET, and a minority exhibited transitions to the 0.6 FRET state, we pursued an
169 unbiased approach to visualize transitions between various FRET states. FRET trajectories were
170 idealized using vbFRET (http://vbfret.sourceforge.net/) and the detected states were plotted in
171 transition density plots (TDPs; Supplemental Figure 1D, 'EC alone') to visualize interconversion among
172 different states and the directionality of those changes (Bronson et al., 2009). In this representation,
173 features that are only present in one half of the plot on either side of the diagonal manifest as
174 asymmetric features and denote irreversible transitions between states on the timescale of the
175 observation. As demonstrated by the TDP, the interconversions between the states were reversible
176 (based on the displayed symmetry of the FRET populations across the diagonal), and were restricted
177 to the high FRET regime. Within the paradigm that FRET changes reflect conformational changes, a
178 straightforward interpretation of these data is that the complex undergoes relatively minor, but
179 reversible, conformational changes.

180 For convenience, here we assign smFRET signals between 0.7–1 as being 'high', 0.4–0.7 as
181 'mid' and 0–0.4 as being 'low'. Trajectories were then sorted as 'static' (black bars; bottom panel

182 Figure 1B) if the signal resided in exactly one state, and dynamic otherwise. Further, dynamic
183 trajectories were parsed into 'high', 'mid' or 'low' categories if all the states visited by the molecule
184 were within the ranges defined above. Molecules exhibiting transitions between two adjacent regimes
185 were classified as 'high–mid' or 'mid–low' as appropriate, and molecules exhibiting transitions among
186 all three regimes were classified in the 'high–low' class. Here, the lower panels represent the
187 percentages of the trajectories in each class, and the total number of molecules observed is indicated
188 in the top right in each panel. Here, for 'EC alone' 65% of the ECs exhibited static FRET signals. 10%
189 exhibited dynamics confined to the high-FRET regime, and 14% and 9% of observations respectively
190 exhibited transitions in the high–mid and mid–low FRET regimes. This visualization serves to provide
191 an overview of static and dynamic behavior by the individual molecules in the data set.

192 To assign states, and to test that the complexes are catalytically functional, we challenged
193 complexes with deoxyNTPs (dNTPs), 'wrong' (or non-complementary) NTPs and the complete set of
194 NTPs. In this assay, transcript elongation is expected to increase the distance between the FRET pair
195 dictated by the geometry of the EC, leading to a drop in FRET. Since dNTPs are selected against
196 incorporation into the transcript (Svetlov et al., 2004), we expected that smFRET signal from the
197 reconstituted ECs would not drop. Consistent with this expectation, complexes exposed to dNTPs
198 revealed a high-FRET population (Figure 1C). In contrast, non-complementary NTPs converted the
199 population of complexes in the high-FRET states to the mid-FRET states (Figure 1D). In both cases, the
200 interconverting populations exhibited symmetric peaks across the diagonal indicating that in the
201 presence of either dNTPs or the wrong NTPs, the ternary complex visits the observed states reversibly
202 (Supplemental Figure 1D).

203 'Walking' the complex by supplying a subset of the 'correct' complementary NTPs revealed
204 gradual conversion from the high-FRET to a mid-FRET population (Supplemental Figure 1E–G). When
205 presented with the complete set of NTPs, the ternary complex exhibited unidirectional conversion of
206 the FRET signal from high- to low-FRET values (Figure 1E and Supplemental Figure 1D). Notably, the
207 TDPs, and indeed the individual traces, often revealed a conversion of the high-FRET states to the low-

208 FRET states *via* the mid-FRET states. Of the 30% of trajectories exhibiting dynamics from the high to
209 the low FRET regime, only 30% of the ECs exhibited a single direct transition, 57% exhibited a step-
210 wise transition to the low-FRET regime by visiting the mid-FRET regime (intermediate; Supplemental
211 Figure 1D, bottom right panel) and the remainder exhibited high-mid dynamic trajectories prior to a
212 single-step transition from a high-FRET state to a low-FRET state. It is possible that trajectories
213 exhibiting a single-step transition conceal a hidden intermediate step with a lifetime below our time
214 resolution.

215 To better characterize the mid-FRET states observed in these experiments, we plotted
216 distributions of lifetimes of individual states in each condition (Table 2, Supplemental Figure 1H and
217 Supplemental Note 2). Here, the elongation complex exhibited a lifetime 5.0 ± 0.8 s, and this did not
218 change appreciably in the presence of dNTPs (5.7 ± 1.5 s) and correct NTPs (5.5 ± 1.3 s) although, a
219 slightly longer lifetime of 7.8 ± 0.9 s was detected in the presence of the wrong NTPs. In conjunction
220 with the TDPs, we infer that despite comparable lifetimes, the exit from the mid-FRET states is favored
221 toward entry into high-FRET states for all conditions except in the presence of the correct NTPs.

222 Collectively, the data indicate that the complexes assembled *in situ* represent functional
223 ternary ECs that exist in a catalytically 'inactive' state and a catalytically 'poised' state. The
224 observations that complexes often first transition to the mid-FRET state prior to addition of NTPs
225 suggests that this represents a catalytically poised state. We tentatively assigned the high-FRET states
226 centered at 0.9 to a catalytically inactive state. This catalytically inactive state may potentially
227 represent a pre-translocated conformation of the EC. Estimates of the expected FRET efficiency based
228 on the distances between the Cy3-Cy5 dye pair for the pre- and post-translocated conformations of
229 the EC are in good agreement with the measured FRET efficiencies measured in our experiments
230 (Supplemental Figure 1B and Supplemental Note 1).

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

232 Next, we tested this assignment of states by assessing the response of the complex to the elongation
233 factor *Ec*NusG (Figure 2A), which is predicted to retain ECs in the catalytically poised state. Addition

234 of NusG indeed resulted in an increased occupancy of mid-FRET states accompanied by a reduction of
235 the high-FRET state (Figures 2B *cf*. 1B). Trajectories revealed a conversion from the high-FRET state
236 into the mid-FRET state and fluctuation around the mid-FRET state (Figure 2B), consistent with our
237 assignment of the mid-FRET as the catalytically poised state.

238 Incubation with incorrect NTPs in the presence of NusG revealed a greater population of
239 dynamic ECs with dynamic behavior across all FRET regimes (Figure 2C and Supplemental Figure 2A).
240 Reversible short-lived transitions to low-FRET states were also evident (Figure 2C and Supplemental
241 Figure 2A). The lifetime of this mid-FRET state (6.9 ± 0.9 s) was similar to that of the EC alone incubated
242 with the incorrect NTPs (7.8 ± 0.9 s) while being marginally longer than that of the EC incubated with
243 NusG alone (4.2 ± 0.4 s) in the absence of any nucleotide (Supplemental Figure 2B).

244 Importantly, the complete set of NTPs revealed efficient, irreversible conversion to the low-
245 FRET regime (0 – 0.4) from the mid-FRET regime (Figure 2D and Supplemental Figure 2A) in the
246 presence of NusG suggesting processive extension of the transcript through nucleotide incorporation
247 and associated forward translocation by RNAP. The lifetime of the EC in the presence of NusG and all
248 correct nucleotides was comparable to the lifetime of this state in the absence of any nucleotide or in
249 the presence of correct nucleotides (Table 2, Supplemental Figures 2B, 1H and Supplemental Note 2).

250 Overall, NusG led to a reduction in the number of static molecules, accompanied by a greater
251 fraction occupying states in the mid-FRET regime in the presence of correct NTPs. Together, these
252 observations are consistent with the role of *Ec*NusG as a suppressor of backtracking of RNAP (Burova
253 et al., 1995; Herbert et al., 2010; Kang et al., 2018). Overall, these results indicate that *Ec*NusG
254 modulates the equilibrium between the catalytically inactive and active states by promoting residence
255 of the complex in a catalytically poised conformation.

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

257 Next, we assessed the interactions of Mfd with ECs armed with the assignment of states in the high-
258 FRET regime as catalytically inactive states, and those in the mid-FRET states as a catalytically poised
259 states. To characterize the binding of Mfd to RNAP, we first measured the affinity of Mfd for pre-

260 assembled ECs or the dsDNA alone using surface plasmon resonance (SPR). Since Mfd interacts poorly
261 with DNA in bulk assays in the absence or presence of ATP, measurements of binding of purified Mfd
262 to either DNA or the EC were carried out in the presence of slowly hydrolyzable ATPγS. SPR
263 measurements revealed a dissociation constant (K_D) of 170 ± 20 nM (Supplemental Figure 2C) for 49-
264 mer dsDNA alone, consistent with measurements by other groups (Brugger et al., 2020; Le et al.,
265 2017). Binding to reconstituted RNAP ECs on a 49-mer exhibited similar affinity with a K_D of 127 ± 6
266 nM (Supplemental Figure 2C). These data are consistent with the model that dsDNA serves as a major
267 binding partner for Mfd-ATPγS in the context of the EC as previously demonstrated (Kang et al., 2021;
268 Le et al., 2017; Park et al., 2002; Shi et al., 2020). It is notable that whereas association curves of Mfd
269 to dsDNA exhibit fast saturation, binding of Mfd to RNAP ECs does not reach full saturation on the
270 timescale of the association phase. This may be attributable to the presence of two types of binding
271 substrates, for example, non-specific binding to ds DNA alone, or specific binding to the EC.

272 We then investigated how interactions of Mfd with RNAP influence the conformational
273 dynamics of the transcription bubble using our smFRET assay (Figure 3A). Consistent with the lack of
274 stable binding of Mfd with DNA in the absence of ATP, only 49% of ECs exhibited observable
275 interactions in the smFRET measurements (n_{obs} = 77, Figure 3B bottom panel). These trajectories
276 exhibited reversible transitions from the high- to mid-FRET states even in the absence of any
277 nucleotide co-factor (Figure 3B and Supplemental Figure 3A 'EC, Mfd'), reflecting failed attempts at
278 stable engagement of Mfd with the EC. Notably, differently from the EC alone (Supplemental Figure
279 1), dynamic trajectories exhibited transitions mainly in the high-mid and high-FRET regimes.

280 Mfd-ATPγS resulted in dynamic ECs visiting additional mid-FRET states (Figure 3C and
281 Supplemental Figure 3A 'EC, Mfd-ATPγS'). ATPγS alone did not account for these observations
282 (Supplemental Figure 3B). Curiously, Mfd-ATPγS exhibits apparently quantized transitions in the
283 transition density plot compared to Mfd alone, indicating that binding of Mfd-ATPγS displaces RNAP
284 forward (Supplemental Figure 3A 'EC, Mfd' *vs.* 'EC, Mfd-ATPγS'). Together the data reveal that Mfd
285 can isomerize RNAP merely upon binding to the nucleotide-starved EC. The efficiency of this

286 isomerization is appreciably greater in the presence of ATPγS.

287 Treatment of *in situ* assembled ECs with Mfd and ATP (1 mM) revealed a greater number of
288 molecules exhibiting dynamics across all three regimes, including molecules that exhibited efficient
289 conversion to the mid- and low-FRET regimes (Figure 3D). To further explore the nature of these
290 transitions, we visualized them in a transition density plot (Supplemental Figure 3A, 'EC, Mfd-ATP'),
291 which revealed a dynamic population in the mid-FRET regime. Measurement of the lifetime of Mfd-
292 ATP in the intermediate state revealed a lifetime of 6.2 ± 0.7 s, comparable to that of Mfd alone (7.1
293 ± 2 s) or in the presence of ATPγS (6.1 ± 1.4 s) (Table 2, Supplemental Figure 3C and Supplemental
294 Note 2). Strikingly, the lifetime of this mid-FRET state did not exhibit a dependence on the
295 concentration of ATP in the range from 10 μM to 10 mM suggesting that stoichiometry amounts of
296 ATP are sufficient to promote conversion to low-FRET states (Table 2 and Supplemental Figure 3C).
297 ATP alone was not responsible for the conversion of the EC to low-FRET states in the absence of Mfd
298 (Supplemental Figure 3D).

299 Compared to the wild-type, the Mfd(R953A) mutant (Chambers et al., 2003), known to be
300 proficient in ATP hydrolysis while strikingly deficient in RNAP displacement and transcription
301 termination exhibited inefficient conversion to the low-FRET states (Figure 3E). Notably, unlike wild-
302 type Mfd, transitions were observed mainly in the high- to mid- FRET regime (Supplemental Figure 3A
303 'EC, Mfd(R953A)-ATP'). The lifetime of Mfd(R953A) in the mid-FRET regime was comparable to that of
304 Mfd in the absence of nucleotide (Supplemental Figure 3C). Intriguingly, Mfd(R953A)-ATP exhibited a
305 pattern of quantized transitions that is similar to that of Mfd-ATPγS in the transition density plots
306 (Supplemental Figure 3A, compare 'EC, Mfd-ATPγS' and 'EC, Mfd(R953A)-ATP'), suggesting that the
307 mutation in this residue serves to reveal distinct aspects of Mfd's binding to the EC.

308 Together, the data indicate that both wild-type and translocase deficient Mfd rescue RNAP
309 from the catalytically inactive states accessible in our substrate to catalytically-poised states, albeit
310 with varying efficiencies. However, only the wild-type Mfd can further process the EC in the presence
311 of ATP. The low-FRET states observed in our system represent circumstances in which the distance

312 between the donor–acceptor pair is outside the dynamic range of efficient FRET as might be expected
313 when the RNAP is forward translocated well past the stall site, and/or the Cy5-labeled RNA transcript
314 is lost. The observation that the EC fails to return to the mid- or high-FRET states in the presence of
315 ATP indicates the occurrence of an irreversible reaction, most consistent with an irreversible loss of
316 RNA. In this model, Mfd(R953A) fails to displace the transcript from the EC, consistent with prior
317 findings (Chambers et al., 2003).

318 To confirm that the exit from the mid-FRET regime observed in the Mfd-EC in the presence of
319 ATP corresponds to loss of the transcript and not photobleaching of the Cy5 dye, we measured the
320 residence time of the Cy5-labeled RNA through direct excitation using a 637 nm laser (see Methods).
321 Here, the distribution of the lifetimes of the Cy5 signal in samples containing RNAP ECs alone on the
322 surface exhibited a single species with a lifetime of 11.3 ± 0.1 s corresponding to the photobleaching
323 lifetime of Cy5 due to direct excitation by the laser. On the other hand, incubation with Mfd in the
324 presence of ATP (1 mM) revealed a distribution that was best fit by two exponentials: a short-lived
325 species with a lifetime of 4.6 ± 0.2 s (amplitude: 52 % of the population) in addition to the longer-lived
326 species detected previously, corresponding to the photobleaching of Cy5 with a mean lifetime of 11.3
327 ± 0.1 s (Figure 3F). This short-lived species exhibits a lifetime that matches well the lifetime of the
328 catalytically poised intermediate measured using smFRET (4.6-6.2 s; Supplemental Figure 3). Further,
329 the fraction of dynamic trajectories observed in the FRET experiments (58%) agreed well with the
330 fraction of the population losing the RNA transcript. Thus, the data demonstrate that exit from this
331 intermediate indeed corresponds to loss of the RNA transcript. In essence this intermediate serves as
332 a 'molecular timer' counting down the time beyond which elongation is no longer an option.

333 Finally, we investigated whether nucleotide cofactor-bound Mfd was sufficient to chaperone
334 the EC into catalytically active conformations. To test whether ECs pre-incubated with Mfd-ATPγS for
335 10 minutes could further incorporate rNTPs, we introduced GTP and UTP into the flow cell and
336 measured the FRET signal. As expected, molecules exhibited a rapid conversion to low-FRET states
337 evidenced in individual trajectories, as well as in the ensemble heat map (Figure 3G). The characteristic

338 lifetime of the intermediate states was recovered from a gamma distribution to be 0.7 ± 0.2 s
339 (Supplemental Figure 3C), almost eight-fold lower than the mean lifetime of the intermediate state
340 for ECs incubated with NTPs (5.5 ± 0.9 s) recovered from a single-exponential distribution. A plausible
341 physical basis for this observation could be the restriction of backtracking following template slippage
342 by the physical presence of Mfd-ATPγS on the upstream edge of the transcription bubble.

343 To quantify the ability of ECs to incorporate nucleotides in the absence or presence of Mfd-
344 ATPγS, we compared histograms of the ensemble FRET states detected in the idealized trajectories for
345 the two conditions (Figure 3H). Fitting the histogram to a sum of Gaussians revealed three populations
346 centered at 0.91, 0.63 and 0.09 for the EC incubated with GTP, UTP and ATP. We assigned these as the
347 EC alone, catalytically poised EC and ECs that have extended the transcript following NTP
348 incorporation, respectively. Similarly, we recovered three populations centered at 0.93, 0.7 and 0.18
349 for ECs that were pre-incubated with Mfd-ATPγS followed by incubation with GTP and UTP. We
350 assigned these three populations as RNAP alone, ECs acted upon by Mfd-ATPγS and ECs that have
351 incorporated two NTPs.

352 In each case, the recovery of the population at 0.9 FRET indicates the presence of catalytically
353 inactive ECs. Further, the recovery of the two populations with mean FRET at 0.63 (95% confidence
354 intervals, lower bound (LB): 0.48, upper bound (UB): 0.77) and 0.7 (LB: 0.65, UB: 0.75) suggests that
355 the FRET signal of the catalytically poised state formed by the EC in the presence of NTPs is
356 indistinguishable from that of the intermediate formed by the EC in the presence of Mfd-ATPγS.
357 Importantly, a third population emerges in time (Figure 3G, Figure 1E and 3H) when NTPs are present.
358 This population exhibited a FRET state of 0.09 (LB: 0.05, UB: 0.12) when the EC was incubated with
359 GTP, UTP and ATP. In comparison, for ECs incubated with Mfd-ATPγS and GTP and UTP, this population
360 exhibited a FRET signal of 0.18 (LB: 0.15, UB: 0.2). The difference in the mean FRET values may reflect
361 the ability of ECs incubated with GTP, UTP and ATP to synthesize a longer transcript compared to ECs
362 incubated with only GTP and UTP.

363 Comparison of the fraction of the population in the low FRET states corresponding to

364 nucleotide incorporation by the RNAP EC or RNAP EC pre-incubated with Mfd-ATPγS (Figure 3H)
365 revealed a two-fold greater efficiency of incorporation after pre-incubation with Mfd-ATPγS: 35% of
366 all FRET states observed in the reaction corresponded to low FRET states when ECs were pre-incubated
367 with Mfd-ATPγS *vs* 15% in the absence of Mfd-ATPγS (Supplemental Note 3). Together the data enable
368 us to conclude that nucleotide-bound Mfd serves to promote the formation of a catalytically poised
369 state that is competent for transcription elongation when presented with NTPs.

370 **Mfd specifically recognizes sense ssDNA**

371 Structures of Mfd bound to dsDNA and ECs containing non-complementary transcription bubbles have
372 revealed a series of translocation intermediates that are formed prior to and leading up to stable
373 engagement of Mfd and initial disruption of the EC (Kang et al., 2021; Shi et al., 2020). Currently, little
374 is known about the intermediate states that follow this initial contact. We set out to identify whether
375 Mfd makes contacts with ssDNA in substrates mimicking the upstream edge of the transcription
376 bubble by employing a combination of approaches: First, we assessed whether Mfd exhibits a
377 preference for binding to sense ('S')/template or anti-sense ('AS')/non-template ssDNA overhangs at
378 junctions, second, we employed smFRET experiments to detect whether and how Mfd contacts
379 template ssDNA, and finally, we visualized Mfd bound to a primed DNA template using cryo-EM.

380 To investigate whether Mfd can recognize ssDNA in substrates mimicking the upstream edge
381 of the transcription bubble, we measured the DNA binding affinity of Mfd to 18-mer dsDNA substrates
382 containing either anti-sense or sense strand overhangs (Figure 4A, B; Table 1). First, we interrogated
383 whether length of the anti-sense strand overhang influenced the affinity of Mfd for the DNA substrate
384 (Figure 4A–C). We measured the dissociation constant of Mfd binding to dsDNA (18-mer) with varying
385 lengths of anti-sense strand overhangs (dT_x in which x = 0, 3, 6, 9, 12, or 15; Figure 4B). In the presence
386 of slowly-hydrolyzable ATPγS, the dissociation constant did not exhibit a dependence on the overhang
387 length (Figure 4C, green trace). Next, we measured binding affinity of Mfd-ATPγS to dsDNA substrates
388 (18-mer) with varying lengths of unpaired 5' sense strand overhangs (dT_x in which x = 0, 3, 6, 9, 12, or
389 15; Figure 4A–C). In this case, the binding curves revealed an inverse relationship between the length

390 of the ssDNA overhang and the strength of interaction: the longer the overhang, the weaker the
391 binding. K_D values increased from 90 ± 4 nM (for 18-mer ds DNA alone) to 220 ± 10 nM (for 18-mer
392 dsDNA with a 15 nt overhang) with a notable inflection point at 6 nt (Figure 4C). Finally, Mfd bound a
393 58-mer ssDNA with a K_D of ~4 μM (Supplemental Figure 4A) which is 20-fold weaker than binding to
394 dsDNA alone, consistent with prior estimates (Selby and Sancar, 1995).

395 In this assay, the R953A mutant similarly exhibited a loss of affinity as a function of sense
396 strand overhang length up to a dT6 overhang in the presence of ATPγS. The dissociation constant for
397 binding a substrate containing a 6 nt sense-strand overhang was 2.3-fold weaker than that for wildtype
398 Mfd. Overhang lengths larger than 6 nt were bound markedly more strongly, and approached the
399 binding affinity of Mfd for these substrates (Figure 4C). These data reveal that only contacts made by
400 Mfd on the template ssDNA overhang serve to antagonize its binding to dsDNA, establishing that the
401 non-hydrolyzable ATP analog-bound conformation of Mfd exhibits strand specificity at DNA forks.
402 Further, the results highlight a curious role of R953 in mediating these contacts on the sense strand
403 and a dependence on the length of the overhang.

404 To further understand these ATP-dependent contacts of Mfd with ssDNA we employed a
405 previously described smFRET assay to study translocase activity (Park et al., 2010). We used a DNA
406 substrate comprised of an 18-mer dsDNA duplex (mimicking upstream DNA) and a 5' dT_{40} sense
407 overhang mimicking the template DNA (Figure 4D) with a FRET pair positioned on either end of the
408 ssDNA overhang (Table 1). Mfd binds this substrate with a K_D of 390 ± 10 nM (see Supplemental
409 Information and Supplemental Figure 4B–D). In smFRET experiments, this substrate exhibited an
410 average FRET of 0.33 ± 0.15 (Figure 4E) in the absence of Mfd, reflecting rapid sampling of various
411 conformations accessed by the entropic coil formed by ssDNA (Figure 4E and Supplemental Figure 4E
412 'template overhang DNA').

413 In the presence of ATPγS, the smFRET signal from the Mfd-bound substrate revealed
414 additional distinct populations at higher FRET values of 0.6 and 0.8 (Figure 4F and Supplemental Figure
415 4E 'Mfd-ATPγS'). Transitions among these states were further enhanced in the presence of ATP as the

416 substrate exhibited a continuous spectrum of mid- and high-FRET states (Figure 4G and Supplemental
417 Figure 4E 'Mfd-ATP'), in addition to the distinct states detected in the presence of ATPγS. In this case,
418 the higher FRET states were only transiently accessed indicating that ATP hydrolysis plays a role in
419 exiting from these high–FRET states to low–FRET states. This observation is consistent with a picture
420 in which Mfd-ATP contacts the ssDNA to bring the dyes close to each other, resulting in a high FRET
421 signal, and subsequent ATP hydrolysis induces a lower–affinity conformation in which Mfd-ADP binds
422 the ssDNA overhang very weakly resulting in a return to the baseline FRET value of 0.33.

423 To verify that the dynamics observed in this assay reflect a dependence on ATP hydrolysis, we
424 studied binding to this substrate by the ATPase-deficient but DNA-binding competent E730Q mutant
425 of Mfd (Deaconescu et al., 2012; Ho et al., 2018); this Mfd(E730Q)-bound template exhibited high-
426 FRET states extremely rarely compared to the wild-type protein (Supplemental Figure 4F, 4G).
427 Therefore, the observation that Mfd(E730Q)-ATP is unable to stably engage with sense ssDNA, unlike
428 Mfd-ATPγS suggests that the Mfd(E730Q)-ATP and Mfd-ATPγS pre-hydrolytic states are sufficiently
429 different that only the wild-type Mfd retains the ability to engage ssDNA efficiently in the ATP-bound
430 state.

431 Unexpectedly, the R953A mutant did not exhibit high FRET states in the presence of ATP,
432 indicating that the ability to hydrolyze ATP alone is insufficient to engage ssDNA on primed DNA
433 templates (Figure 4H). Therefore, we conclude that both, the ability to hydrolyze bound ATP and the
434 presence of an intact R953 are essential for Mfd to engage the ssDNA template DNA overhang. We
435 propose that R953 serves to 'hook' sense ssDNA, and ATP binding serves to 'pull' it closer to the
436 junction resulting in the high FRET signal detected in this assay with wild-type Mfd. In turn, ATP
437 hydrolysis leads to exit from the high-FRET state reflecting slippage of the ssDNA from the motor
438 domain. We note that the recent structures of Mfd translocating on dsDNA (Kang et al., 2021) indicate
439 that R953 along with key residues in the translocation motif (TRG: translocation in RecG (Mahdi et al.,
440 2003)) of Mfd, specifically R929 and Q963, contacts the anti-sense strand in dsDNA. From these
441 structures it is not immediately obvious how the sense strand overhang is contacted by Mfd in the

442 smFRET assays. The results presented here uncover a new role for R953 in making critical contacts
443 with sense ssDNA distinct from its previously described role in translocation on dsDNA. This role
444 explains the observed deficiency of this mutant of Mfd in coupling ATP hydrolysis to translocation of
445 RNAP.

446 Intriguingly, each of the three Mfd variants also exhibited a distinct sub-population of
447 molecules centered at ~0.1 FRET. Here, whereas ATPγS promoted the formation of this state almost
448 immediately, Mfd (or mutants) in the presence of ATP entered this state in a time window ~5–20 s
449 after introduction into the flow cell (Figure 4F-H and Supplemental Figure 4H). The presence of the
450 low-FRET peak and the excursions to high-FRET states suggested that multiple independent binding
451 modes may underlie the FRET signal observed here.

452 **Mfd melts dsDNA in primed-DNA substrates**

453 To further understand how Mfd binds sense overhang–containing DNA, we visualized Mfd
454 bound to the dye-labeled smFRET substrate by cryo-EM (Figure 5A, Table 1). Full-length Mfd bound to
455 the primed-template substrate in the presence of transition-state analog ADP-AlF_x was subjected to
456 single particle cryo-EM analysis (Supplemental Figure 5A) and a ~5.2 Å resolution map was obtained
457 (real resolution ~8 Å, Supplemental Figure 5B, C). This reconstruction yielded EM density that could at
458 best accommodate the motor domains of Mfd (residues 575–956 representing domains 5 and 6);
459 other domains are presumably flexible and could not be seen in the EM density maps. In addition the
460 reconstruction could accommodate 18 bp of dsDNA and 11 nt of the 5' dT (sense) overhang (Figure
461 5B).

462 The cryo-EM map reconstructed here suggests the dsDNA on the upstream side of the motor
463 domains and the placement of the motor domains with respect to dsDNA to be consistent with
464 previously reported complexes (Brugger et al., 2020; Kang et al., 2021) (Figure 5B). However, EM
465 density for residues 917 onward, containing the TRG motif, appeared to be absent from the side of
466 domains 5 and 6, and the EM density of the DNA as it passes from the upstream to the downstream
467 side becomes inconsistent with dsDNA (Figure 5B). Using molecular dynamics flexible fitting

468 (Singharoy et al., 2016) to equilibrate the structure in density suggests that the TRG motif has dropped
469 into the path of the dsDNA (yellow ribbon, Figure 5C), resulting in the partial melting of the duplex.
470 The motif forms an arch that interacts with the template DNA (Figure 5C). There are precedents for
471 conformational plasticity in the TRG motif: the pair of helices forming the bulk of the TRG motif are
472 almost orthogonal in the apo-structure of Mfd (PDB ID 2EYQ; Figure 5C, orange) with the sulfonate
473 group from 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid bound between the helices, and are
474 parallel in the complex with ATP- and ADP-bound conformations on upstream dsDNA (PDB ID 6X43
475 and 6X4W). These movements are attributable to both to shifts in domains 5 and 6 relative to each
476 other and to movements of the helices relative to each other. The conformation of domains 5 and 6
477 suggested here may represent a mode of engagement whereby Mfd grabs and stabilizes the template
478 DNA in the upstream edge of the transcription bubble, supporting the inferences drawn from the
479 biophysical investigations.

480 The molecular model derived from the cryo-EM investigations presents two aspects of Mfd
481 binding that can be experimentally validated. Despite the poor resolution of the cryo-EM data which
482 limits confident assignment of the position of R953, the model is consistent with and supports the
483 observation from the smFRET studies that R953 mediates scrunching of the 5' ssDNA overhang (Figure
484 4).

485 Additionally, the model suggests that Mfd partially melts about five nucleotides of the dsDNA
486 near the junction forming a 'TRCF bubble' (Figure 5D). An additional attempt to obtain an improved
487 higher resolution reconstruction with an unmodified dT_6 sense overhang-containing 18-mer ds
488 template yielded a low-resolution reconstruction of limited utility (data not shown). Considering these
489 limitations, we used an orthogonal approach to test the model prediction that binding of Mfd in the
490 presence of a non-hydrolyzable ATP analog-leads to melting of the dsDNA in sense-overhang
491 containing dsDNA substrates.

492 2-Aminopurine (2AP), is a fluorescent analog of adenine that has been previously used to
493 study protein-induced local conformational changes in DNA (Jones and Neely, 2015). 2AP readily base

494 pairs with dT and its fluorescence is quenched in dsDNA. We reasoned that local melting of DNA could
495 be observed by monitoring the dequenching of the fluorescence signal that occurs when 2AP is
496 exposed to the solvent. To test whether nucleotide-bound Mfd destabilizes primed DNA templates,
497 we designed two sets of substrates (18 bp double-stranded with 6 nt of 5' sense strand overhang) –
498 a first set comprised of completely complementary double stranded substrates ('ds') with 2AP at
499 either position 11 or 12 (counted from the 5' end on the sense strand of the substrate; analogous
500 positions are highlighted in the model in Figure 5D), and a matched set where the antisense strand
501 contains a non-complementary sequence resulting in a 'bubble' opposite the introduced 2AP at
502 position 11 or 12 (Figure 5D).

503 First, we measured the difference in the emission spectra between the bubble and ds
504 substrates (2 μM each) for each of the substrates containing 2AP at position 11 or 12 (Figure 5E black
505 trace) upon excitation at 330 nm. This difference spectrum demonstrates that the fluorescence of 2AP
506 in the bubble substrates is greater than that in the double-stranded substrates, consistent with the
507 expectation that 2AP is quenched in dsDNA. This signal reflects the maximum fluorescence
508 enhancement resulting from exposure of 2AP upon local melting of ds DNA to form a bubble.

509 Next, the DNA substrate (2 μM, either ds or bubble) was incubated with Mfd (4 μM) in the
510 presence or absence of ATPγS (0.5 mM), and emission spectra were measured at steady state.
511 Incubation of the 2AP bubble substrates with Mfd and ATPγS resulted in a fluorescence enhancement
512 (Figure 5E, purple trace) for both substrates. Importantly, in the presence of Mfd and ATPγS, both 2AP-
513 containing dsDNA substrates exhibited an enhanced fluorescence (Figure 5E, green and yellow traces).
514 The increase in fluorescence was greatest when 2AP was introduced at position 12, and was
515 indistinguishable from the maximum fluorescence enhancement on this substrate. Together, the data
516 provide independent confirmation of the prediction from the cryo-EM derived molecular model that
517 Mfd melts DNA substrates containing a sense-strand overhang in the presence of non-hydrolyzable
518 ATP analogs. Consistent with this scenario, Mfd-ATPγS binds a 49-mer dsDNA substrate containing a
519 non-complementary bubble of 9 nucleotides with an affinity of 395 ± 7 nM as assessed by SPR

520 (Supplemental Figure 5D).

521 Together, the findings converge on a 'melt, hook and pull' model to describe how Mfd may
522 contact DNA substrates mimicking the upstream edge of the transcription bubble: SPR experiments
523 reveal that Mfd specifically recognizes the sense strand and this binding antagonizes binding to dsDNA.
524 Mutation of R953 weakens this interaction. Further, the cryo-EM data and 2AP fluorescence
525 experiments suggest that Mfd binds and partially melts the dsDNA near the junction, contacting the
526 unpaired template ssDNA *via* the conserved R953 residue in the TRG motif. Finally, the smFRET
527 experiments reveal that Mfd dynamically cycles between high- and low-affinity states in an ATP-
528 hydrolysis dependent manner on the template ssDNA overhang, pulling and scrunching the template
529 strand in this process, in an R953 dependent manner.

530 **Discussion**

531 In this work, we set out to understand dynamic modulation of EC activity by the action of interacting
532 partners. We detected that NusG promotes the residence of the EC in catalytically poised states. We
533 established that a 'molecular timer' intermediate is formed by Mfd, and ATP-hydrolysis-dependent
534 exit from this state determines the fate of the elongation complex. Our attempts to understand this
535 intermediate resulted in the discovery of a novel DNA-binding mode in which Mfd binds sense-strand
536 ssDNA in a transcription bubble mimic *via* a hook residue. Together, the results provide
537 unprecedented insight into the transcription termination reaction and enable us to propose a
538 comprehensive model describing how events unfold after initial engagement of the elongation
539 complex with the transcription-repair coupling factor.

540 Interactions of Mfd with the EC are complex. ATP-dependent remodeling of the EC by Mfd
541 proceeds *via* multiple stages: In the first stage, Mfd recognizes a distressed polymerase and is
542 recruited to the EC (Ho et al., 2018; Howan et al., 2012; Le et al., 2017; Park et al., 2002; Proshkin and
543 Mironov, 2016) through a conserved RNAP-interacting domain (Park et al., 2002) (Figure 6A). Mfd can
544 recognize extensively backtracked (Le et al., 2017), reversibly paused as well as stalled complexes
545 (Proshkin and Mironov, 2016). In the second stage, EC-tethered Mfd undergoes a series of concerted

546 conformational changes (Smith et al., 2012) that enable it to translocate on upstream dsDNA and
547 'load' onto the EC (Kang et al., 2021; Shi et al., 2020) (Figure 6B). In the third stage, Mfd either
548 reactivates paused transcription (Le et al., 2017; Park et al., 2002) or terminates transcription (Howan
549 et al., 2012; Selby and Sancar, 1993a). Alongside these activities, Mfd recruits its binding partners
550 $UvrA_2B$ to the site during transcription-coupled DNA repair (Fan et al., 2016; Ghodke et al., 2020; Ho
551 et al., 2018, 2020; Selby and Sancar, 1993a).

552 Our studies shed light on hitherto mysterious aspects of Mfd-RNAP interactions central to this
553 reaction and present a new model for resolution of a critical intermediate. The smFRET-based
554 elongation modulation assay revealed that Mfd can engage the EC in an ATP independent manner,
555 albeit infrequently, and manifest transient, reversible conformational changes in the EC. Stable
556 association requires ATP or at minimum non-hydrolyzable ATP, likely required to impart a dsDNA
557 binding conformation in the motor domain. Indeed, a recent study reported the structural basis of
558 ATP-dependent translocation of Mfd following initial recognition of distressed elongation complexes
559 leading to docking of Mfd with the upstream edge of the transcription bubble (Kang et al., 2021). Our
560 findings reinforce these previous findings (Figure 6 A – C).

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

568 The suggestion from the cryo-EM derived molecular model that Mfd-ADP-AlF_x^-might bind and
569 melt primed DNA templates to unwind dsDNA (Figure 5), and partial confirmation of this model in the
570 2AP fluorescence experiments provides insight into this transition. Only 13 nucleotides of the 18-mer
571 DNA template are consistent with being dsDNA, with the remainder being partially melted. The lack

572 of observations of dsDNA melting in a recently reported structure (Brugger et al., 2020) suggests that
573 Mfd requires access to unpaired ssDNA *via* critical residues in the TRG motif to melt the dsDNA.
574 We propose that in the context of the upstream dsDNA, ATP- or non-hydrolyzable analog-
575 bound Mfd partially melts about five nucleotides of dsDNA to produce a bubble, henceforth referred
576 to as the 'TRCF bubble' (Figure 5). Considering the three nucleotide proximity of the binding site of
577 Mfd to the upstream edge of the transcription bubble (Kang et al., 2021), the upstream DNA may
578 further melt allowing the TRCF bubble to fuse with the transcription bubble to produce the isomeric
579 TRCF-transcription ('TRCF-T') bubble (Figure 6D). Since prior experiments did not detect bubble
580 formation by Mfd on dsDNA lacking an overhang (Howan et al., 2012), a more likely possibility is that
581 Mfd merely extends the transcription bubble by further melting the upstream edge, resulting in
582 essentially a TRCF-T bubble. While our study was under review, an independent study reported that
583 'a highly negatively supercoiled domain forms between Mfd and RNAP', consistent with our report of
584 a TRCF-T bubble (Portman et al., 2021).
585 Although available structures of Mfd-[ATP or non-hydrolyzable analog]-EC do not report
586 melting of the upstream DNA or a TRCF-T bubble, and a structure of Mfd-EC bound to the TRCF-T
587 bubble is not as yet forthcoming, precisely such an intermediate has been observed in the Rad26-Pol
588 II structure where an extended TRCF-T bubble of 17 nucleotides is observed (Xu et al., 2017). Similarly,
589 human CSB was also reported to bind large transcription bubbles in a recent study (Kokic et al., 2021).
590 Our cryo-EM derived model suggests that R953 lines the path of the template overhang
591 (Figure 5). The observation that Mfd(R953A)-ATP fails to stably engage the ssDNA overhang in our
592 smFRET study supports a model in which this residue (potentially along with other residues in the
593 translocation motif) serves to stabilize the fork at the downstream edge of the TRCF bubble, or sense
594 ssDNA in the TRCF-T bubble. We propose that binding of R953 to the ssDNA template occurs
595 simultaneously with the melting of the upstream dsDNA. Evidently, along with a role in DNA
596 translocation, R953 has a newly discovered role in hooking and stabilizing ssDNA. Remarkably, a
597 similar element responsible for hooking the non-template DNA in the upstream fork in the case of CSB

598 was recently identified in a study published while ours was in revision (Kokic et al., 2021). Here, the
599 authors demonstrated that residue F796 in CSB serves as a hook by inserting itself in the upstream
600 fork in the non-template strand. In the yeast homolog Rad26, K613/R614 in the wedge motif together
601 serve to hook the template strand (Xu et al., 2017). While the wedge motif itself is absent in Mfd, R953
602 appears to play a functionally analogous role as a hook in the prokaryotic model. In all three cases,
603 this hook serves to couple ATP binding and hydrolysis to remodeling of the EC. These similarities – the
604 presence of a hooking element and destabilization of the upstream DNA – between the prototypical
605 members of this class of transcription-coupled DNA repair factors establish a more profound
606 conservation of mechanism of action.

607 We propose that the formation of the TRCF-T bubble results in forward translocation of the
608 EC (or at minimum restricts its backtracking) leading to the formation of the catalytically poised state
609 of the RNAP. Here, in the presence of the correct NTP and an intact R953, resumption of transcription
610 would lead to faster forward translocation of RNAP, leaving behind Mfd-ATP bound to the TRCF bubble
611 (Figure 6E). In effect this finding explains how Mfd can 'release' distressed polymerases in the so-
612 called 'release and catch-up' model for transcription rescue (Le et al., 2017). Indeed, this prediction of
613 our model was verified in the experiment demonstrating that Mfd bound to ATPγS leads to a
614 significantly truncated residence time in the intermediate state accompanied by loss of FRET
615 consistent with transcription rescue (Figure 3).

616 To understand what happens once the molecular timer intermediate (Figure 6D) forms, we
617 must first consider that exit from this intermediate is signified by a loss of FRET in the presence of ATP
618 but no other nucleotides. Our measurements of the residence time of the transcript through direct
619 excitation of the Cy5 matched the lifetime of the intermediate state, indicating that loss of FRET
620 corresponds to loss of the transcript from the complex. This interpretation is supported by previous
621 observations that loss of RNA from the EC occurs rapidly after Mfd binds the EC (Graves et al., 2015).

622 An important consequence of entry into this gating intermediate is that a finite time window
623 is available beyond which transcription restart is no longer possible. The prolonged residence in the

624 catalytically poised state prior to irreversible remodeling of the EC by Mfd may serve to provide an
625 opportunity to restart transcription at sites of pausing. This observation explains how Mfd can
626 orchestrate transcription restart (Le et al., 2017) or termination (Howan et al., 2012) on distressed
627 polymerases, discriminating paused ECs that can be substrates for transcription restart from
628 irreversibly or DNA damage-stalled ECs that are substrates for transcription termination.

629 It can be deduced that resolution of this gating intermediate during transcription termination
630 proceeds *via* at least partial, and potentially complete, shrinking of the TRCF-T bubble leaving behind
631 Mfd bound to the TRCF bubble. How might these events play out? Our results demonstrate that the
632 lifetime of this state is only ~6 s at 24°C. While entry into this state is dependent on the presence of
633 ATP, the observation that this lifetime is independent of ATP concentration in the range from 10 μM
634 to 10 mM suggests that hydrolysis of stoichiometric amounts of ATP, may be sufficient to promote
635 exit from this state. Further, an intact R953 is required for exit from this state. Since dissociation of
636 the 9-nt RNA:DNA hybrid (possessing a melting temperature significantly below RT) is expected to be
637 spontaneous at this temperature in the absence of stabilizing interactions with the EC, the loss of the
638 transcript signals the resolution of the molecular timer, accompanied by the formation of an
639 intermediate where RNAP is extensively remodeled, and the TRCF-T bubble is partially or completely
640 rewound. The cryo-EM derived model presented here sheds light on how Mfd may scrunch template
641 DNA in this process.

642 Mfd recruits the repair machinery to the site of distressed polymerases during transcription-
643 coupled nucleotide excision repair. During global genomic nucleotide excision repair, the damage
644 recognition enzyme $UvrA_2$ partially melts putative damage containing dsDNA and loads two copies of
645 the damage verification enzyme UvrB at the site (Kisker et al., 2013). Two UvrB helicases then examine
646 the unpaired ssDNA with single-nucleotide resolution to precisely identify the damaged strand (Jaciuk
647 et al., 2020; Pakotiprapha et al., 2012). In contrast, during transcription-coupled DNA repair, Mfd
648 recruits $UvrA_2B$ to the site of the distressed polymerase (Ghodke et al., 2020; Ho et al., 2020).
649 Compared to global genomic nucleotide excision repair, only a single copy of UvrB is recruited to a site

650 of damage by Mfd through interactions with $UvrA_2$. Thus, sequestration of specifically the damage-
651 containing template by Mfd supports the intriguing possibility that Mfd possesses the intrinsic ability
652 to faithfully convey the identity of the damaged strand to complete the handoff to a single UvrB *via*
653 the adaptor $UvrA_2$. Further elucidation of transcription-coupled DNA repair intermediates will shed
654 light on this model.

655 **Acknowledgments**

656 We thank Rachel Ann Mooney and Robert Landick (UW Madison) for generously sharing expression
657 vectors for RNA polymerase and purified NusG used in these studies. HG is supported by NIH Grant
658 1RM1GM130450 and acknowledges UOW Small Project Grant (2020/SPGA-S/04). ND, AJO, PJL, HG
659 acknowledge support by the Australian Research Council (DP210100365). AMvO acknowledges
660 support by the NIH (1RM1GM130450) and Australian Research Council (DP210100167).

661 **Author Contributions**

662 Data curation: BPP, ZX, SJ, SB, HG; Data analysis: BP, ZQ, SJ, HG, AJO, PJL; Writing—Original draft: HG;
663 Writing—review and editing: all authors; Resources: JB, ND and AvO; Conceptualization: BP and HG;
664 Supervision: HG.

665 **Declaration of Interests**

666 No competing interests.

667

668 **Materials and Methods**

669 **Proteins used in this study**

670 Purification of RNA polymerase

671 RNA polymerase was purified essentially as described previously (Svetlov and Artsimovitch, 2015).
672 pIA900 was a generous gift from Irina Artsimovitch (Addgene #104401)(Svetlov and Artsimovitch,
673 2015). pIA900 was transformed into BL21(λDE3)*recA*$^-$ cells. A single colony was isolated and streaked
674 on an ampicillin plate (100 mg L^{-1}). Four clones were then tested for small scale protein expression (25
675 mL) and were found to produce similar levels of overexpressed RNAP. One colony was then used for
676 overproduction in 6 L of LB medium, inoculated with an overnight culture and shaken for 2.5 h at 37
677 °C to reach A_{595} = 0.6, followed by induction with a final concentration of 1 mM IPTG for an additional
678 3 h, followed by pelleting. Overproduced protein was purified from cells following the exact protocols
679 described in (Svetlov and Artsimovitch, 2015). Following chromatographic separation over the Mono-
680 Q column, two fractions were isolated, one corresponding to the RNAP core complex, and a second
681 smaller fraction corresponding to the holoenzyme. For this work, the RNAP core complex was isolated
682 at a concentration of 35.7 mg mL^{-1} corresponding to 92 μM in storage buffer (50 mM Tris/HCl pH 7.6,
683 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol and 50% *v*/*v* glycerol). Purified protein was stored at
684 -80 °C following snap freezing with liquid N_2. Aliquots in use were stored at –20°C.

685 Overproduction and purification of Mfd variants

686 Sequence verified *mfd-ypet, mfd(E730Q)-ypet* and *mfd(R953A)-ypet* genes were obtained from
687 Aldevron. Plasmids for overexpression of Mfd, Mfd(E730Q) and Mfd(R953A) were created as follows:
688 first, *mfd* or mutant gene inserts (lacking the *ypet* gene) were amplified using PCR from the custom
689 synthesized constructs using the primers (pETMCSII_Mfd_F and pETMCSII_Mfd_R; Table 1). PCR
690 products were digested with with *Bam*HI and *Eco*RI and purified using a Qiagen miniprep kit. The *mfd*
691 (or mutant) fragments were inserted between the same restriction sites in vector pSH1017 (Hamdan
692 et al., 2002), previously isolated using electrophoretic separation and an extraction from a 1% (*w*/*v*)
693 agarose gel. Ligations were performed with T4 DNA ligase at 16 °C overnight, followed by

694 transformation into *E. coli* DH5α cells; colonies were screened for the presence of insert by colony
695 PCR and intact *mfd* sequences in plasmids pHG4137 (coding for wildtype Mfd), pHG42305
696 (Mfd(R953A)) and pHG42306 (Mfd(E730Q)) were confirmed by sequencing.

697 A generalized procedure for the expression and purification of Mfd variants used in this study
698 is best exemplified by the methods used in the production of Mfd(E730Q). *E. coli* strain
699 BL21(λDE3)*recA*/pHG42306 (E730Q) was grown at 37°C in LB medium supplemented with thymine
700 (50 mg L^{-1}) and ampicillin (200 mg L^{-1}) to A_{600} = 0.6. To induce Mfd(E730Q) overproduction, 1 mM IPTG
701 was added to the shaking culture. Cultures were grown for a further 3 h, and then chilled in ice. Cells
702 were harvested by centrifugation (11,000 x *g*; 6 min), frozen in liquid N_2 and stored at –80 °C.

703 After thawing, cells (17.1 g, from 6 L of culture) were resuspended in 120 mL lysis buffer (50
704 mM Tris/HCl, pH 7.6, 2 mM dithiothreitol, 0.5 mM EDTA, 20 mM spermidine) and lysed by two
705 passages through a French press operated at 12000 psi. Cell debris was removed from the lysate by
706 centrifugation (35,000 x *g*; 30 min) to yield the soluble Fraction I. Proteins in Fraction I that were then
707 precipitated by addition of solid ammonium sulfate (0.38 g mL^{-1}), followed by stirring for 30 min, were
708 collected by centrifugation (38,000 x *g*; 30 min) and dissolved in 45 mL buffer A (50 mM Tris/HCl, pH
709 7.6, 2 mM dithiothreitol, 0.5 mM EDTA, 10% *v*/*v* glycerol) containing 180 mM NaCl. The solution was
710 dialyzed against 2 L of the same buffer to yield Fraction II.

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

716 Fraction III was loaded at 1 mL min^{-1} onto the same DEAE column that had been equilibrated
717 with buffer A. After the column was washed with 80 mL of the same buffer, Mfd(E730Q) was eluted
718 at 2.5 mL min^{-1} using a linear gradient (700 mL) of 0 – 500 mM NaCl in buffer A. The Mfd(E730Q)
719 eluted in a single, relatively narrow, peak centered around 80 mM NaCl. Fractions containing

720 Mfd(E730Q) (40 mL) were pooled and dialyzed against four changes of 1.2 L of buffer B (50 mM
721 Na/MES, pH 6.5, 2 mM dithiothreitol, 0.5 mM EDTA, 10% *v/v* glycerol) to yield Fraction IV.
722 Fraction IV (40 mL) was loaded at 1.5 mL min^{-1} onto a SP Sepaharose FastFlow column (Cytiva;
723 2.5 x 13 cm) that had been equilibrated with buffer B. After washing the column with 80 mL of buffer
724 B, Mfd(E730Q) was eluted at 2.5 mL min^{-1} using a linear gradient (700 mL) of 0–500 mM NaCl in buffer
725 B. The Mfd(E730Q) eluted in a single peak centered around 100 mM NaCl. Fractions containing highly
726 purified Mfd(E730Q) were pooled and dialyzed against four changes of 1.2 L of storage buffer (40 mM
727 Tris/HCl, pH 7.6, 100 mM NaCl, 4 mM dithiothreitol, 0.5 mM EDTA, 30% *v/v* glycerol) to yield final
728 product containing 130 mg of protein at a final concentration of 46.7 μM. The same protocol yielded
729 51.5 mg of Mfd(R953A) at a final concentration of 18 μM. Final protein preparations were snap frozen
730 in liquid N_2 and stored at –80 °C.
731 In case of wildtype Mfd, a modified purification procedure was used whereby the fractions
732 containing Mfd from the second DEAE (40 mL) column were pooled and directly applied at 1 mL min^{-}
733 ^{1} onto a 2.5 x 10 cm column of heparin Sepharose 4B that had been equilibrated with buffer A + 50
734 mM NaCl. Mfd was eluted at 1 mL min^{-1}, in a broad peak centered around ~120 mM NaCl, using a
735 linear gradient (450 mL) of 50–400 mM NaCl in buffer A. Fractions containing highly purified Mfd were
736 pooled and dialyzed against 2 L of storage buffer (40 mM Tris/HCl, pH 7.6, 100 mM NaCl, 4 mM
737 dithiothreitol, 0.5 mM EDTA, 25% *v/v* glycerol) to yield 115 mg of protein at a final concentration of
738 40 μM. Purified Mfd was snap frozen in liquid N_2 and aliquots were stored at –80 °C.
739 NusG was a generous gift from Prof Robert Landick (University of Wisconsin-Madison) (Kang
740 et al., 2018).
741

742 **Oligonucleotides and DNA substrates**

743 All oligonucleotides were purchased from Integrated DNA Technologies (IDT, Singapore). Primed DNA
744 template for smFRET and cryo-EM was first annealed by hybridizing bio_AS18_Cy5 and Cy3_dT40_S18
745 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
746 and 1.05-fold excess of the bio_AS18_Cy5 complementary strand. Sequences are presented in Table
747 1.

748 **Single molecule FRET measurements**

749 <u>smFRET TIRF microscope setup</u>

750 A home built objective-type TIRF microscope based on a Nikon TiE2 model was used to record single
751 molecule movies. FRET was measured by excitation with a 532 nm laser and the emissions at 555 and
752 637 nm were collected using a band-pass filter at 555 nm and a long-pass filter at 650 nm. Scattered
753 light was removed by using a 560 nm long pass filter. Cy3 and Cy5 signals were separated by a 638 nm
754 dichroic using photometrics dual view (DV-2) and both signals were focused onto a CCD camera
755 (Hamamatsu C9 100–13), simultaneously. Data were collected at 5 frames s^{-1}. A power density
756 between 50–150 W cm^{-2} was used in typical experiments.

757 <u>Preparation of flow cells for imaging</u>

758 Coverslips were prepared as described previously (Johnston et al., 2020). Quartz coverslips were
759 cleaned with 100% ethanol and 1 mM KOH. Aminosilanization of cleaned coverslips was carried out
760 using a 1% (*v*/*v*) (3-aminopropyl)triethoxy silane (Alfa Aesar, A10668, UK) solution in acetone.
761 PEGylation was carried out by incubating a mixture of biotinPEG-SVA MW 5000 and mPEG-SVA MW
762 5000 (Laysan Bio, AL) at a ratio of 1:10 prepared in 50 mM MOPS pH 7.5 solution on the top of the
763 silanized coverslip for 3–4 h. PEGylated coverslips were stored under dry nitrogen gas at −20 °C.

764 Neutravidin solution was prepared in transcription buffer (1 × TB; 20 mM Tris/HCl pH 8.0, 40
765 mM KCl, 5 mM $MgCl_2$ and 1 mM β-mercaptoethanol) and spread on the top of dry a PEGylated
766 coverslip followed by a 10 min incubation. Sample flow chambers were created by sandwiching
767 polydimethylsiloxane (PDMS) on the top of the streptavidin coated coverslip. Then, blocking buffer (1

768 × TB + 0.25% (*v*/*v*) Tween 20) was injected into the channel to reduce non-specific binding of proteins
769 on the surface followed by 10–15 min incubation. This was then followed by washing with 1 × TB to
770 remove unbound Tween 20.

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

772 Protocols described by (Komissarova et al., 2003) were adapted for *in situ* reconstitution of ECs in
773 single-molecule FRET studies. Template DNA (1 μM) was annealed to RNA primer (3 μM) in 1 × TB by
774 heating to 65 °C for 5 min followed by 3 °C drops in temperature every 2 min until the reaction
775 temperature reached 27 °C, followed by a hold step at 25 °C. Typically, 200 μL of reaction were
776 prepared at a time. To form the RNAP:template (T70_31Cy3):RNA(R15_4Cy5) complex, 0.22 μL of 92
777 μM RNAP was incubated with 2 μL of template:RNA hybrid in 1 × TB at room temperature for 30 min.
778 Meanwhile, 200 μL of 100 pM of NT70_bio (Table 1) were immobilized on the neutravidin-coated
779 surface. Following this, the RNAP:template:RNA mixture was diluted serially 6000 × in a final
780 suspension of 200 μL of 1 × TB and introduced into the flow cell and incubated for at least 5 min to
781 promote annealing of the template DNA to the biotinylated non-template DNA resulting in formation
782 of the EC. At the end of this incubation step, the flow cell was washed with 200 μL of 1 × TB + 1 M KCl
783 and subsequently with 1 × TB supplemented with an oxygen-scavenging system (OSS) consisting of
784 protocatechuic acid (2.5 mM) and protocatechuate-3, 4-dioxigenase (50 nM) to reduce photo-
785 bleaching of the fluorophores and 2 mM Trolox to reduce photo-blinking of dyes, followed by imaging.
786 All elongation experiments in Figures 1 and 2 were performed at 100 μM of each NTP or dNTP (where
787 indicated). smFRET experiments involving NusG were performed at 500 nM of NusG in 1 × TB (Figure
788 2). smFRET experiments involving wildtype Mfd (Figures 3 and 4) were performed at 500 nM of Mfd
789 in 1 × TB. Experiments involving Mfd(R953A) (Figure 3 and 4) and Mfd(E730Q) (Figure 4) were
790 performed at 1 μM of Mfd mutant in 1 × TB . Concentrations of ATP and ATPγS were at 1 mM and 0.5
791 mM respectively unless otherwise specified.

792 Primed DNA template (Figure 4):

793 Primed DNA substrate (bio_AS18_Cy5: Cy3_dT40_S18) used in smFRET experiments in Figure 4 was

794 diluted to 50 pM in 200 μL of 1 × TB prior to immobilization in the flow cell.

795 637 nm transcript lifetime experiments

796 Lifetime of the Cy5 dye was measured by directly exciting *in situ* assembled ECs with a 637 nm laser
797 illumination with a flux of 4 W cm^{-2} (140 mW Vortran, Sacramento, CA, USA) with 200 ms time
798 resolution. Emissions were collected using a dichroic mirror (Dio1-R405/488/561/635, Semrock,
799 Rochester, NY, USA). 100-150 molecules were imaged for each condition. Images were corrected for
800 non-homogeneous laser illumination, and the fluorescence of the molecules (corrected for
801 background) in the entire field of view was calculated on a per pixel basis as a function of time using
802 Fiji, and is presented as mean and standard deviation in Figure 3.

803 A single exponential decay process best described the observed decay of the signal for the EC alone,
804 but insufficiently described the decay in case of ECs incubated with Mfd and ATP. Therefore, we
805 proceeded with globally fitting the data with a sum of two exponential decay processes. Global fitting
806 was performed by fitting the signals to equation (1) for the two conditions (subscript 1 ≡ EC alone, and
807 2 ≡ EC incubated with Mfd and ATP):

808 $$y_i(t) = A_{1,i}e^{-\frac{t}{\tau}} + A_{2,i}e^{-\frac{t}{\tau_{2,i}}} \text{ ..... (1)}$$

809 while specifying $A_{2,1}$ = 0 and $\tau_{2,1} = 0$ for the signal measured from the elongation complex alone.
810 Here, τ represents the photobleaching rate, and $\tau_{1,1}$ describes additional processes influencing the
811 lifetime of the Cy5. Global fitting was performed in OriginPro 2018 using the Levenberg-Marquardt
812 iteration algorithm and iterations were terminated when Chi-square reached a tolerance of 10^{-9}
813 signifying convergence. Optimized values are for EC alone (A_1 = 0.973 ± 0.005; τ_1 = 11.3 ± 0.1 s) and
814 for EC incubated with Mfd and ATP (A_1 = 0.49 ± 0.02, τ_1 = 11.3 ± 0.1 s; A_2 = 0.52 ± 0.02, τ_2 = 4.6 ± 0.2
815 s).

816 **smFRET Data Analysis**

817 Single-molecule data acquisition was carried out using NIS element software. Then, single-molecule
818 intensity time trajectories and FRET-time trajectories were generated in MASH-FRET software
819 (https://rna-fretools.github.io/MASH-FRET). Heat maps represent aggregated FRET signal for FRET-

820 ting molecules detected in the experiment. Overlaid contour plots represent population deciles, *i.e,*
821 the smallest contour represents the top 10th percentile of the population, the next largest contains
822 the 20th percentile and so on. The projection on the time axis represents the number of observations
823 over time, and the projection on the FRET axis represents the ensemble FRET signal. Transition density
824 plots (TDPs) represent transitions between states fit to individual, idealized trajectories using vbFRET.
825 Plotted here are kernel density plots of the 2D histograms of the data representing transitions
826 between consecutive states. Data analysis and figures are plotted using custom written Python scripts
827 provided in the supplemental methods.

828 **Surface plasmon resonance**

829 Activation of SPR chip:

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

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

836 The 49-mer dsDNA substrate was prepared by hybridizing NT_49_bio (final concentration of 1 μM)
837 with two-fold excess of TS_49 (Table 1) in 1 × TB in a final volume of 120 μL by heating in a heat block
838 at 90 °C for 5 min followed by slow cooling until room temperature was reached. NT_49_bio:TS_49
839 dsDNA substrate was resuspended in 1 × SPR buffer (30 mM Tris/HCl, pH 7.6, 40 mM KCl, 5 mM $MgCl_2$,
840 0.005% (*v*/*v*) surfactant P20, 0.25 mM EDTA, 0.5 mM dithiothreitol) to a final concentration of 2.5 nM
841 (of the biotinylated oligo) and introduced onto the SPR chip for immobilization, followed by a wash
842 step with 1M KCl. The signal from the dsDNA template corresponded to 106 response units (RU).
843 Measurement of the Mfd binding affinity for the immobilized dsDNA ligand was performed as
844 described below.

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

846 The RNAP:template (TS_49):RNA(RNA_15) complex was formed using the procedure described for *in*
847 *situ* reconstitution of ECs for smFRET measurements. Following this, biotinylated non-template strand
848 (NT49_bio; 10 μL of 10 μM working stock in water) was added to the mixture, which was incubated at
849 room temperature for 15 min. Next, 50 μL of nickel NTA resin suspension were washed and
850 equilibrated with 1 × TB. This resin resuspended in 50 μL of 1 × TB was then added to the mixture
851 containing the RNAP, template:RNA hybrid and non-template strand followed by incubation at 37 °C
852 for 15 min to allow binding of His-tagged RNAP complexes to the resin. The remaining steps were then
853 performed at room temperature. First, supernatant from this suspension was isolated by
854 centrifugation on a mini-benchtop centrifuge for 10 s. The pelleted resin was washed with 150 μL of
855 1 × TB supplemented with 1 M KCl for 2 min. This suspension was then subjected to centrifugation
856 and the supernatant was discarded. Following this, 150 μL of elution buffer (1 × TB + 500 mM
857 imidazole) were then applied to the resin followed by gentle mixing by pipetting. Finally, the
858 suspension was centrifuged and the eluate containing reconstituted elongation complexes was
859 collected. All reactions containing RNAP were performed in LoBind microcentrifuge tubes (Eppendorf).
860 The eluate was then applied to a size exclusion column (Wyatt SEC analytical column 030S5) previously
861 equilibrated with 30 mM Tris/HCl, pH 7.6, 150 mM KCl, 15% (*v*/*v*) glycerol, 0.5 mM EDTA, 1 mM
862 dithiothreitol to separate biotinylated elongation complexes from RNAP alone and collected in 200 μL
863 fractions. Fraction 7 (F7) was found to contain the elongation complex as verified by smFRET studies.
864 Immobilization of these reconstituted ECs was performed by resuspending 40 μL of F7 in 200 μL of
865 SPR buffer and introduction into the flow cell. Immobilization was stopped when the signal reached
866 1380 RU over the background. This was followed by treatment with 1M KCl (10 μL min^{-1}) that served
867 to displace incorrectly assembled or non-specifically bound RNAP from the chip surface. The resulting
868 signal corresponded to 550 RU over the background. ECs thus immobilized were found to be stable
869 for several days. Measurement of Mfd binding affinity for the immobilized EC ligand was performed
870 as described below.

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

872 Primed DNA templates were assembled *in situ* (on the chip surface) in two stages. First, the
873 biotinylated strand was immobilized in three of the four channels in the chip. Single-stranded
874 biotinylated DNA substrates (bio_dT40_S18, bio_AS18, and bio_AS18_Cy5; Table 1) were diluted to 1
875 nM in SPR buffer and immobilized in channels 2, 3, and 4 respectively. This corresponded to 150 RUs
876 of bio_dT40_S18 (channel 2), 75 RU of bio_AS18 (channel 3), and 80 RU of bio_AS18_Cy5 (channel 4).
877 Next, the complementary non-biotinylated strand (2 μM in 1 × SPR buffer) was introduced into the
878 flow cell to hybridize the template *in situ* at 5 μL min^{-1}. This led to a further increase in signal
879 corresponding to 39 RU (channel 2, complementary strand: AS18), 138 RU (channel 3, complementary
880 strand: dT40_S18) and 115 RU (channel 4, complementary strand: Cy3_dT40_S18) upon complete
881 hybridization. Following this step, the flow cell was washed with 1 M $MgCl_2$. Measurement of Mfd
882 binding affinity to immobilized primed DNA template was performed as described below.

883 Single-stranded biotinylated DNA substrate (bio_AS18) was diluted to 1 nM in SPR buffer and
884 immobilized (~48 RU) onto the flow cell of the SA chip by flowing at 5 μL min^{-1}. This was followed by
885 flowing in the complementary non-biotinylated strand (5 μM in 1 × SPR buffer) with defined lengths
886 of the ssDNA $(dT)_{0,3..15}$ overhang until the signal reached saturation indicating complete hybridization
887 (typically corresponding to a comparable increase in RU). Following this step, the flow cell was washed
888 with 1 M $MgCl_2$. Measurement of Mfd binding affinity was performed as described below.

889 <u>Titration of Mfd (Supplemental Figures 2 and 4)</u>:

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

897 Two conditioning steps of flushing the flow cell with 300 mM KCl for 30 s at 10 μL min^{-1} were carried

898 out prior to injection of Mfd. Mfd-ATPγS was flowed in for 150 s in the association phase, and
899 dissociation was observed for 200 s in 1 × SPR buffer lacking any Mfd or nucleotide cofactor.
900 Analysis of sensorgrams:
901 Sensorgrams were subtracted from the signal from unmodified flow cell (flow channel 1), then zero-
902 subtracted (injection at 0 μM Mfd or mutant) and R_{eq} values, generated by averaging response values
903 in the sensorgram's steady-state region from the appropriate Mfd concentration range, fit against
904 [Mfd] using a 1:1 steady state affinity (SSA) model incorporated in the BIAevaluation software 4.0.1
905 (Cytiva):
906 $$R_{\mathrm{eq}} = R_{\mathrm{max}}\left(\frac{[A]}{[A]+K_{\mathrm{D}}}\right) \text{.....} (2)$$
907 where R_{max} corresponds to the response when all the immobilized DNA ligands on the surface are
908 saturated with the analyte A, K_D is the dissociation constant, and [A] is the concentration of analyte
909 (Mfd or mutant) in solution. In general, raw sensorgrams and fit curves (where applicable) were
910 exported using Excel, then plotted in Origin, and final Figures prepared in Adobe Illustrator CC.
911 **Sample preparation for electron microscopy**
912 Electron microcopy data acquisition:
913 Mfd stock (40 μM; 30 μL) was dialyzed overnight in 1 L of 1 × TB at 4 °C in a Slida-A-lyzer mini dialysis
914 unit with 3,000 MWCO (Thermo Fisher) to reduce the amount of glycerol in the sample. Following this,
915 protein was recovered and concentrated to a 1.67 mg mL^{-1}. For the sample with ADP-AlF_x, to the
916 mixture of Mfd (7.8 μL at 1.6 mg mL^{-1}) and fluorescent oligonucleotides (10 μL of 10 μM stock), ADP,
917 NaF and $AlCl_3$ were added to final concentrations of 2, 5 and 0.5 mM, respectively. This reaction
918 mixture was incubated at 37°C for 5 min to allow binding of Mfd to the template. The samples were
919 clarified by centrifugation at 20,000 × g for 10 min at 4 °C before grid preparation. Samples (3 μL) were
920 applied to glow-discharged Quantifoil R 1.2/1.3, Au 400 grids. The grids were blotted at 6 °C for 3.5 s
921 at 100% humidity with no extra blot force in an FEI Vitrobot Mark IV and plunged into liquid ethane.
922 The grids were stored in liquid nitrogen until use.
923 **Electron microcopy data acquisition and image processing**

924 Data were collected using Thermo Fisher Talos Arctica microscope at 200 kV equipped with a Falcon
925 III (Thermo Fisher) camera in electron counting mode. Micrographs were acquired at a nominal
926 defocus range of −1.0 to −1.9 μm, with a pixel size of 0.74 Å and a total dose of 50 e^{-} $Å^{-2}$ accumulated
927 over 50 s and fractionated across 50 movie frames.
928 Data were processed in cryoSPARC v2 (Punjani et al., 2017). For the sample with ADP-AlF_x, 472 images
929 were manually selected after motion correction and CTF estimation for final processing. A total of
930 110562 particles were selected after 2D classification and subjected to 3D classification in *ab initio*
931 reconstruction followed by heterogeneous refinement using four *ab initio* classes. Particles (31718)
932 from class 3 were used for 3D refinement, which resulted in a 3D reconstruction at 5.2 Å resolution
933 (0.143 FSC gold standard; Supplemental Figure 5A-C).

934 Model fitting

935 Molecular dynamics with explicit solvent was used to fit the model of Mfd on primer DNA into the
936 cryo-EM density map (Singharoy et al., 2016). The starting model was comprised of the Mfd motor
937 domain (PDB ID 2EYQ) with idealized B-DNA with the same sequence as used for the cryo-EM
938 experiments but with a dT_{12} overhang instead of dT_{40}. The system was equilibrated for 10 ns. The
939 resulting model was then used as the basis for modeling the complex in COOT (Emsley et al., 2010)
940 with minimal distortions to protein secondary structure and DNA. The model was then subjected to
941 restrained positional refinement in phenix real_space_refine (Liebschner et al., 2019). The final model
942 is comprised of 382 residues from Mfd and 47 nucleotide residues. The model was visualized in
943 Chimera (Pettersen et al., 2004).

944 **Measurement of 2AP fluorescence**

945 Double-stranded (ds) and bubble substrates used in 2AP experiments were hybridized at a
946 concentration of 20 μM in 1 × TB by heating to 95 °C on a heat block for 5 minutes followed by turning
947 off the heat source and allowing the reaction to cool down to room temperature overnight.

948 dT6_S18_v2_11_2AP or dT6_S18_v2_12_2AP were hybridized with either AS18_v2_2AP or
949 AS18_v2_2AP_bubble (sequences are provided in Table 2) to create four substrates: (a) 11_2AP_ds ≡
950 dT6_S18_v2_11_2AP: AS18_v2_2AP (b) 11_2AP_bubble ≡ dT6_S18_v2_11_2AP:
951 AS18_v2_2AP_bubble (c) 12_2AP_ds ≡ dT6_S18_v2_12_2AP: AS18_v2_2AP and (d) 12_2AP_bubble ≡
952 dT6_S18_v2_12_2AP: AS18_v2_2AP_bubble. Starting with a stock concentration of 40 μM of Mfd and
953 20 μM of DNA substrates, protein and DNA were diluted 10-fold into 1 × TB in a total sample volume
954 of 120 μL and 100 mM ATPγS stock was supplemented to a final concentration of 0.5 mM as
955 appropriate. For purposes of background subtraction, we collected emission spectra of Mfd (4 μM)
956 alone, Mfd (4 μM) and ATPγS (0.5 mM), or ds or bubble substrates alone (2 μM). All samples were
957 incubated for at least 20 minutes at room temperature (24 °C) prior to measurements.
958 Spectra were collected on a Cary Eclipse in FP8300 emission mode. Samples were deposited in a clean
959 quartz cuvette with a total sample volume of 120 μL. Samples were excited at 330 nm (slit width 5
960 nm) and emission spectra were collected in the range from 350 nm – 460 nm with a slit width of 5 nm
961 and a scan rate of 30 nm/min. The averaging time for each measurement was set as 1 s and data were
962 collected at an interval of 0.5 nm. The PMT voltage was set to 800.

963 Two independent repeats were collected for each of the ds and bubble substrates (11_2AP_ds,
964 11_2AP_bubble, 12_2AP_ds, 12_2AP_bubble). Each spectrum was blanked for buffer, and the
965 difference spectrum was calculated by subtracting the blanked spectrum of the ds substrate from that
966 of the bubble substrate. Fluorescence enhancement was calculated by subtracting the signal
967 measured in the presence of Mfd and ATPγS from the signal measured in the presence of Mfd alone
968 for both substrates. The fluorescence emission of either ds or bubble DNA substrate (2 μM) was
969 measured in the presence of Mfd (4 μM) alone or with 0.5 mM ATPγS and blanked by subtracting the
970 corresponding emission spectrum of a sample of either Mfd (4 μM) or with 0.5 mM ATPγS in 1 × TB,
971 and lacking the DNA substrate. Three repeats were performed for 11_2AP_ds, 11_2AP_bubble and
972 12_2AP_bubble and two repeats were performed for the 12_2AP_ds substrate. The average

973 difference spectrum was then calculated at each wavelength in the emission spectrum and a five point
974 moving average was further used to smooth this signal for purposes of visualization. In the figures,
975 individual corrected measurements are plotted as points, and the average is plotted as a line.

976 **References**

977 Abdelkareem, M., Saint-Andre, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J., and
978 Weixlbaumer, A. (2019). Structural Basis of Transcription: RNA Polymerase Backtracking and
979 Its Reactivation. Mol Cell *75*, 298-309 e294.
980 Andrecka, J., Lewis, R., Bruckner, F., Lehmann, E., Cramer, P., and Michaelis, J. (2008). Single-
981 molecule tracking of mRNA exiting from RNA polymerase II. Proc Natl Acad Sci U S A *105*, 135-
982 140.
983 Belogurov, G.A., and Artsimovitch, I. (2019). The Mechanisms of Substrate Selection,
984 Catalysis, and Translocation by the Elongating RNA Polymerase. J Mol Biol *431*, 3975-4006.
985 Borukhov, S., Sagitov, V., and Goldfarb, A. (1993). Transcript cleavage factors from *E. coli*. Cell
986 *72*, 459-466.
987 Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., Jr., and Wiggins, C.H. (2009). Learning Rates
988 and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-
989 Molecule FRET Data. Biophys J *97*, 3196-3205.
990 Brugger, C., Zhang, C., Suhanovsky, M.M., Kim, D.D., Sinclair, A.N., Lyumkis, D., and
991 Deaconescu, A.M. (2020). Molecular determinants for dsDNA translocation by the
992 transcription-repair coupling and evolvability factor Mfd. Nat Commun *11*, 3740.
993 Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L., and Gottesman, M.E. (1995). *Escherichia coli*
994 NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. Journal of
995 bacteriology *177*, 1388-1392.
996 Byrd, A.K., and Raney, K.D. (2012). Superfamily 2 helicases. Front Biosci (Landmark Ed) *17*,
997 2070-2088.
998 Chambers, A.L., Smith, A.J., and Savery, N.J. (2003). A DNA translocation motif in the bacterial
999 transcription-repair coupling factor, Mfd. Nucleic Acids Res *31*, 6409-6418.
1000 Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J., and
1001 Darst, S.A. (2006). Structural Basis for Bacterial Transcription-Coupled DNA Repair. Cell *124*,
1002 507-520.
1003 Deaconescu, A.M., Sevostyanova, A., Artsimovitch, I., and Grigorieff, N. (2012). Nucleotide
1004 excision repair (NER) machinery recruitment by the transcription-repair coupling factor
1005 involves unmasking of a conserved intramolecular interface. Proc Natl Acad Sci U S A *109*,
1006 3353-3358.
1007 Duan, M., Selvam, K., Wyrick, J.J., and Mao, P. (2020). Genome-wide role of Rad26 in
1008 promoting transcription-coupled nucleotide excision repair in yeast chromatin. Proc Natl
1009 Acad Sci U S A *117*, 18608-18616.
1010 Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of
1011 Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.
1012 Fan, J., Leroux-Coyau, M., Savery, N.J., and Strick, T.R. (2016). Reconstruction of bacterial
1013 transcription-coupled repair at single-molecule resolution. Nature *536*, 234-237.

1014 Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R.H., and Bustamante, C.J. (2018). Pause
1015 sequences facilitate entry into long-lived paused states by reducing RNA polymerase
1016 transcription rates. Nat Commun *9*, 2930.
1017 Ghodke, H., Ho, H.N., and van Oijen, A.M. (2020). Single-molecule live-cell imaging visualizes
1018 parallel pathways of prokaryotic nucleotide excision repair. Nat Commun *11*, 1477.
1019 Graves, E.T., Duboc, C., Fan, J., Stransky, F., Leroux-Coyau, M., and Strick, T.R. (2015). A
1020 dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and
1021 fluorescence. Nat Struct Mol Biol *22*, 452-457.
1022 Haines, N.M., Kim, Y.I., Smith, A.J., and Savery, N.J. (2014). Stalled transcription complexes
1023 promote DNA repair at a distance. Proc Natl Acad Sci U S A *111*, 4037-4042.
1024 Hamdan, S., Bulloch, E.M., Thompson, P.R., Beck, J.L., Yang, J.Y., Crowther, J.A., Lilley, P.E.,
1025 Carr, P.D., Ollis, D.L., Brown, S.E*., et al.* (2002). Hydrolysis of the 5'-p-nitrophenyl ester of TMP
1026 by the proofreading exonuclease (epsilon) subunit of Escherichia coli DNA polymerase III.
1027 Biochemistry *41*, 5266-5275.
1028 Herbert, K.M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., and Block, S.M. (2010). *E. coli*
1029 NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward
1030 Translocation of RNA Polymerase. J Mol Biol *399*, 17-30.
1031 Ho, H.N., van Oijen, A.M., and Ghodke, H. (2018). The transcription-repair coupling factor Mfd
1032 associates with RNA polymerase in the absence of exogenous damage. Nat Commun *9*, 1570.
1033 Ho, H.N., van Oijen, A.M., and Ghodke, H. (2020). Single-molecule imaging reveals molecular
1034 coupling between transcription and DNA repair machinery in live cells. Nat Commun *11*, 1478.
1035 Howan, K., Smith, A.J., Westblade, L.F., Joly, N., Grange, W., Zorman, S., Darst, S.A., Savery,
1036 N.J., and Strick, T.R. (2012). Initiation of transcription-coupled repair characterized at single-
1037 molecule resolution. Nature *490*, 431-434.
1038 Jaciuk, M., Swuec, P., Gaur, V., Kasprzak, J.M., Renault, L., Dobrychlop, M., Nirwal, S., Bujnicki,
1039 J.M., Costa, A., and Nowotny, M. (2020). A combined structural and biochemical approach
1040 reveals translocation and stalling of UvrB on the DNA lesion as a mechanism of damage
1041 verification in bacterial nucleotide excision repair. DNA Repair (Amst) *85*, 102746.
1042 Johnston, C.L., Marzano, N.R., Paudel, B.P., Wright, G., Benesch, J.L.P., van Oijen, A.M., and
1043 Ecroyd, H. (2020). Single-molecule fluorescence-based approach reveals novel mechanistic
1044 insights into human small heat shock protein chaperone function. J Biol Chem.
1045 Jones, A.C., and Neely, R.K. (2015). 2-Aminopurine as a fluorescent probe of DNA
1046 conformation and the DNA-enzyme interface. Q Rev Biophys *48*, 244-279.
1047 Kang, J.Y., Llewellyn, E., Chen, J., Olinares, P.D.B., Brewer, J., Chait, B.T., Campbell, E.A., and
1048 Darst, S.A. (2021). Structural basis for transcription complex disruption by the Mfd
1049 translocase. Elife *10*.
1050 Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T.V., Artsimovitch, I., Landick, R.,
1051 and Darst, S.A. (2018). Structural Basis for Transcript Elongation Control by NusG Family
1052 Universal Regulators. Cell *173*, 1650-1662 e1614.
1053 Kisker, C., Kuper, J., and Van Houten, B. (2013). Prokaryotic Nucleotide Excision Repair. Cold
1054 Spring Harb Perspect Biol *5*, a012591.
1055 Kokic, G., Wagner, F.R., Chernev, A., Urlaub, H., and Cramer, P. (2021). Structural basis of
1056 human transcription-DNA repair coupling. Nature.
1057 Komissarova, N., and Kashlev, M. (1997). Transcriptional arrest: Escherichia coli RNA
1058 polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc
1059 Natl Acad Sci U S A *94*, 1755-1760.
1060 Komissarova, N., Kireeva, M.L., Becker, J., Sidorenkov, I., and Kashlev, M. (2003). Engineering

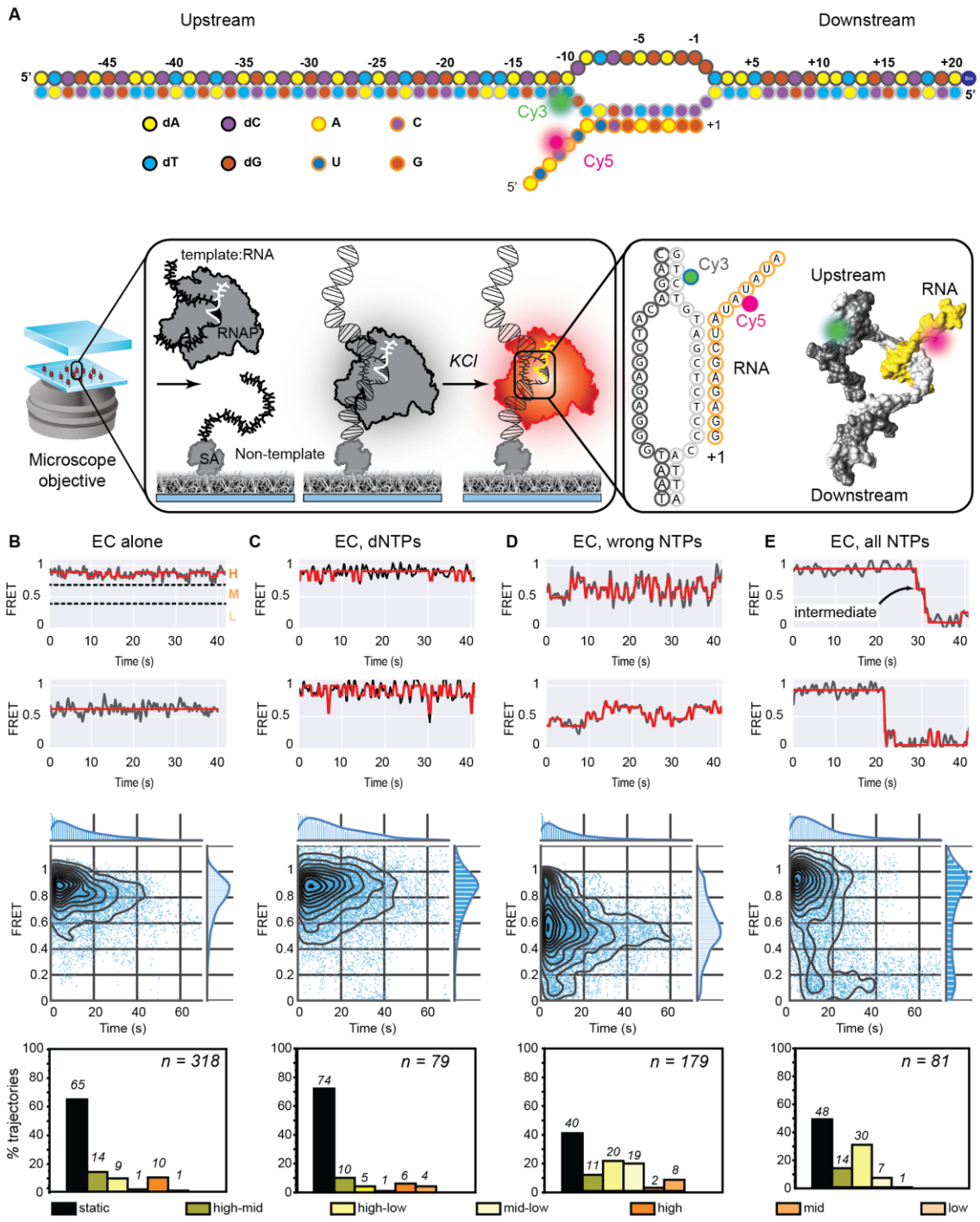
1061 of elongation complexes of bacterial and yeast RNA polymerases. Methods Enzymol *371*, 233-
1062 251.
1063 Le, T.T., Yang, Y., Tan, C., Suhanovsky, M.M., Fulbright, R.M., Jr., Inman, J.T., Li, M., Lee, J.,
1064 Perelman, S., Roberts, J.W.*, et al.* (2017). Mfd Dynamically Regulates Transcription via a
1065 Release and Catch-Up Mechanism. Cell.
1066 Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung,
1067 L.W., Jain, S., McCoy, A.J.*, et al.* (2019). Macromolecular structure determination using X-rays,
1068 neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol *75*, 861-
1069 877.
1070 Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q., and Lloyd, R.G. (2003). A model for dsDNA
1071 translocation revealed by a structural motif common to RecG and Mfd proteins. EMBO J *22*,
1072 724-734.
1073 Manelyte, L., Kim, Y.I., Smith, A.J., Smith, R.M., and Savery, N.J. (2010). Regulation and Rate
1074 Enhancement during Transcription-Coupled DNA Repair. Mol Cell *40*, 714-724.
1075 Mayne, L.V., and Lehmann, A.R. (1982). Failure of RNA synthesis to recover after UV
1076 irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma
1077 pigmentosum. Cancer Res *42*, 1473-1478.
1078 Mellon, I., and Hanawalt, P.C. (1989). Induction of the *Escherichia coli* lactose operon
1079 selectively increases repair of its transcribed DNA strand. Nature *342*, 95-98.
1080 Pakotiprapha, D., Samuels, M., Shen, K., Hu, J.H., and Jeruzalmi, D. (2012). Structure and
1081 mechanism of the UvrA-UvrB DNA damage sensor. Nat Struct Mol Biol *19*, 291-298.
1082 Park, J., Myong, S., Niedziela-Majka, A., Lee, K.S., Yu, J., Lohman, T.M., and Ha, T. (2010). PcrA
1083 helicase dismantles RecA filaments by reeling in DNA in uniform steps. Cell *142*, 544-555.
1084 Park, J.S., Marr, M.T., and Roberts, J.W. (2002). *E. coli* Transcription Repair Coupling Factor
1085 (Mfd protein) Rescues Arrested Complexes by Promoting Forward Translocation. Cell *109*,
1086 757-767.
1087 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and
1088 Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and
1089 analysis. J Comput Chem *25*, 1605-1612.
1090 Portman, J.R., Brouwer, G.M., Bollins, J., Savery, N.J., and Strick, T.R. (2021). Cotranscriptional
1091 R-loop formation by Mfd involves topological partitioning of DNA. Proc Natl Acad Sci U S A
1092 *118*.
1093 Proshkin, S.A., and Mironov, A.S. (2016). Stalled RNA Polymerase Is a Target of the Mfd Factor.
1094 Mol Biol (Mosk) *50*, 381-384.
1095 Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for
1096 rapid unsupervised cryo-EM structure determination. Nat Methods *14*, 290-296.
1097 Ragheb, M.N., Merrikh, C., Browning, K., and Merrikh, H. (2021). Mfd regulates RNA
1098 polymerase association with hard-to-transcribe regions in vivo, especially those with
1099 structured RNAs. Proc Natl Acad Sci U S A *118*.
1100 Selby, C.P., and Sancar, A. (1993a). Molecular Mechanism of Transcription-Repair Coupling.
1101 Science *260*, 53-58.
1102 Selby, C.P., and Sancar, A. (1993b). Transcription-Repair Coupling and Mutation Frequency
1103 Decline. Journal of bacteriology *175*, 7509-7514.
1104 Selby, C.P., and Sancar, A. (1995). Structure and Function of Transcription-Repair Coupling
1105 Factor. I. Structural Domains and Binding Properties. J Biol Chem *270*, 4882-4889.
1106 Selby, C.P., and Sancar, A. (1997). Cockayne syndrome group B protein enhances elongation
1107 by RNA polymerase II. Proc Natl Acad Sci U S A *94*, 11205-11209.

1108 Shi, J., Wen, A., Zhao, M., Jin, S., You, L., Shi, Y., Dong, S., Hua, X., Zhang, Y., and Feng, Y. (2020).
1109 Structural basis of Mfd-dependent transcription termination. Nucleic Acids Res.
1110 Singharoy, A., Teo, I., McGreevy, R., Stone, J.E., Zhao, J., and Schulten, K. (2016). Molecular
1111 dynamics-based refinement and validation for sub-5 A cryo-electron microscopy maps. eLife
1112 *5*.
1113 Smith, A.J., Pernstich, C., and Savery, N.J. (2012). Multipartite control of the DNA translocase,
1114 Mfd. Nucleic Acids Res *40*, 10408-10416.
1115 Smith, A.J., Szczelkun, M.D., and Savery, N.J. (2007). Controlling the motor activity of a
1116 transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. Nucleic
1117 Acids Res *35*, 1802-1811.
1118 Svetlov, V., and Artsimovitch, I. (2015). Purification of bacterial RNA polymerase: tools and
1119 protocols. Methods Mol Biol *1276*, 13-29.
1120 Svetlov, V., Vassylyev, D.G., and Artsimovitch, I. (2004). Discrimination against
1121 Deoxyribonucleotide Substrates by Bacterial RNA Polymerase. J Biol Chem *279*, 38087-38090.
1122 Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J.H.
1123 (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's
1124 syndrome and preferential repair of active genes. Cell *71*, 939-953.
1125 van der Weegen, Y., Golan-Berman, H., Mevissen, T.E.T., Apelt, K., Gonzalez-Prieto, R.,
1126 Goedhart, J., Heilbrun, E.E., Vertegaal, A.C.O., van den Heuvel, D., Walter, J.C.*, et al.* (2020).
1127 The cooperative action of CSB, CSA, and UVSSA target TFIIH to DNA damage-stalled RNA
1128 polymerase II. Nat Commun *11*, 2104.
1129 van Gool, A.J., Verhage, R., Swagemakers, S.M., van de Putte, P., Brouwer, J., Troelstra, C.,
1130 Bootsma, D., and Hoeijmakers, J.H. (1994). RAD26, the functional S. cerevisiae homolog of the
1131 Cockayne syndrome B gene ERCC6. EMBO J *13*, 5361-5369.
1132 Venema, J., Mullenders, L.H., Natarajan, A.T., van Zeeland, A.A., and Mayne, L.V. (1990). The
1133 genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA
1134 damage in transcriptionally active DNA. Proc Natl Acad Sci U S A *87*, 4707-4711.
1135 Wang, B., and Artsimovitch, I. (2020). NusG, an Ancient Yet Rapidly Evolving Transcription
1136 Factor. Front Microbiol *11*, 619618.
1137 Xu, J., Lahiri, I., Wang, W., Wier, A., Cianfrocco, M.A., Chong, J., Hare, A.A., Dervan, P.B.,
1138 DiMaio, F., Leschziner, A.E.*, et al.* (2017). Structural basis for the initiation of eukaryotic
1139 transcription-coupled DNA repair. Nature *551*, 653-657.
1140

1141

1142 **Figure legends**

Figure 1



1143

1144

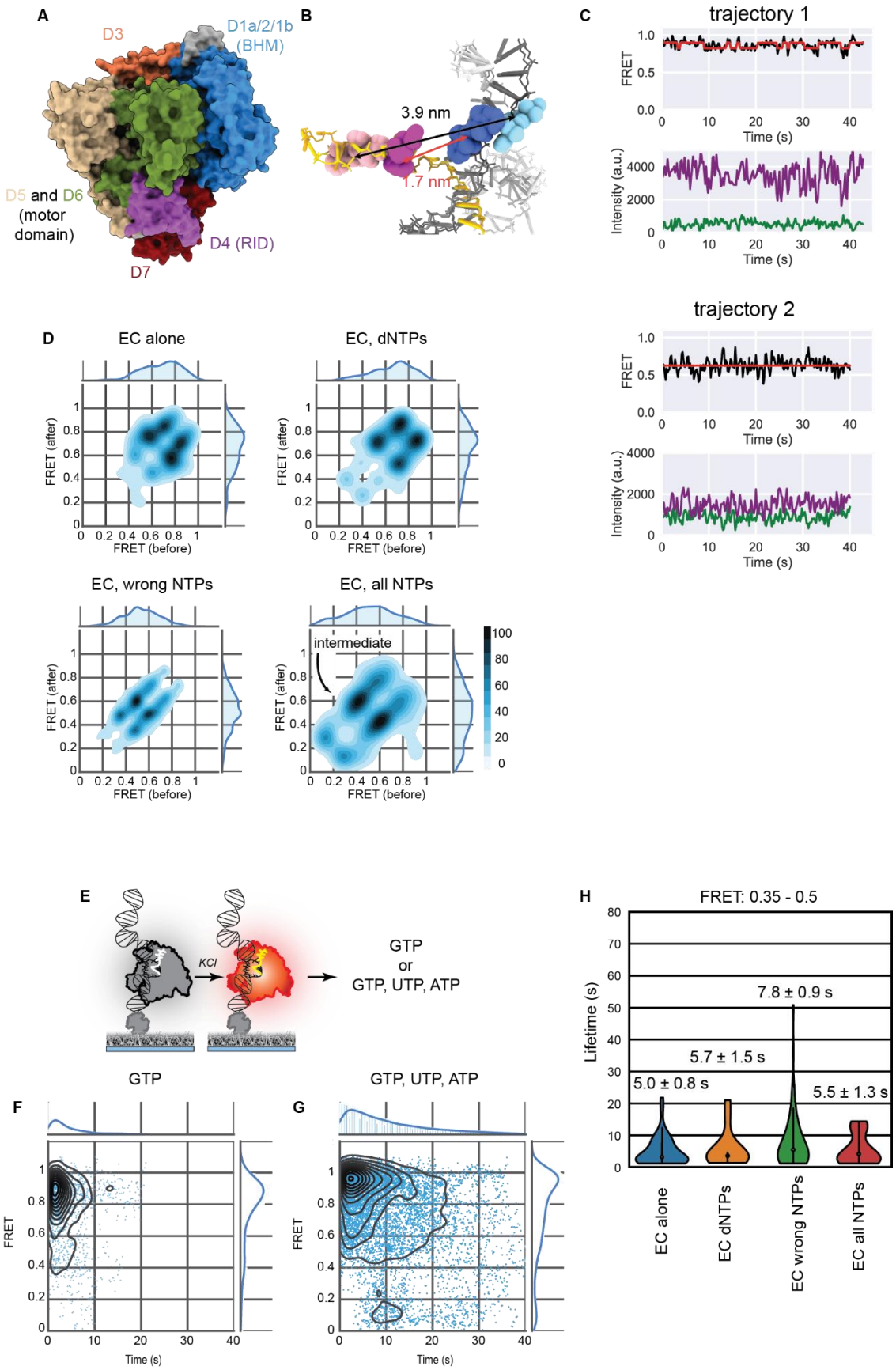
1145 **Figure 1: Reconstitution of a functional EC.**

1146 (A) Schematic of the transcription bubble used in this study. The sequence of the four next correct
1147 incorporations is G, U, A and A. ECs were assembled on dye-labeled nucleic acids by flowing in pre-
1148 incubated RNAP-template:RNA complexes into a flow cell with immobilized biotinylated non-template
1149 strand (EC shown in grey). Following initial immobilization, a high-salt wash step was used to remove
1150 poorly hybridized ECs from the surface. Here, only complexes containing both the Cy3-labeled
1151 template strand and Cy5-labeled transcript are observed when stably associated with the biotinylated
1152 non-template strand (shown in orange).

1153 (B-E) Two example FRET trajectories (black) and idealized fits (red) (top panel), temporal heat maps
1154 (middle panel) and column plots indicating percentages of trajectories exhibiting static or dynamic
1155 behavior (exhibiting transitions between or within high, mid, or low FRET regimes) (bottom panel) are
1156 provided for (B) ECs alone (n = 318 molecules) or in the presence of (C) dNTPs (n = 79 molecules), (D)
1157 non-complementary ('wrong') NTPs (n = 179 molecules) and (E)the full set of NTPs (n = 81 molecules)
1158 respectively.

1159

Supplemental Figure 1



1160

1161 **Supplemental Figure 1: Walking the EC with subsets of correct rNTPs.** ***Related to Figure 1.***

1162 (A) Crystal structure of apo Mfd (PDB 2EYQ) colored by functional modules: UvrB homology module
1163 (blue), domain 3 (coral), domain 4 (RNAP interacting domain; purple), motor domains 5 (tan) and 6
1164 (olive), auto-inhibitory domain7 (dark red).

1165 (B) Schematic of the *E coli* transcription bubble (non-template: light gray, template: dark gray, RNA:
1166 gold) with label positions highlighted. Blue spheres represent Cy3 and magenta/pink spheres
1167 represent Cy5 inserted in the backbone of the nucleic acids. The distances between the cyanine dyes
1168 are indicated for the pre-translocated register (1.7 nm) and post-translocated register (3.9 nm) (see
1169 also Supplemental Note 1).

1170 (C) Two example FRET trajectories of the EC alone (as shown in Figure 1B) and corresponding intensity
1171 traces for the Cy3 donor signal (green) and Cy5 acceptor (magenta). Trajectories are truncated at the
1172 photobleaching step.

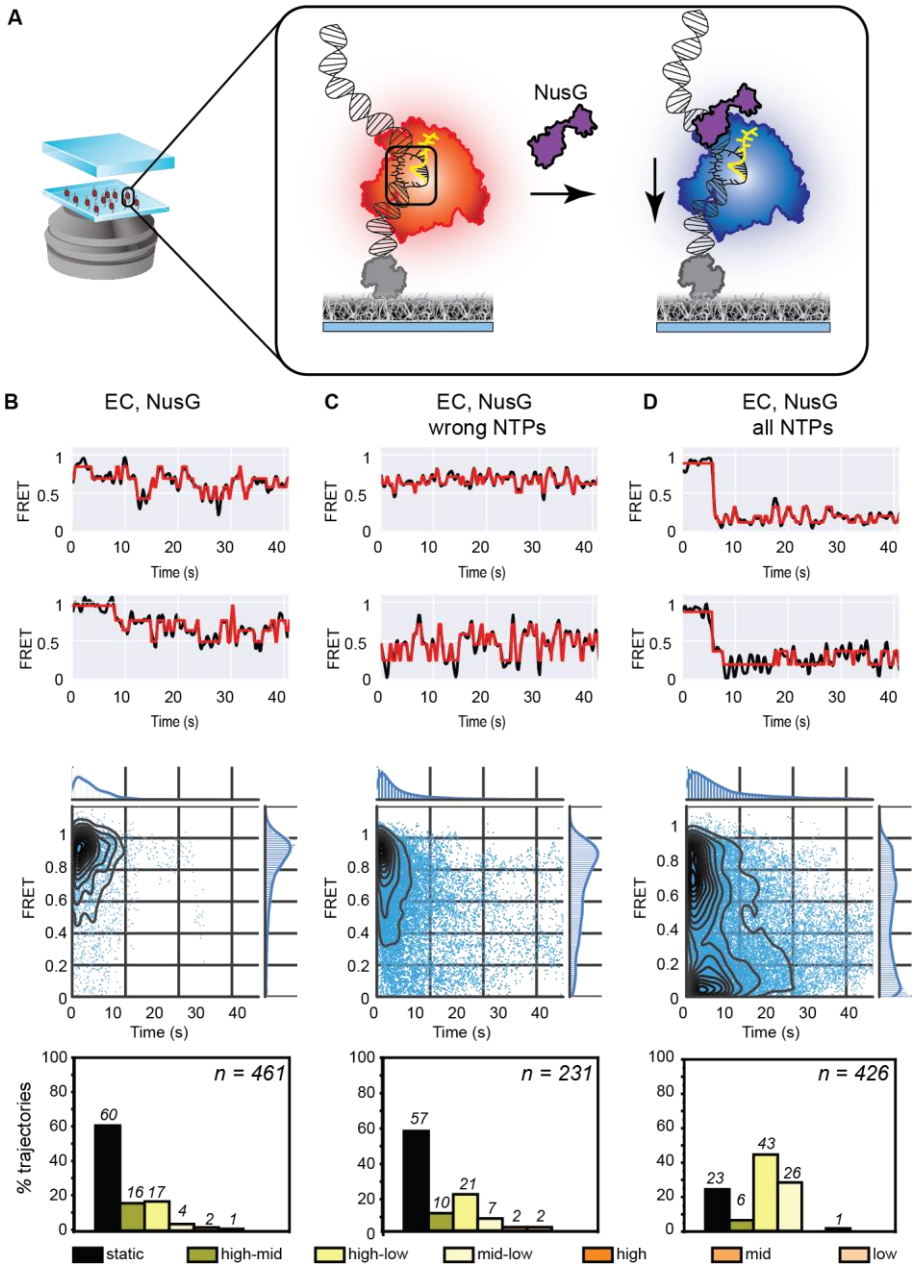
1173 (D) Transition density plots for RNAP alone (n = 318 molecules), in the presence of dNTPs (n = 79
1174 molecules), with non-complementary ('wrong') NTPs (n = 179 molecules) and the full set of NTPs (n =
1175 81 molecules) respectively.

1176 (E) Schematic of *in situ* assembled FRET-pair labeled ECs incubated with GTP (the next correct NTP) (n
1177 = 63 molecules) or with GTP, ATP and UTP (n = 87 molecules). Temporal heat maps of the reaction for
1178 experiments conducted in the presence of GTP (F) or GTP, ATP and UTP (G) are presented here.

1179 (H) Lifetimes of FRET intermediates observed in indicated conditions in the FRET range from 0.35 to
1180 0.5. See Supplemental Note 2 and Table 2.

1181

Figure 2



1182

1183

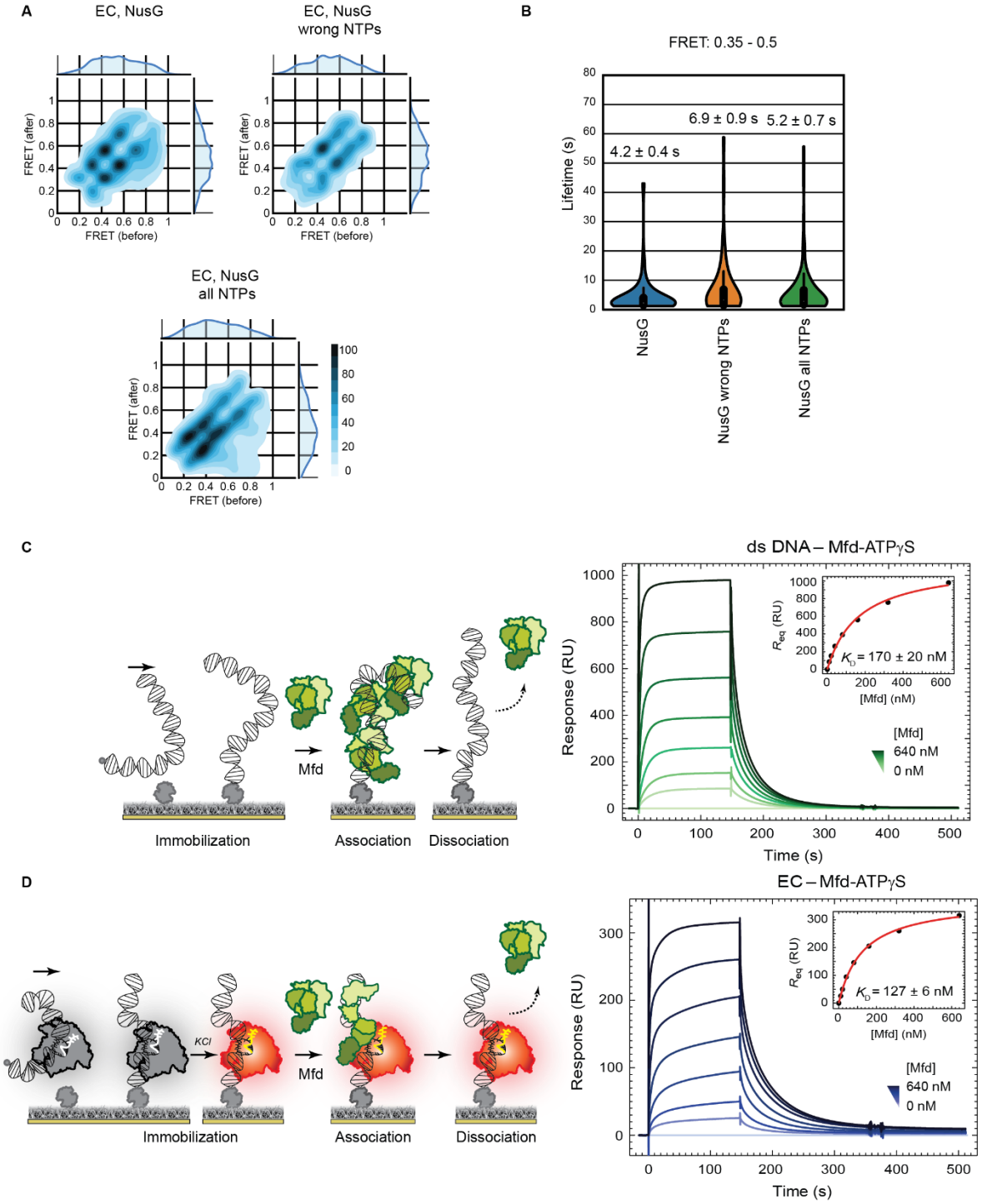
1184 **Figure 2: Interactions between NusG and the EC**

1185 (A) Schematic of FRET-pair labeled ECs incubated with NusG (purple).

1186 (B-D) Two example FRET trajectories (black) and idealized fits (red) (top panel), temporal heat maps
1187 (middle panel) and column plots indicating percentages of trajectories exhibiting static or dynamic
1188 behavior (exhibiting transitions between or within high, mid, or low FRET regimes) (bottom panel) are
1189 provided for (B) EC and NusG alone (n = 461 molecules), (C) EC and NusG in the presence of wrong
1190 NTPs (n = 231 molecules), and (D) EC and NusG with the full set of NTPs (n = 426 molecules). NusG
1191 promotes the residence of the EC in a catalytically poised state, and enhances transitions to low FRET
1192 states.

1193

Supplemental Figure 2



1194

1195

1196 **Supplemental Figure 2: Measurement of binding affinity of Mfd for dsDNA and the EC.** ***Related to***

1197 ***Figures 2 and 3.***

1198 (A)Transition density plots for EC and NusG alone (n = 461 molecules), EC and NusG in the presence

1199 of wrong NTPs (n = 231 molecules), and EC and NusG with the full set of NTPs (n = 426 molecules).

1200 (B) Lifetimes of FRET intermediates observed in indicated conditions in the FRET range from 0.35 to

1201 0.5. See Supplemental Note 2 and Table 2.

1202 (C, D) Schematic and SPR sensorgrams show association (150 s) and dissociation phases of serially-

1203 diluted 10 – 640 μM Mfd-ATPγS, including a 0 μM control for (C) biotinylated 49-mer dsDNA, and (D)

1204 *in vitro* reconstituted EC on biotinylated 49-mer DNA flowed into the SPR chip. Responses at

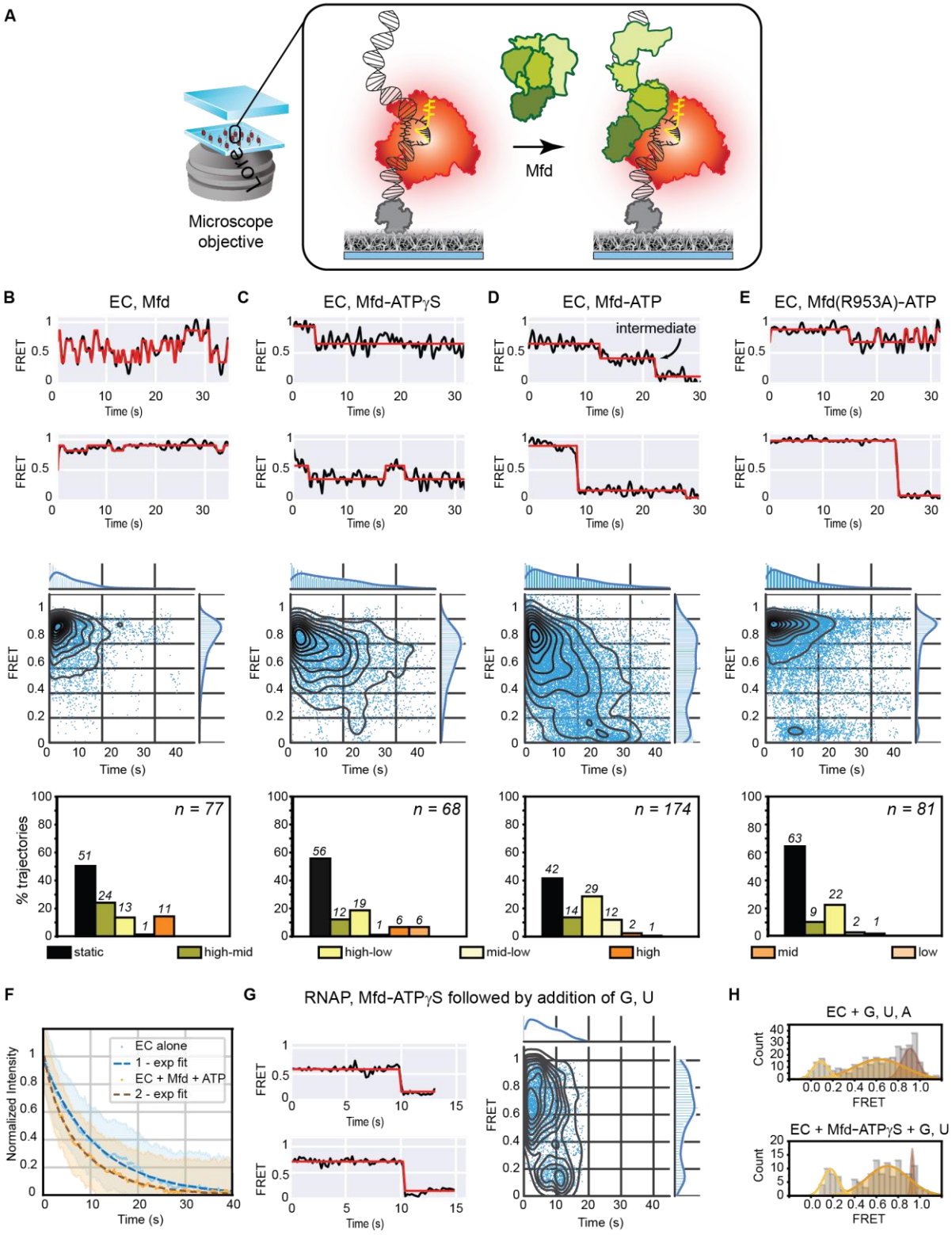
1205 equilibrium averaged over the steady region of sensorgrams (R_{eq}) were fit (insets) using a steady-state

1206 affinity (SSA) model (Equation 1, Materials and Methods) to derive values of K_D (as indicated). Errors

1207 are S.E. of the fit.

1208

Figure 3



1209

1210

1211 **Figure 3: Interactions between Mfd and the EC.**

1212 (A) Schematic of FRET-pair labeled, surface-immobilized ECs incubated with Mfd in solution.

1213 (B-E) Two example FRET trajectories (black) and idealized fits (red) (top panel), temporal heat maps
1214 (middle panel) and column plots indicating percentages of trajectories exhibiting static or dynamic
1215 behavior (exhibiting transitions between or within high, mid, or low FRET regimes) (bottom panel) are
1216 provided for (B) EC and Mfd alone (n = 77 molecules), (C) EC and Mfd in the presence of 0.5 mM ATPγS
1217 (n = 68 molecules), (D) EC and Mfd in the presence of 1 mM ATP (n = 174 molecules) and (E) EC and
1218 DNA-binding competent but translocase-deficient Mfd(R953A) in the presence of 1 mM ATP (n = 81
1219 molecules).

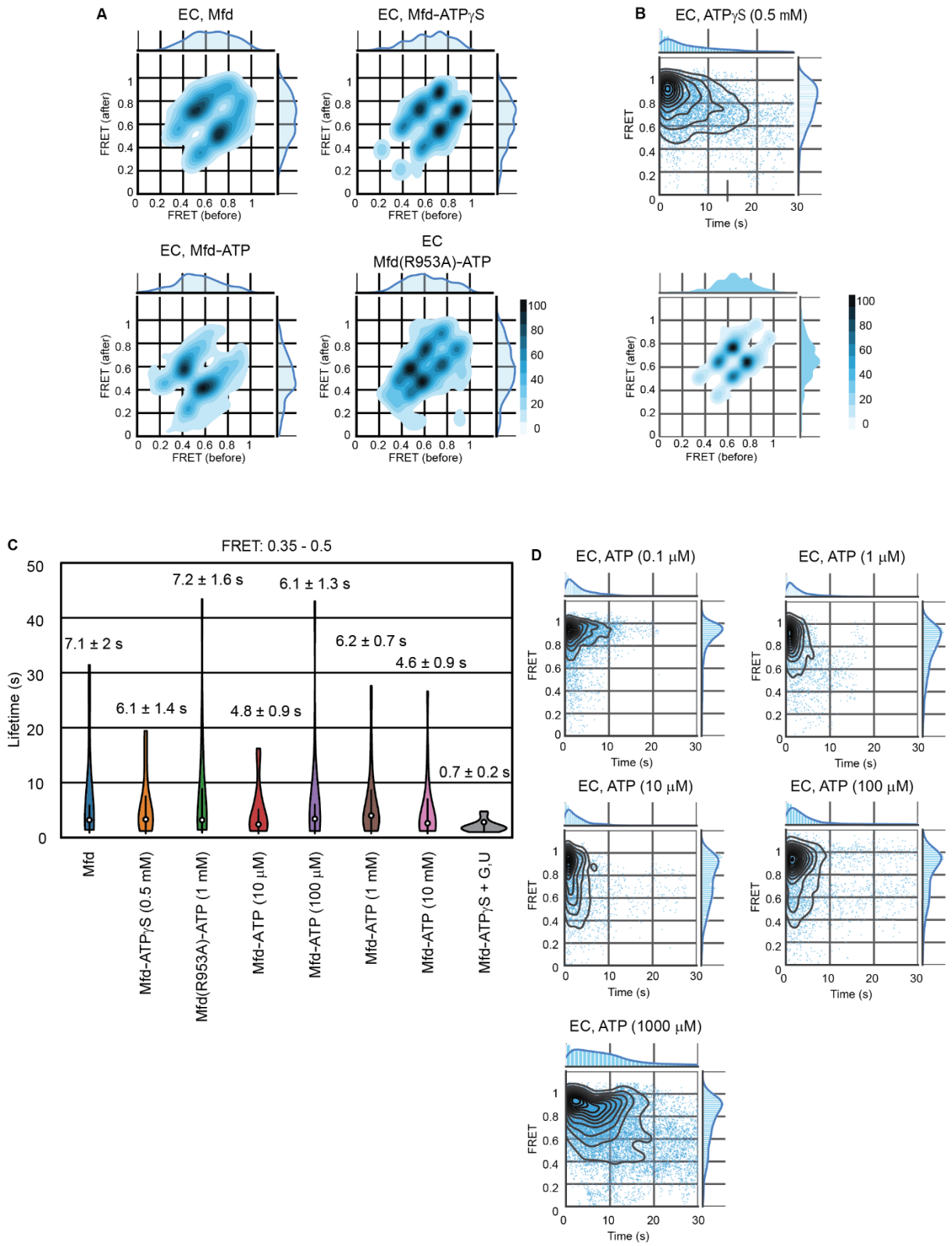
1220 (F) Normalized intensity of Cy5 fluorescence in the field of view following direct red laser excitation of
1221 the EC alone (blue circles) or incubated with Mfd and 1mM ATP (yellow circles). Single (blue) and
1222 double (brown) – exponential fits to the EC alone or when incubated with Mfd and 1 mM ATP are
1223 indicated as dashed lines. Error bars represent standard error of the fit.

1224 (G) Example FRET trajectories of ECs incubated with Mfd and ATPγS followed by addition of 100 μM
1225 each of GTP and UTP. Corresponding heat map shows an overview of the transition to low-FRET states
1226 upon incorporation of GTP and UTP.

1227 (H) Histograms (gray bars) of FRET states accessed by the EC when incubated with 100 μM each of
1228 GTP, UTP and ATP (top panel), or pre-incubated with Mfd-ATPγS (10 min) followed by 100 μM each of
1229 GTP and UTP. Each histogram was fit to the sum of three individual Gaussian curves (see Supplemental
1230 Note 3).

1231

Supplemental Figure 3



1232

1233

1234 **Supplemental Figure 3: Interactions of ATP with the EC alone or in the presence of Mfd.** ***Related to***

1235 ***Figure 3.***

1236 (A) Transition density plots for EC and Mfd alone (n = 75 molecules), EC and Mfd in the presence of 1

1237 mM ATPγS (n = 68 molecules), EC and Mfd in the presence of 1mM ATP (n = 174 molecules), and EC

1238 and Mfd(R953A) in the presence of 1 mM ATP (n = 234 molecules).

1239 (B) Temporal heat map (upper panel) and transition density plot (lower panel) of FRET pair labelled

1240 RNAP EC in the presence of 0.5 mM ATPγS alone (n = 124 molecules).

1241 (C) Lifetimes of FRET intermediates observed in indicated conditions in the FRET range from 0.35 to

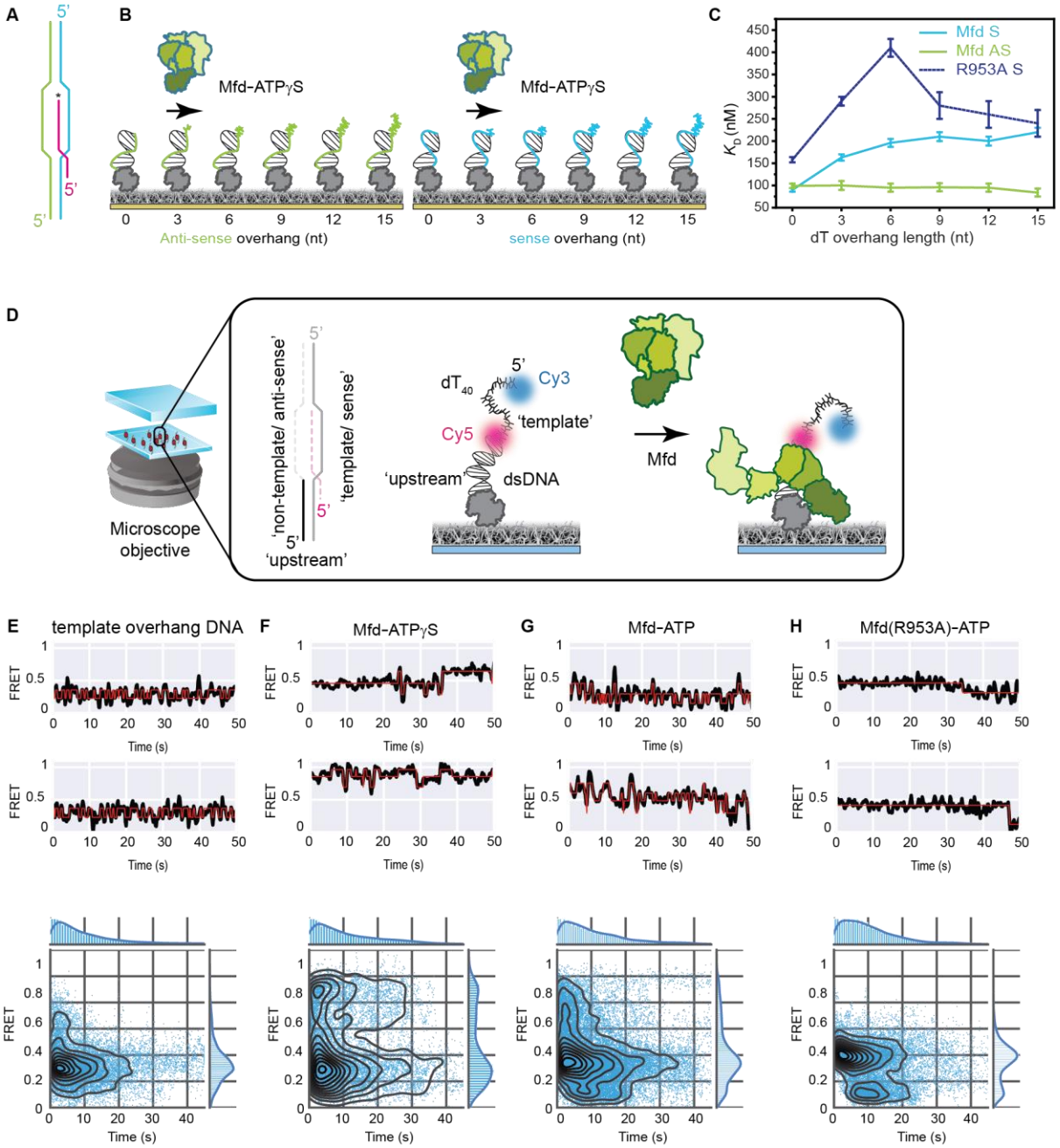
1242 0.5. See Supplemental Note 2 and Table 2.

1243 (D) Temporal heat maps of FRET pair labelled RNAP EC when titrated with ATP in the range of 0.1 μM

1244 to 1 mM (n = 151, 82, 81, 76, 157 molecules respectively) as indicated.

1245

Figure 4



1246

1247

1248 **Figure 4: Interactions between Mfd and a primed-DNA template**

1249 (A) Schematic of transcription bubble.

1250 (B) Schematic of SPR experiments to assess the strength of interaction between Mfd-ATPγS or R953A-
1251 ATPγS and primed DNA templates with 18 bp dsDNA and either anti-sense or sense strand overhangs
1252 (dT_3, dT_6, dT_9, dT_{12} or dT_{15} overhangs).

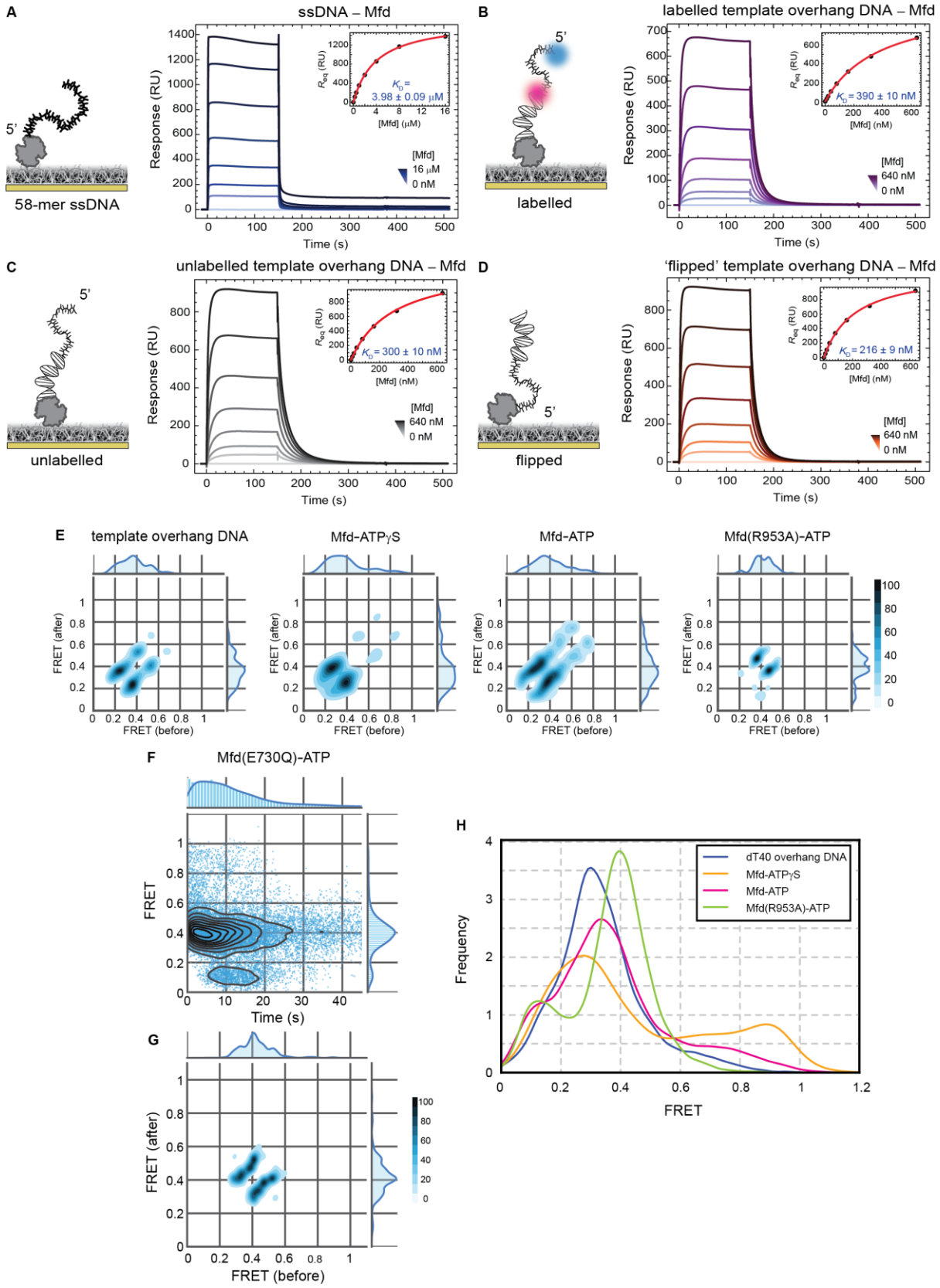
1253 (C) Plot depicting K_D of interaction as a function of overhang length for sense strand overhang (blue)
1254 or anti-sense strand overhang (green) containing DNA substrates. Error bars represent S.E.

1255 (D) Schematic of FRET-pair labelled primed DNA substrate comprised of an 18-mer dsDNA and a 5' 40-
1256 mer poly-dT overhang to assess interactions of Mfd with ssDNA.

1257 (E-H) Two example FRET trajectories (top panel) and temporal heat maps (bottom panel) are
1258 presented for (E) primed DNA alone (n = 100 molecules), (F) Mfd in the presence of ATPγS (n = 94
1259 molecules), (G) Mfd in the presence of 1 mM ATP (n = 203 molecules), and (H) Mfd(R953A) in the
1260 presence of 1 mM ATP (n = 105 molecules).

1261

Supplemental Figure 4



1262

1263

1264 **Supplemental Figure 4: Interactions between Mfd and a primed-DNA template.** ***Related to Figure 4.***

1265 Schematic and SPR sensorgrams show association (150 s) and dissociation phases of serially-diluted
1266 10 – 640 μM Mfd-ATPγS, including a 0 μM control for (A) 58-mer ssDNA containing the dT_{40} sequence,
1267 (B) FRET pair labeled primed DNA substrate (biotinylated 18-mer dsDNA with dT_{40} overhang), (C)
1268 primed DNA substrate (biotinylated 18-mer dsDNA with dT_{40} overhang) and (D) "flipped" primed DNA
1269 substrate (18-mer dsDNA with biotinylated dT_{40} overhang). Responses at equilibrium averaged over
1270 the steady region of sensorgrams (R_{eq}) were fit (insets) using a steady-state affinity (SSA) model to
1271 derive values of K_D (as indicated). Errors are S.E. of the fit.

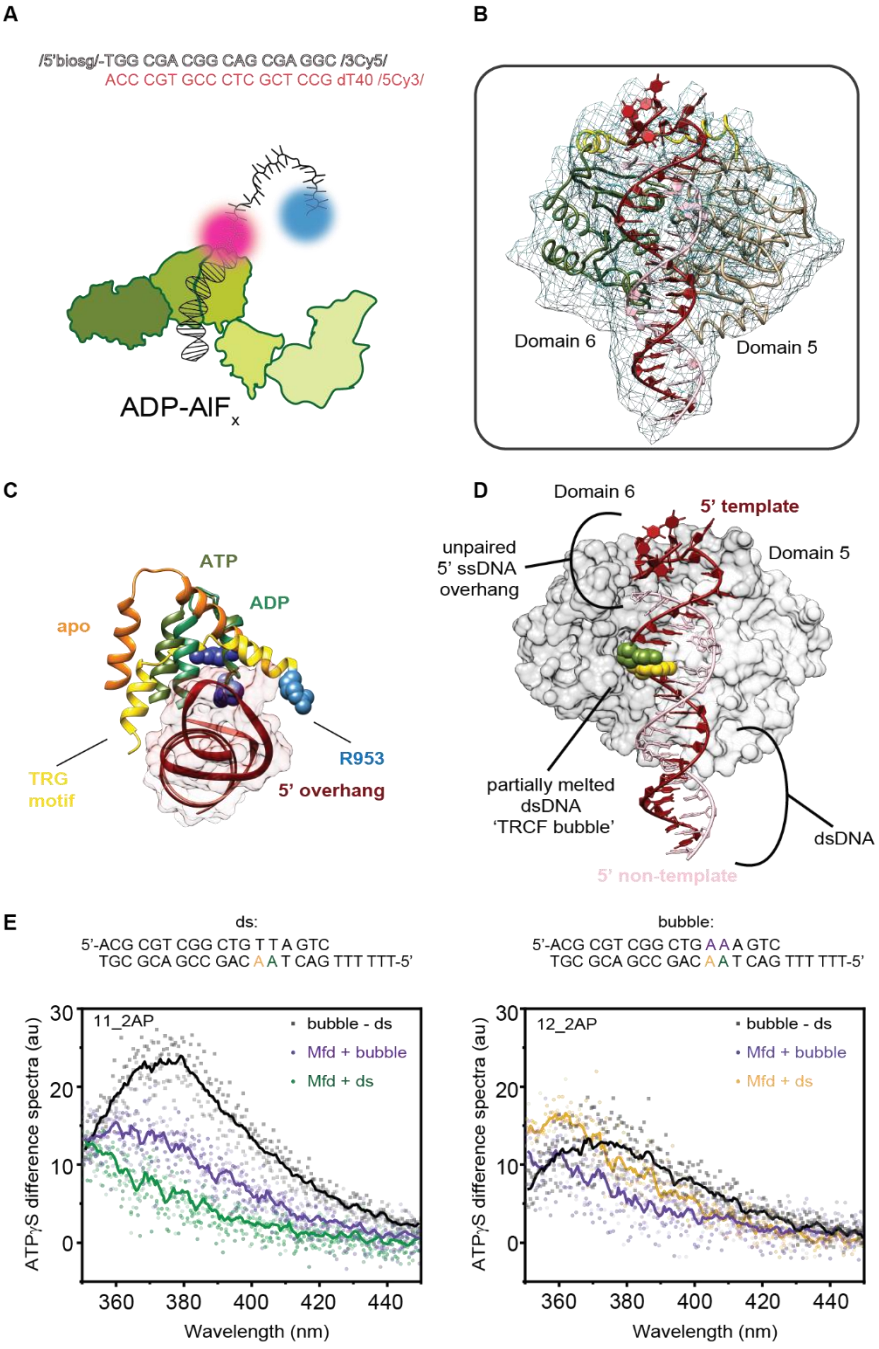
1272 (E) Transition density plots for FRET-pair labeled primed DNA alone (n = 100 molecules), or bound to
1273 Mfd in the presence of ATPγS (n = 94 molecules), or ATP (n = 203 molecules), and Mfd(R953A) in the
1274 presence of ATP (n = 105 molecules) are presented here. The DNA substrate is the same as described
1275 in Figure 4.

1276 (F, G) Temporal heat map (F) and transition density plot (G) for ATPase mutant Mfd(E730Q) binding
1277 to FRET pair labelled primed DNA template used in Figure 4 (n = 103 molecules).

1278 (H) For convenience, kernel density estimations of the ensemble FRET distributions read as projections
1279 along the ordinate in the heat-maps presented in Figure 4 are presented to enable direct visual
1280 comparisons.

1281

Figure 5



1282

1283

1284 **Figure 5: Interactions between Mfd and primed DNA**

1285 (A) Sequence of primed DNA substrate and schematic of complex subjected to cryo-EM.

1286 (B) Front view of the molecular model derived from the cryo-EM reconstruction showing Mfd (residues
1287 575-956) including domains 5 (tan) and, 6 (olive), sense (dark red) and anti-sense (pink) DNA. DNA
1288 (template: residues 28–58; non-template: 1–18) are shown in ribbon representation to illustrate the
1289 path of the DNA.

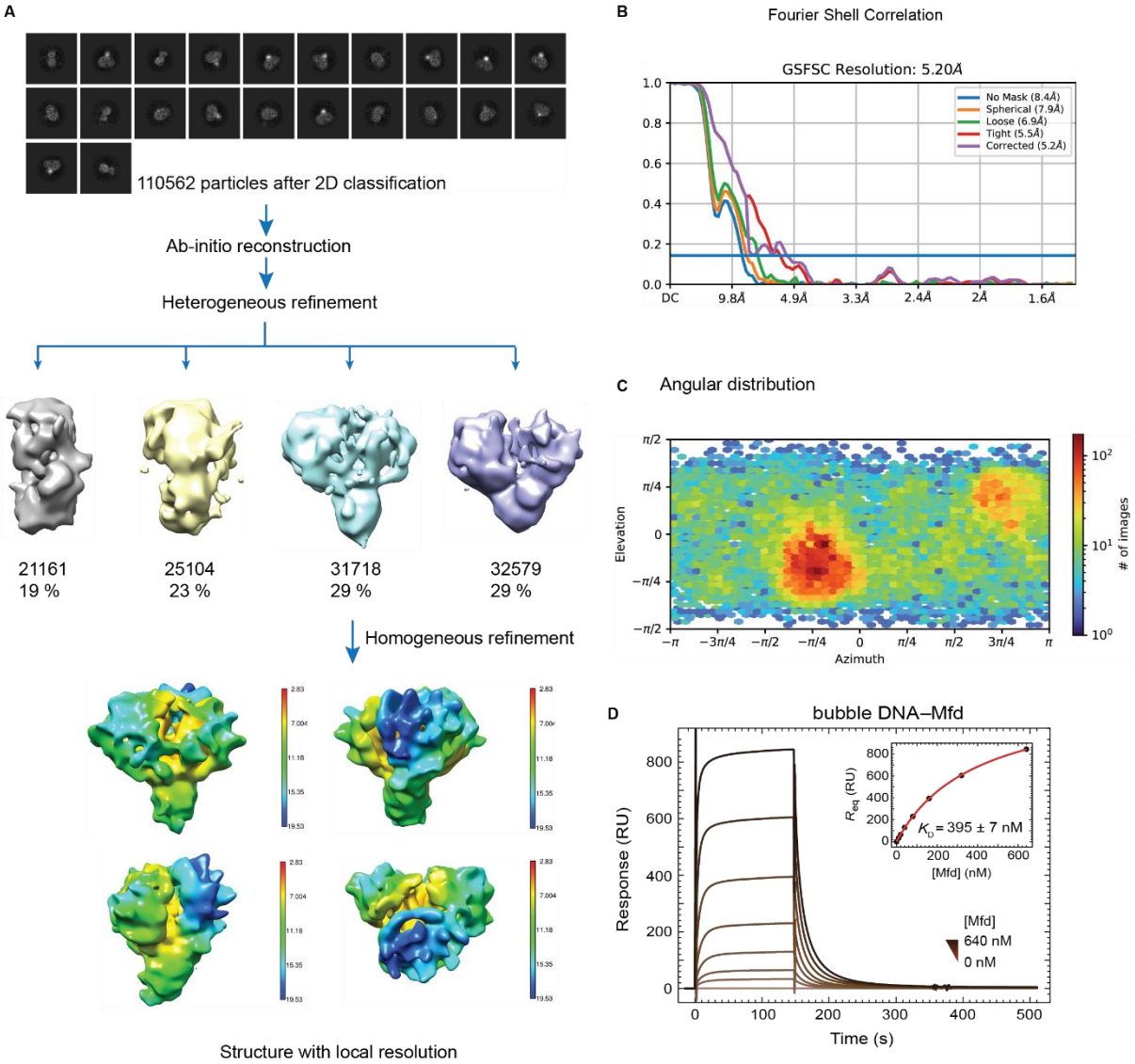
1290 (C) Comparisons of the conformations of the TRG motif in apo (orange; PDB 2EQY); ADP-bound (green,
1291 PDB 6X4W); ATP-bound (olive; PDB 6X43) and on primed-DNA (yellow, ADP.AlFx) in structures aligned
1292 to the motor domain in PDB 6X43. R953 is shown as blue spheres. The primed DNA substrate is shown
1293 in ribbon and surface representation.

1294 (D) The model suggests dsDNA near the junction is melted to form the TRCF bubble. Position 11 from
1295 the 5' end of the template is shown in green, and position 12 is shown in yellow on the sense strand.

1296 (E) Difference spectra are shown for the fluorescence of 2AP in bubble substrates compared to the
1297 respective ds substrates in the absence of Mfd or ATPγS (black). Difference spectra showing increase
1298 in fluorescence upon addition of ATPγS to a sample containing Mfd and either the bubble substrate
1299 (purple) or ds substrate (panels show data for substrate containing 2AP at position 11(left; green in
1300 panel D) or 12 (right; yellow in panel D)). Lines indicate five-point moving averages, and individual
1301 squares/circles indicate data from 2-3 independent repeats.

1302

Supplemental Figure 5



1303

1304

1305 **Supplemental Figure 5: EM reconstruction of Mfd.** ***Related to Figure 5.***

1306 (A) Overview of image processing using CryoSPARC. A set of 110562 particles were selected after 2D
1307 classification for *ab initio* reconstruction and heterogeneous refinement leading to four classes of 3D
1308 reconstructions. Of these, 31718 particles that yielded the most complete reconstruction were
1309 subjected to homogenous refinement to generate the map presented here. The colormap applied to
1310 the reconstruction represents local resolution.

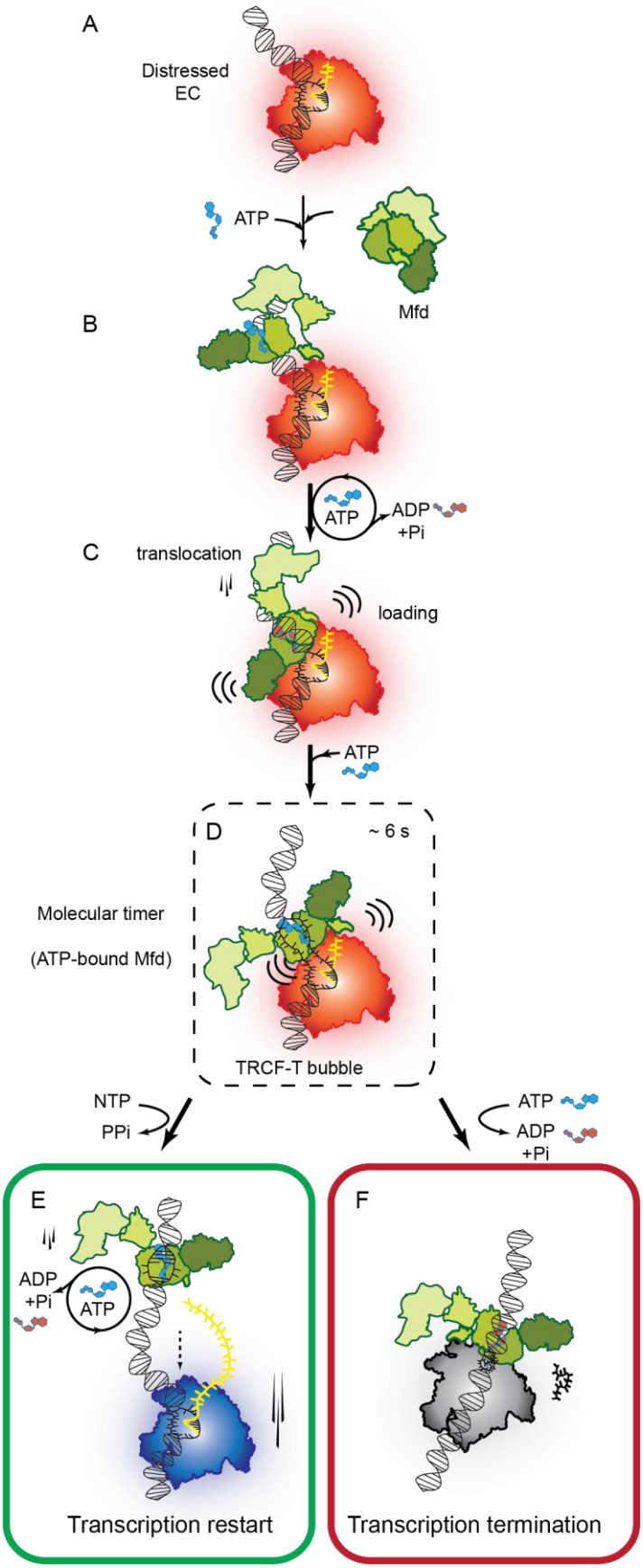
1311 (B) Fourier shell correlation curves of the reconstruction reported an apparent resolution of 5.2 Å
1312 (gold-standard 0.143 FSC).

1313 (C) Angular distribution of particle projections

1314 (D) Measurement of the dissociation constant of Mfd for a 49-mer dsDNA template containing a 9
1315 nucleotide non-complementary bubble where the complementary residues were replaced by dT_9. The
1316 errors reflect S.E of the fit.

1317

Figure 6



1318

1319 **Figure 6: Model for Mfd-mediated transcription modulation**

1320 Mfd interacts with the (A) elongation complex weakly in the absence of ATP, however, (B, C) ATP
1321 binding and hydrolysis are required for stable association of Mfd with the EC. Following, a brief period
1322 of ATP-hydrolysis dependent translocation of Mfd on upstream dsDNA, Mfd loads on to the EC. This
1323 results in chaperoning of the EC from catalytically inactive state(s) into a catalytically poised molecular
1324 timer intermediate (D) with a lifetime of ~6 s. In this ATP-bound state, we propose that Mfd melts the
1325 upstream DNA to form an extended transcription-repair coupling factor – transcription ("TRCF-T")
1326 bubble. In this state R953 contacts the template ssDNA in the TRCF-T bubble. Melting of this upstream
1327 DNA is sufficient to promote residence of the EC in a catalytically poised state. (E) In the presence of
1328 NTPs on an undamaged template, transcription can be resumed. (F) On damaged templates or under
1329 conditions of nucleotide starvation or in the presence of impediments to the progress of the EC, the
1330 bound ATP is hydrolyzed to promote transcription termination. In our experiments this is seen as the
1331 loss of RNA. We propose that in either case continued translocation occurs in a manner in which the
1332 ATPase activity of Mfd translocates the TRCF bubble forward until Mfd interacts with its binding
1333 partners in the repair machinery.

1334

1335 # Supplemental Information

1336

1337 **Mechanism of transcription modulation by the transcription-repair coupling factor**

1338 Bishnu Paudel[1,2], Zhi-Qiang Xu[1,2], Slobodan Jergic[1,2], Aaron J Oakley[1,2], Nischal Sharma[1,2],

1339 Simon HJ Brown[1,2], James C Bouwer[1,2], Peter J Lewis[1,3], Nicholas E Dixon[1,2], Antoine M van

1340 Oijen[1,2], Harshad Ghodke[1,2]*

1341 [1]Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong,

1342 Wollongong, Australia

1343 [2] Illawarra Health and Medical Research Institute, Wollongong, Australia

1344 [3]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia

1345 **Correspondence:**

1346 Harshad Ghodke, Molecular Horizons and School of Chemistry and Molecular Bioscience, University

1347 of Wollongong, Wollongong, New South Wales, 2522, Australia

1348

1349 **Supplemental Note 1**

1350 To assess whether the calculated distances are consistent with the FRET state assignments for the pre-
1351 vs. post- translocated states, we built an *in silico* model (Supplemental Figure 1B) based on available
1352 structures of the *E. coli* elongation complex (Abdelkareem et al., 2019) and structures of the Cy3 and
1353 Cy5 dyes (Liu and Lilley, 2017). The distances between the Cy3 and Cy5 dyes in the pre- and post-
1354 translocated states were then estimated to be 1.7 nm and 3.9 nm respectively. Using these distances,
1355 and available estimates for R_0 for Cy3–Cy5 dye pairs introduced into protein-DNA complexes in the
1356 range of 4.7 – 6.5 nm (Levitus and Ranjit, 2011), we estimate FRET efficiencies for the pre- and post-
1357 translocated states to be: 0.99 and 0.75 (for R_0 = 4.7 nm) and 0.99 and 0.95 (for R_0 = 6.5 nm). The
1358 measured FRET efficiencies are ~0.9 and ~0.6 (Figure 1), which agree well with the lower estimates
1359 for R_0.

1360 **Supplemental Note 2**

1361 To measure the lifetime of the intermediate state we first chose to analyze trajectories with
1362 intermediates in the mid-FRET regime. However, examination of individual trajectories revealed that
1363 this mid-FRET regime is populated by molecules exhibiting high-mid FRET state dynamics as well as
1364 mid-low FRET dynamics. We therefore selected molecules in the 0.35–0.5 FRET range which mostly
1365 accommodated molecules exhibiting mid-low FRET dynamics. The mean lifetimes (and standard error)
1366 of the intermediates states were best estimated by fitting the distributions to single-exponential fits.
1367 In the case of one notable condition – Mfd in the presence of ATPγS followed by incubation with
1368 correct nucleotides – the lifetime distributions was found to be better fit by a gamma distribution. In
1369 this case, the scale parameter was used to report the lifetime (Table 2).

1370 **Supplemental Note 3**

1371 To compare the efficiency of transcription by the elongation complex following incubation with Mfd-
1372 ATPγS upon addition of 100 μM each of GTP and UTP to that in the presence of GTP, UTP and ATP, we
1373 plotted histograms of the FRET states observed in the system (gray bars in Figure 3L). The histogram
1374 for each condition was best fit with a sum of three Gaussian terms of the form:

1375 $$f(x) = ae^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1376 Here a_i represents the amplitude, μ_i represents the mean of the distribution and σ_i the standard
1377 deviation of the ith distribution. In the absence of Mfd-ATPγS, the three Gaussians (a_i, μ_i, σ_i) were:
1378 (25.36, 0.91, 0.11); (16.5, 0.63, 0.33) and (15.14, 0.09, 0.10) with an R-square of 0.88. In the presence
1379 of Mfd-ATPγS and rGTP, rUTP, the three Gaussians were: (17.15, 0.93, 0.19), (10.94, 0.70, 0.27) and
1380 (9.82, 0.18, 0.10) with an R-square of 0.85. For each Gaussian the area was calculated as:

1381 $$A_i = \sqrt{2\pi} a_i \sigma_i$$

1382 Next, each area was normalized by the sum of the areas of the three Gaussians to obtain the fractional
1383 area occupied by each Gaussian. Efficiencies were calculated by comparing the fractions of the
1384 populations corresponding to each of the Gaussians across the two conditions. Since transcript
1385 elongation is reflected in the low-FRET states, we compared the fraction of the population in the
1386 Gaussian with the smallest mean FRET.

1387 **Supplemental Note 4**

1388 SPR measurements of the strength of interaction between Mfd and this DNA substrate in the presence
1389 of ATPγS revealed a dissociation constant that is somewhat higher (K_D = 300 ± 10 nM) than that of
1390 dsDNA (170 ± 20 nM; Supplemental Figure 4B–D compared to Supplemental Figure 2C). To monitor
1391 changes in the distance between the junction and the 5' end, we introduced a FRET pair on either end
1392 of the ssDNA overhang (Figure 4D). SPR measurements revealed an affinity of 390 ± 10 nM for the
1393 binding of Mfd to the FRET-pair labelled substrate (Supplemental Figure 4B).

1394

1395 **Table 1: Nucleic Acid substrates 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
SPR (Figure 4)	S18	GCC TCG CTG CCG TCG CCA
SPR	dT3_S18	TTT GCC TCG CTG CCG TCG CCA

(Figure 4)		
SPR (Figure 4)	dT6_S18	TTT TTT GCC TCG CTG CCG TCG CCA
SPR (Figure 4)	dT9_S18	TTT TTT TTT GCC TCG CTG CCG TCG CCA
SPR (Figure 4)	dT12_S18	TTT TTT TTT TTT GCC TCG CTG CCG TCG CCA
SPR (Figure 4)	dT15_S18	TTT TTT TTT TTT TTT GCC TCG CTG CCG TCG CCA
SPR (Figure 4)	S18_bio	GCC TCG CTG CCG TCG CCA /3Bio/
SPR (Figure 4)	AS18_dT3	TGG CGA CGG CAG CGA GGC TTT
SPR (Figure 4)	AS18_dT6	TGG CGA CGG CAG CGA GGC TTT TTT
SPR (Figure 4)	AS18_dT9	TGG CGA CGG CAG CGA GGC TTT TTT TTT
SPR (Figure 4)	AS18_dT12	TGG CGA CGG CAG CGA GGC TTT TTT TTT TTT
SPR (Figure 4)	AS18_dT15	TGG CGA CGG CAG CGA GGC TTT TTT TTT TTT TTT
2AP Bulk fluorescence	dT6_S18_v 2_11_2AP	TTT TTT GAC T/i2AmPr/A CAG CCG ACG CGT

(Figure 5)		
2AP Bulk fluorescence (Figure 5)	dT6_S18_v2_12_2AP	TTT TTT GAC TA/i2AmPr/ CAG CCG ACG CGT
2AP Bulk fluorescence (Figure 5)	AS18_v2_2AP	ACG CGT CGG CTG TTA GTC
2AP Bulk fluorescence (Figure 5)	AS18_v2_2AP_bubble	ACG CGT CGG CTG AAA GTC

1396

1397

1398 **Table 2: Lifetimes of mid-FRET states**

	Condition	Distribution	Fit Parameter				
			Mean (τ) [Exp]/ scale parameter (Gamma)	S.E.	Shape	S.E.	log likelihood
1	EC (n_states = 42)	exp	5.0	0.8			−109.9
2	EC dNTPs (n_states = 14)	exp	5.7	1.5			−38.4
3	EC wrong (n_states = 70	exp	7.8	0.9			−213.7
4	EC right (n_states = 19)	exp	5.5	1.3			−51.6
5	NusG (n_states = 99)	exp	4.2	0.4			−242.1
6a	NusG wrong (n_states = 57)	exp	6.9	0.9			−166.9
7a	NusG correct (n_states = 129)	exp	5.9	0.5			−356.965
8	Mfd (n_states = 13)	exp	7.1	2.0			−38.5

9	Mfd-ATPγS (n_states = 20)	exp	6.1	1.4			−56.1
10	Mfd(R953A) (n_states = 21)	exp	7.2	1.6			−62.3
11	Mfd-ATP [10μM] (n_states = 29)	exp	4.8	0.9			−74.9
12	Mfd-ATP [100μM] (n_states = 23)	exp	6.1	1.3			−64.5
13	Mfd-ATP [1 mM] (n_states = 77)	exp	6.2	0.7			−216.9
14	Mfd-ATP [10 mM] (n_states = 27)	exp	4.6	0.9			−68.1
15	ATPgS_rG_rU (n_states = 17)	Gamma	0.7	0.2	5.20	1.73	−29.6

Python scripts for Data Analysis

<u>1. Concatenating .dat files</u>

```
import os
from natsort import natsorted
import pandas as pd
import seaborn as sns
import matplotlib as plt
import glob
path = ("")  # location of .dat files
all_files = glob.glob(path + "/*.dat")
li = []
for filename in all_files:

    df = pd.read_fwf(filename, index_col=None)
    df1 = pd.DataFrame(df)
    df1.columns = ["Time(s)", "Donor", "Acceptor", "FRET", "Idealized-FRET"]
    li.append(df1)

frame = pd.concat(li, axis=0, join = 'outer', ignore_index=False)
frame = pd.DataFrame(frame)
frame.to_csv(" ", index=None) # provide location to save .csv file
```

2. Heatmap (Time-FRET density map)

```
from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import pandas as pd
import numpy as np
from matplotlib import rc

filename = “ ”   # provide path of .csv files for heat map saved from concate scripts above
x = pd.read_csv(filename)
x.columns = ['Time (s)', 'Donor', 'Acceptor', 'FRET', 'Idealized FRET']
x = x.loc[(x['FRET']<=1) & (x['FRET']>0)]
T = x['Time (s)']/5
FRET = x['FRET']
sns.set(style = 'whitegrid', font_scale = 1.5)
sns.jointplot(x = T, y = FRET, data=x, kind="kde", fill=True, gridsize = 100, color = 'skyblue', reverse =
False, xlim = (0,160), ylim = (0, 1.1), shade_lowest = False)
plt.show()
```

3. Transition Density Plot (TDP)

```
from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import pandas as pd
import numpy as np

filename = "_dwelldata.txt"
A = pd.read_table(filename, header = None)
print(A.head(5))
A.columns = ['Molecule', 'FRET before transition', 'FRET after transition', 'Time']
print(A.head(5))
# Creating bins
x = A['FRET before transition']
y = A['FRET after transition']
z = A['Molecule']
x_min = np.min(x)
x_max = np.max(x)

plt.figure(figsize = (6, 6))
kdeplot = sns.jointplot(x, y, ratio = 6, gridsize = 100, xlim = (0, 1.2), kind = 'kde', ylim = (0, 1.2), color =
'skyblue', shade_lowest = False)

kdeplot.ax_joint.set_xticks([0, 0.2, 0.4, 0.6, 0.8, 1.0])
kdeplot.ax_joint.set_yticks([0, 0.2, 0.4, 0.6, 0.8, 1.0])

plt.show()
```

Plot Traces

```
import seaborn as sns
from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import pandas as pd
from scipy.signal import savgol_filter

import numpy as np
filename = " "  # provide path either .dat file or .csv file
x = np.loadtxt(filename)
D = x[:, 1]
A = x[:, 2]
F = x[:, 3]
IF = x[:, 4]
Time = x[:, 0]/5
# smoothing
F = savgol_filter(F, 5, 2)
A = savgol_filter(A, 5, 2)
D = savgol_filter(D, 5, 3)
```

```
#Intensity-time plot
plt.figure(figsize = (5, 2))
sns.set(style = 'darkgrid', font_scale = 1.5)
plot1 = plt.figure(1)
sns.lineplot(Time, D, color = 'blue')
sns.lineplot(Time, A, color = 'purple')
plt.xlabel("Time (s)")
plt.ylabel("Intensity (a.u.)")
plt.xlim(0, 42, 10)
plt.ylim(-200, 6000, 2000)

#FRET-time plot
plt.figure(figsize = (5, 2))
plot2 = plt.figure(2)
plt.xlim(0, 42, 10)
plt.ylim(0, 1.1, 0.2)
sns.lineplot(Time, F, color = 'black')
sns.lineplot(Time, IF, color = 'red')
plt.xlabel("Time (s)")
plt.ylabel("FRET")
plt.legend()
plt.show()
```

4. Exponential Fit

```
from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import pandas as pd
import numpy as np
from matplotlib import rc
import os
from natsort import natsorted
import glob
from scipy.optimize import curve_fit
import sympy as sym

#Load data using path for the list of files
path = src_path

all_files = glob.glob(path + "/*.dat")

li = []

# filtering data using for loop
for filename in all_files:

    df = pd.read_fwf(filename, index_col=None)
    df1 = pd.DataFrame(df)
    df1.columns = ["Time(s)", "Donor", "Acceptor", "FRET", "Idealized-FRET"]
    df2 = df1[(df1["FRET"]>0.35) & (df1["FRET"]<0.5)]

    length = len(df2)/5

    li.append(length)
# converting list inot dataframe
li1 = pd.DataFrame(li)
li2 = li1[li1[:]>1]
li3 = li2.dropna()

li3.to_csv(".csv", index=None) # provide path
# Making counts of histogram
hist = np.histogram(li3, density=True)[0]

# choosing x length to fit entire histogram
x = np.arange(0, len(hist), 1)

x_fit = x + x/2

# Defining single exponential function
def exp_fit(x, a, b):
    return a*np.exp(-b*x)

# plotting and fitting data
```

```
plt.figure(figsize=(5,4))
fit = curve_fit(exp_fit, x_fit, hist, p0=[2,0.5]) # parameter should optimize here
fit_eq = fit[0][0]*np.exp(-fit[0][1]*x_fit)
plt.bar(x_fit, hist, align="edge")
plt.plot(x_fit, fit_eq, color="red")
plt.xlabel("time (s)")
plt.ylabel("Density")
plt.xlim(0,40)

plt.show()
```

5. Violin plots

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("midFRET_kinetics.csv")
sns.set_style("whitegrid")
plt.figure(figsize=(5,5))
sns.violinplot(data=df, cut=0)
plt.xticks(rotation=45, ha="right")
plt.ylabel("Time (s)")

plt.savefig(save_path, dpi=600)
plt.savefig(save_path, dpi=600)
```

```
6: Trajectory sort
from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from scipy.stats import expon
import math
import pandas as pd
import numpy as np
from matplotlib import rc
import os
import glob
from scipy.optimize import curve_fit
import sympy as sym
from shutil import copyfile

# dataset_name = "500nM_NusG_alone"
# dataset_name = "500nM_NusG_all_correct_NTPs"
# dataset_name = "500nM_NusG_wrong_NTPs"

# dataset_name = "RNAP_alone"
# dataset_name = "RNAP_dNTPs"
# dataset_name = "RNAP_wrong_NTPs"
# dataset_name = "RNAP_all_NTPs"

# dataset_name =  "500nM_Mfd"
# dataset_name = "Mfd_ATPgS"
# dataset_name = "Mfd_10uMATP"
# dataset_name = "Mfd_100uMATP"
# dataset_name = "Mfd_1mMATP"
# dataset_name = "Mfd_10mMATP"
dataset_name = "RNAP_Mfd_ATPgS_rG_rU"
# dataset_name = "MfdR953A_1mMATP"



#define range for FRET values
a = 0.4
b = 0.7
spf = 0.2          #seconds per frame
#Load data using path for the list of files
src_path = **source path**
#Save file path
save_path_static = os.path.join(src_path, 'static')

if not os.path.isdir(save_path_static):
    os.mkdir(save_path_static)

save_path_dynamic = os.path.join(src_path, 'dynamic')
if not os.path.isdir(save_path_dynamic):
    os.mkdir(save_path_dynamic)
```

```
save_path_dynamic_high = os.path.join(save_path_dynamic, 'dynamic_high')
if not os.path.isdir(save_path_dynamic_high):
    os.mkdir(save_path_dynamic_high)

save_path_dynamic_mid = os.path.join(save_path_dynamic, 'dynamic_mid')
if not os.path.isdir(save_path_dynamic_mid):
    os.mkdir(save_path_dynamic_mid)

save_path_dynamic_low = os.path.join(save_path_dynamic, 'dynamic_low')
if not os.path.isdir(save_path_dynamic_low):
    os.mkdir(save_path_dynamic_low)

save_path_dynamic_high_mid = os.path.join(save_path_dynamic, 'dynamic_high_mid')
if not os.path.isdir(save_path_dynamic_high_mid):
    os.mkdir(save_path_dynamic_high_mid)

save_path_dynamic_high_low = os.path.join(save_path_dynamic, 'dynamic_high_low')
if not os.path.isdir(save_path_dynamic_high_low):
    os.mkdir(save_path_dynamic_high_low)

save_path_dynamic_mid_low = os.path.join(save_path_dynamic, 'dynamic_mid_low')
if not os.path.isdir(save_path_dynamic_mid_low):
    os.mkdir(save_path_dynamic_mid_low)


#Open files
all_files = glob.glob(src_path + "/*.dat")

li_states_all = []
li_lifetimes_all = []
li_dFRET_all = []

# filtering data using for loop
for filename in all_files:
# for filename in all_files[29:30]:
        # print(filename)
    # if os.path.basename(filename) == '40.dat':
        with open(filename, 'r') as file:
            data = np.loadtxt(filename)                                #["Time(s)", "Donor", "Acceptor", "FRET",
"Idealized-FRET"]
        # print("Trajectory data")
        # print(data)
        #Sort trajectories
        #does trajectory have photobleaching step included? Classify trajectory as 'static' if only one other
state exists, and it survives for less-equal 5 frames
        unique_states, counts = np.unique(data[:,4], return_counts = True)          #this seems to return
unique_states in ascending order[!]
        # print("Unique states are: "+str(unique_states))
        # print(counts)
        if len(unique_states) == 1:                                    #do nothing in case of trajectories with a
```

```
single step
        print("static - do nothing")
    elif (len(unique_states) == 2) and (counts[0]<5):                    #do nothing in case of trajectories
with a single step
        print("static_pb - do nothing")
    else:
        print("analyse dynamic trajectory")
        for state in unique_states:
            # print(counts[np.where(unique_states == state)[0]])
            if counts[np.where(unique_states == state)[0]]>1:                #apply the requirement that a
state has to last at least 2 frames
                data_state = data[data[:,4]==state]
                # print(data_state)
                # print(state)
                # print("-----------------------------------------------------------------")
                frames = data_state[:,0]                              #extract timeprints for frames
                states = data_state[:,4]

                frames_1 = np.zeros(frames.shape)                     #Create a new array shifted by 1
                frames_1[1:] = frames[:-1]
                # print(frames)
                # print(frames_1)
                del_frames = frames-frames_1
                # print(del_frames)

                change_pts = np.where(del_frames>1)
                # print("Change_pts: " + str(change_pts))
                ch_pts = np.asarray(change_pts)
                ch_pts_1 = np.zeros(ch_pts.shape)
                ch_pts_1[0][:-1] = ch_pts[0][1:]
                # print(ch_pts)
                # print(ch_pts_1)
                lifetimes = ch_pts_1[0][:-1] - ch_pts[0][:-1]                #this is a tuple corresponding to
number of instances of the intermediate being formed
                # print("lifetimes are: " + str(lifetimes))

                # Identify whether the next state has higher or lower FRET
                li_dFRET = [5]*len(lifetimes)
                index = 0
                for val in range(1, len(change_pts[0])):
                    index = int(frames[change_pts[0][val]-1]-1)
                    index_1 = index+1
                    dFRET = data[index_1][4]-data[index][4]
                    if dFRET>0:
                        li_dFRET[val-1] = 1
                    else:
                        li_dFRET[val-1] = 0


                # # Create lists of entries
                li_lifetimes = lifetimes.tolist()
```

```
            li_state = [state]*len(lifetimes)

            # Extend master list by adding entries of states
            li_states_all.extend(li_state)
            li_lifetimes_all.extend(li_lifetimes)
            li_dFRET_all.extend(li_dFRET)
            # print("Next intermediate")
            # print("-----------------------------------------------------------------")

df = []
df = pd.DataFrame()
df['States'] = li_states_all
df['lifetimes'] = li_lifetimes_all
df['lifetimes'] = spf*df['lifetimes']
df['dFRET'] = li_dFRET_all
df['in_range'] = (df['States']>0.4)&(df['States']<0.7)
df = df.sort_values('dFRET')
df = df[df['lifetimes']>spf]
# df.to_csv(os.path.join(src_path, 'all_state_lifetimes.csv'), index=None)
# print(df.head)
sns.displot(df, x = "lifetimes", hue = "in_range", kind = "hist")
ax2 = plt.twinx()
sns.ecdfplot(df, x = "lifetimes", hue = "in_range", ax = ax2)
plt.xlim(0, 10)

# sns.displot(df, x = "lifetimes", hue = "dFRET", kind = "hist")
# ax2 = plt.twinx()
# sns.ecdfplot(df, x = "lifetimes", hue = "dFRET", ax = ax2)
# plt.xlim(0, 10)

plt.title(dataset_name)
```