The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation

[1,2]Yuri A. Nedialkov, [1]Kristopher Opron, [1]Fadi Assaf, [3]Irina Artsimovitch, [4]Maria L. Kireeva, [4]Mikhail Kashlev, [5]Robert I. Cukier, [2]Evgeny Nudler and [1]Zachary F. Burton

[1]Department of Biochemistry and Molecular Biology, Michigan State University, E. Lansing, MI 48824-1319
[2]Department of Biochemistry, New York University Medical Center, New York, NY 10016, USA
[3]Department of Microbiology; The Ohio State University; Columbus, Ohio USA
[4]Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute, Frederick, MD 21702-1201
[5]Department of Chemistry, Michigan State University, E. Lansing, MI 48824-1319

Supplementary Text

F773V is not necessarily a "fast" RNAP

In Figure S1, wild type, F773V and Y772A RNAPs are compared for bulk elongation at 200 and 1 μM NTPs. Elongation was from G8 or G7, as indicated. TECs were assembled in vitro with G7 (5'-^{32}P-CGAGAGG) and walked to G8 and extended (left panel) or extended from G7 (right panel). F773V is not a purely "fast" RNAP. Compared to wild type and Y772A RNAPs, F773V pauses more strongly at some template positions (i.e. C22) and less strongly at others (i.e. A18, A24) [1].

Termination at bacteriophage λ intrinsic terminator tR2

To distinguish individual gel bands and to correlate bands with the RNA sequence, a long polyacrylamide gel was done of wild type Ec RNAP transcriptional termination or read through at λ tR2 (Figure S2A) [2, 3]. The experiment was done as in Figure 3, except using transcription buffer containing 40 mM rather than 150 mM KCl. At 40 mM KCl, termination efficiency is higher for wild type RNAP at 200 μM NTPs (~38 % efficiency versus ~29 % efficiency at 150 mM KCl (Figure 3)). In Figure S2B, the expected RNA structures of λ tR2 and *hisT* intrinsic terminators are shown. The *hisT* terminator can form a longer RNA stem (14 bp compared to 8 bp for λ tR2) with a smaller loop (AUCAU compared to UAGUAAUC) and has a long U stretch lacking interruptions (UUUUUUUUU compared to UUUUUAUUU) causing *hisT* to be a stronger intrinsic terminator than λ tR2 [4, 5].

In Figure S3, wild type and I774A RNAP are compared for termination at λ tR2. I774A is less efficient than wild type RNAP at termination at 200 μM NTPs but efficient at 1 μM NTPs.

Downstream border exonuclease III mapping

To demonstrate the generality of Ec RNAP downstream border mapping with exonuclease III (Exo III), an experiment with RNAP stalled at position A39 is shown (Figure S4; see also Figure 8). Ec RNAP-σ^{70} holoenzyme was initiated from the bacteriophage T7 A1 promoter with ApUpC trinucleotide, GTP and ATP. TECs were stalled at G10 and adsorbed onto Ni-NTA beads (Supplementary Methods). Immobilized G10 TECs were washed and then radiolabeled by extension to A32 with α^{32}P-CTP, GTP and ATP. Washed A32 TECs were walked to G35 and washed again. G35 TECs were treated with *Eco*

RV endonuclease to remove ^{32}P radiolabel from the DNA template strand. After washing, TECs were released from Ni^{2+} beads with imidizole. ATP, CTP and Exo III were added together to extend the TEC to A39 and to map the A39 downstream border.

β' F773V is similar to wild type and Y772A RNAPs in the absence of GMP for resting translocation state and backtracking at A39. In the presence of GMP, however, F773V is shifted to the hyper- and post-translocated states and strongly resists backtracking. For F773V, therefore, GMP strongly stimulates forward translocation and inhibits backtracking.

Molecular dynamics simulations

All atom molecular dynamics simulations of wild type Tt RNAP TECs with closed (PDB 2O5J) or open (PDB 2PPB) trigger loop conformations [6, 7] were used to probe the functional contacts of Ec RNAP YFI residues β' Y772 (*Thermus thermophilus* (Tt) RNAP β' Y1070) and F773 (Tt RNAP F1071) [8, 9]. A catalytic TEC with loaded ATP substrate was analyzed because the significance of YFI residues is likely to be enhanced in the catalytic structure. The open TEC structure was analyzed as a reasonable comparison to the closed TEC structure and to observe switching involving the RNAP link domain [8]. Simulations were analyzed to generate hypotheses to help explain the activities of the Ec RNAP Y772A and F773V substitutions. Molecular dynamics simulations of Tt RNAP Y1070A and F1071V (the homologous residues) are currently in progress. The bridge helix Ec β' 772-YFI-774 and the nearby glycine hinge (β' 778-GARKG-782) couple the motions of the bridge helix, the link domain, the extended fork and the F-loop. The link domain interacts with the active site and the NTP substrate (i.e. through link residue β R678), so bridge helix dynamics is expected to be coupled to RNAP catalysis.

Proposed contacts and functions of Y772

In the catalytic TEC, Ec β' Y772 is proposed to closely approach and to potentially interact with β' S776 (bridge helix), β D674 (main chain hydrogen bond), V668, F669, D675, A676, N677, R678, L680 (β link domain), β F1187 and β' R744 (F loop; ionic interaction with β D675 (link domain)) (Figure S5). In Figure S5, the numbering is for (Tt) RNAP because the simulation was done with the Tt RNAP closed trigger loop TEC structure (PDB 2O5J) [7]. Multiple-sequence alignments are shown in the figure to convert between Tt and Ec numbering and to compare the conservation of residues across species. The Ec β' Y772A substitution is proposed to limit contact to the link domain, changing the interaction of R678 with the triphosphate tail of a NTP substrate in the RNAP active site, so Y772 is expected to couple bridge helix bending and motion to the active site [10]. The Y772A substitution eliminates hydrogen bonding between Y772 and the main chain oxygen of β D674, breaking a key interaction coupling YFI and the bridge helix glycine hinge to the active site through the link domain. Bridge helix to active site communication is braced by ionic interaction between β' R744 (base of the F loop) and β D675 (link domain). We posit that bulky hydrophobic residues β' 772-YFI-774 couple bending at the adjacent β' 778-GARKGL-783 glycine hinge on the bridge helix to the link domain (contacted by Y772), the extended fork region (contacted by F773) and the F loop (contacted by I774), causing YFI to move with surrounding protein including the active site. Substitution of Y772 to alanine, therefore, is expected to reduce linkage between the bridge helix and the link domain uncoupling one pathway of connection between YFI and the active site.

According to simulations of the open trigger loop Tt RNAP TEC (PDB 2PPB, with streptolydigin removed for simulation) [7], the link domain residue Tt β R557 (Ec R678) may switch its contacts from the NTP substrate in the active site to participate in pyrophosphate release (Figure S6). Specifically, in the open trigger loop TEC after simulation, Tt RNAP link domain β R557 (Ec R678) switches its contact from the γ-phosphate of the substrate ATP to form an ion pair to β D686, which in the original 2PPB crystal structure, in the closed trigger loop TEC 2O5J structure and in 2O5J simulations contacts Mg-II. β D686 and E685 comprise the invariant β 685-ED-686 motif that interacts with Mg-II in multi-subunit RNAPs. We posit that during elongation, β R557 (Ec β R678) of the link domain, which is coupled in its motion to YFI through β' Y1070 (Ec β' Y772), functions in catalysis, in the closed trigger loop TEC. In the open trigger loop TEC, β R557 (Ec β R678) of the link domain may have a role in pyrophosphate release and translocation. The molecular dynamics modeling done here did not allow for the breaking of covalent bonds, so, during simulation, pyrophosphate was unable to dissociate from ATP to maintain contact with Tt β R557.

Proposed contacts and functions of F773

F773 is a bulky hydrophobic residue that makes most of its atomic contacts to other residues on the bridge helix (Ec β' V769, I774 and H777) and the extended fork (β V550, H554, Y555 and P560) (Figure S7). The F773V substitution is expected to reduce these contacts altering how bending at the nearby β' 778-GARKGL-783 glycine hinge on the bridge helix affects the conformation of the fork and the F loop, with which the β' 778-GARKGL-783 glycine hinge interacts. Mutation of F773, therefore, is expected to alter bridge helix dynamics and bending and also the conformation of surrounding protein with some longer range effects (i.e. up to about 20 Å from F773).

Supplementary methods

Exonuclease III mapping

Templates for Exo III RNAP downstream border footprinting were based on a PCR fragment containing the bacteriophage T7 A1 promoter with a modified transcribed sequence downstream. The DNA template was prepared by PCR amplification, purified, and end-labeled with polynucleotide kinase and [γ-^{32}P]-ATP (MP Biomedicals). TECs were assembled by addition of RNAP holoenzyme, template DNA and NTP mix (10 μM ApUpC (Integrated DNA Technologies), 50 μM GTP and ATP (GE Healthcare)) and stalled at G10. G10 TECs were captured on Ni-NTA agarose beads. After washing beads 3X, 50 μM ATP and GTP and 80 μCi α-^{32}P-CTP were added to radiolabel the RNA and to extend to A32. Stalled A32 TECs were washed and treated with heparin to remove unbound RNAP. A32 TECs were treated with *Eco*RV endonuclease to remove the final 9 bp at the downstream end of the DNA template, thereby removing the radioactive label from the DNA template strand, leaving only the non-template strand labeled at the 5' end. The TECs were then walked to position G35 with 5 μM UTP, CTP and GTP, washed 5X, and eluted from the beads with 100 mM imidazole in 20 μl with 1 % bovine serum albumin. G35 TECs were then diluted to 300 μl Exo III reaction buffer and divided into separate reactions (10 μl). Finally, ExoIII was added for the indicated amount of time in combination with 5 μM ATP and CTP to walk to position A39. Reactions were quenched with urea buffer (8M Urea, 20 mM EDTA, 1X TBE, 0.5 % Brilliant Blue R and 0.5 % Xylene Cyanol FF) and Exo III digestion samples were resolved

on 6 % polyacrylamide-7M urea gels and RNA samples were resolved on 20 % polyacrylamide-7M urea gels.

Tt RNAP TEC simulations

All atom molecular dynamics simulations of Tt RNAP closed (PDB 2O5J with some missing chain taken from PDB 2O5I) and open trigger loop TECs (PDB 2PPB) in explicit water with counterions were done as previously described [8, 9]. Simulations were done with all RNAP TEC subunits included ($\alpha_2\beta\beta'\omega$). For simulation, AMPcPP (α-, β-methylene ATP) in the original crystal structures was substituted with ATP**.** Because we were not certain how to predict the states of an important histidine on the trigger loop, Tt RNAP β' H1242 was not protonated during simulations [11, 12]. Thio-substitutions in nucleic acids used to block RNA cleavage by Tt RNAP in crystals were not present in the protein data bank files and were not considered in modeling [6, 7]. For the closed-Mg simulation, to maintain catalytic geometry during simulation, extra bonds were assigned according to reference [7]. For the closed-Mg simulation, Mg-I was covalently bonded to the RNA 3'O, β' D739 OD1, β' D741 OD1, β' D743 OD2 and the α-phosphate O. Mg-II was covalently bonded to α-, β- and γ-phosphate oxygens and to β' D739 O2. Bond lengths were taken from the original 2O5J crystal structure and were maintained with a force constant of 200 kcal/mol/Å^2. For the open TEC simulation (lacking restraints), extra bonds to Mg atoms were omitted. The expectation was that, as observed in simulations, the closed-Mg TEC would maintain reasonable catalytic geometry close to the original crystal structure and that the open TEC might diverge from catalytic geometry [8].

Simulations were done on the NSF XSEDE (Extreme Science and Engineering Discovery Environment) national computer system using the University of Texas, Austin, Sun Constellation Linux Cluster TACC (Texas Advanced Computer Center) Ranger computer. To parallelize large computer jobs, simulations were done using the program NAMD (Not (just) Another Molecular Dynamics Program) [13] using AMBER (Assisted Model Building with Energy Refinement) [14] force fields.

Equilibration consists of 500–1000 steps of minimization followed by instantaneous heating to 320 K and a gradual release of harmonic restraints from 5 to zero kcal/mol/Å^2 over 500 ns. This is followed by 10.5 ns of production simulation run at 2 fs/step with the SETTLE algorithm and TIP3P explicit water. A Langevin thermostat and barostat were used to maintain 320 K and 1 atm. Simulations were run using NAMD version 2.8 and with Amber ff10 force fields from Amber 11. Non-bonded (electrostatic and Van der Waals) forces were calculated out to 10 Å with switching at 8 Å for smoothing the cutoff. For the calculation of electrostatic forces, Particle Mesh Ewald summation was used with periodic boundary conditions [13, 14].

Molecular images were prepared using the program Visual Molecular Dynamics (VMD) [15]. Representative PDB files from simulations will be made available on request to ZFB (burton@cns.msu.edu).

Figure legends

Figure S1. F773V RNAP has altered pause selection. Transcription elongation for Y772A, F773V and wild type RNAPs is shown from G8 (200 μM NTPs; panel A) or G7 (1 μM NTPs; panel B).

Figure S2. λ tR2 termination by wild type Ec RNAP. A) A long polyacrylamide gel was run to resolve RNA bands for the T7A1 promoter-λ tR2 terminator template, as in Figure 3. Transcription buffer

contained 40 mM KCl. B) Predicted RNA secondary structures for the λ tR2 and *hisT* intrinsic terminators are compared.

Figure S3. RNA synthesis by wild type and β' I774A Ec RNAP. Elongation was from A32 using the T7A1 promoter-λ tR2 terminator template, as in Figure 3. Transcription buffer contained 150 mM KCl.

Figure S4. RNAP A39 TEC downstream border Exo III mapping indicating the slow post→pre and pre→backtracked transitions. A) Promoter-initiated G35 Ec RNAP TECs were extended to A39 and mapped with Exo III. B) Exo III mapping (15 s incubation at 25 C) and RNA gel data. 1 mM GMP was added to the indicated reactions as a non-incorporatable GTP analogue. Triplicate samples are shown.

Figure S5. Ec Y772 (Tt β' Y1070) is coupled to the active site and the F loop through the link domain. The image was prepared using the program Visual Molecular Dynamics [15] from a late simulation snapshot derived from a molecular dynamics simulation of a Tt RNAP TEC with a closed trigger loop and active site restraints on the Mg^{2+} atoms I and II. Tt RNAP β' G1076 and G1080 are shown in space-filling representation to identify the dynamic bridge helix glycine hinge 1076-GARKGG-1081. Some of the chain is indicated in "secondary structure" representation: helix is magenta; turn is cyan; coil is white; extended chain is gold. Some atoms are colored according to chemistry: blue is nitrogen; cyan is carbon; red is oxygen; gold is phosphorous; white is hydrogen; pink is Mg. An arrow indicates Y1070. Amino acids indicated in red in alignments make direct or functional contacts with Y1070.

Figure S6. Proposed switching of Tt link domain β R557 (Ec β R678) to participate in pyrophosphate release (left panel; open trigger loop TEC) or catalysis (right panel; closed trigger loop TEC). ATP, Mg, YFI and bridge helix glycine hinge residues β' G1076 and G1080 are in space-filling representation. Link domain β 553-DDANR-557 and Mg-II binding β 835-ED-836 are in stick representation.

Figure S7. Atomic contacts of F773. The image was colored as in Figure S4. Two views are shown.

Supplementary references:

[1] V. Svetlov, G.A. Belogurov, E. Shabrova, D.G. Vassylyev, I. Artsimovitch, Allosteric control of the RNA polymerase by the elongation factor RfaH, Nucleic acids research, 35 (2007) 5694-5705.
[2] K.S. Wilson, P.H. von Hippel, Transcription termination at intrinsic terminators: the role of the RNA hairpin, Proc Natl Acad Sci U S A, 92 (1995) 8793-8797.
[3] I. Gusarov, E. Nudler, The mechanism of intrinsic transcription termination, Molecular cell, 3 (1999) 495-504.
[4] H.M. Johnston, W.M. Barnes, F.G. Chumley, L. Bossi, J.R. Roth, Model for regulation of the histidine operon of Salmonella, Proceedings of the National Academy of Sciences of the United States of America, 77 (1980) 508-512.
[5] V. Epshtein, C.J. Cardinale, A.E. Ruckenstein, S. Borukhov, E. Nudler, An allosteric path to transcription termination, Molecular cell, 28 (2007) 991-1001.
[6] D.G. Vassylyev, M.N. Vassylyeva, A. Perederina, T.H. Tahirov, I. Artsimovitch, Structural basis for transcription elongation by bacterial RNA polymerase, Nature, 448 (2007) 157-162.
[7] D.G. Vassylyev, M.N. Vassylyeva, J. Zhang, M. Palangat, I. Artsimovitch, R. Landick, Structural basis for substrate loading in bacterial RNA polymerase, Nature, 448 (2007) 163-168.
[8] M.L. Kireeva, K. Opron, S.A. Seibold, C. Domecq, R.I. Cukier, B. Coulombe, M. Kashlev, Z.F. Burton, Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase, BMC biophysics, 5 (2012) 11.

[9] S.A. Seibold, B.N. Singh, C. Zhang, M. Kireeva, C. Domecq, A. Bouchard, A.M. Nazione, M. Feig, R.I. Cukier, B. Coulombe, M. Kashlev, M. Hampsey, Z.F. Burton, Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase, Biochimica et biophysica acta, 1799 (2010) 575-587.
[10] R.O. Weinzierl, The Bridge Helix of RNA polymerase acts as a central nanomechanical switchboard for coordinating catalysis and substrate movement, Archaea, 2011 (2011) 608385.
[11] L.T. Da, D. Wang, X. Huang, Dynamics of Pyrophosphate Ion Release and Its Coupled Trigger Loop Motion from Closed to Open State in RNA Polymerase II, J Am Chem Soc, 134 (2012) 2399-2406.
[12] X. Huang, D. Wang, D.R. Weiss, D.A. Bushnell, R.D. Kornberg, M. Levitt, RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription, Proc Natl Acad Sci U S A, 107 (2010) 15745-15750.
[13] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, Journal of Computational Chemistry, 26 (2005) 1781-1802.
[14] D.A. Case, T.E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J Comput Chem, 26 (2005) 1668-1688.
[15] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, Journal of molecular graphics, 14 (1996) 33-38, 27-38.

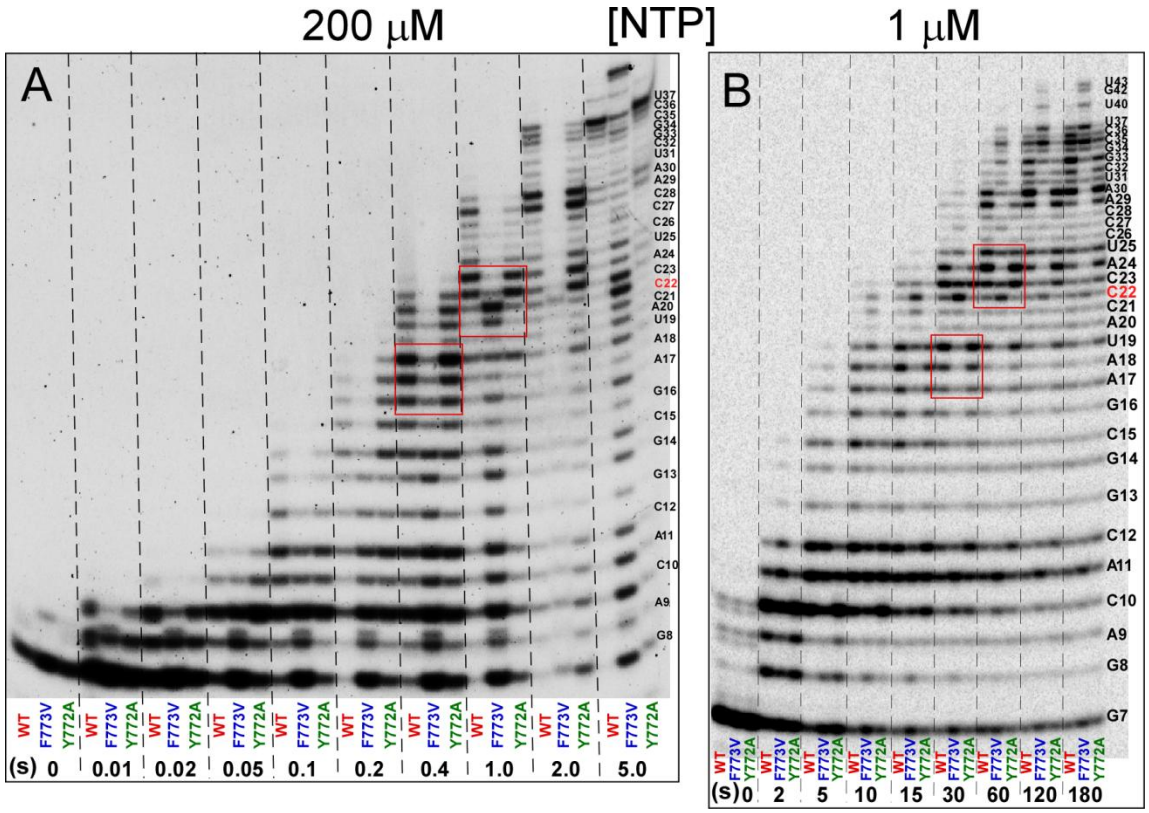

200 μM
[NTP]
1 μM
A
B
WT
F773V
Y772A
(s) 0 0.01 0.02 0.05 0.1 0.2 0.4 1.0 2.0 5.0
(s) 0 2 5 10 15 30 60 120 180

Figure S1

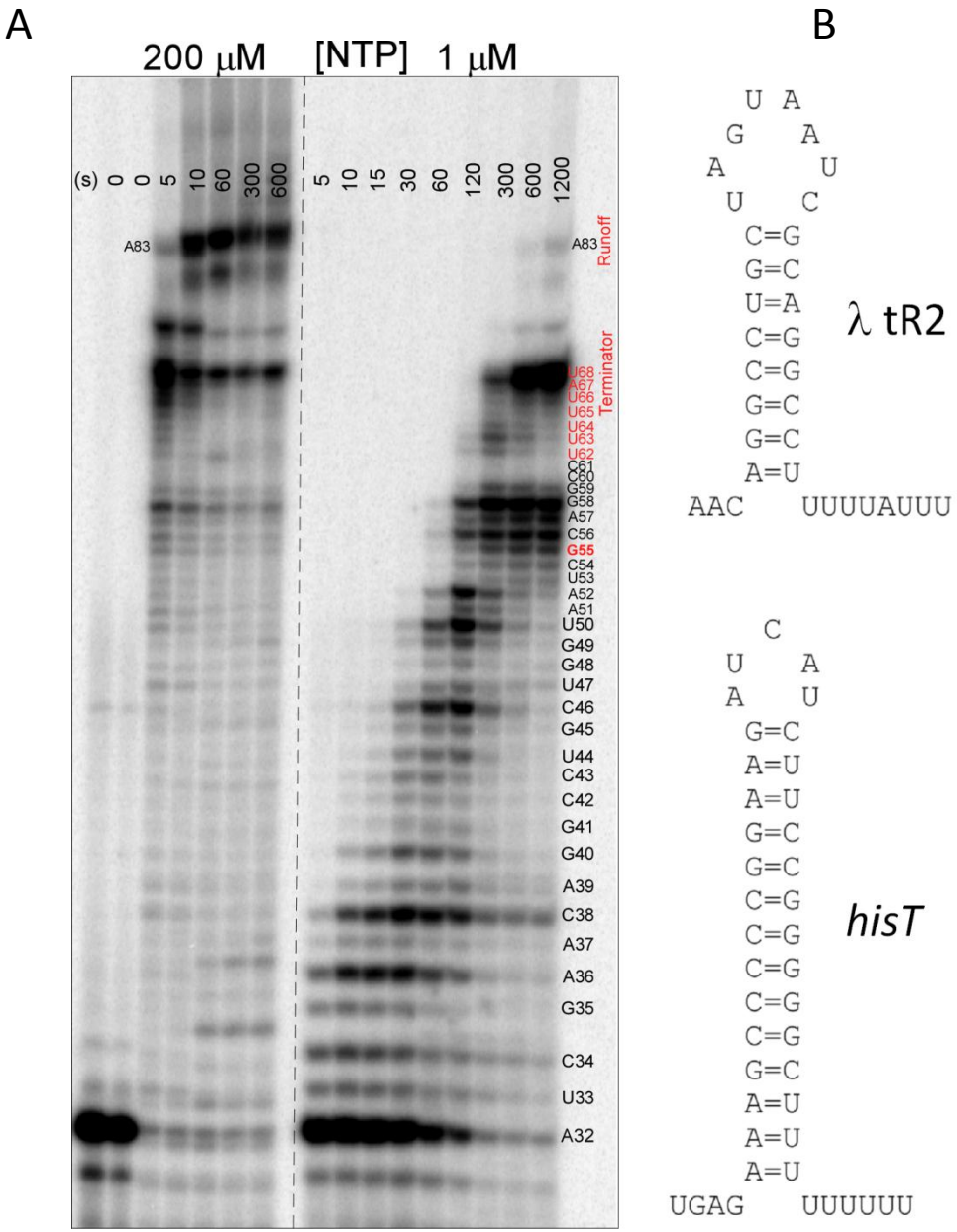

A
200 μM
[NTP]
1 μM
(s) 0 0 5 10 60 300 600
5 10 15 30 60 120 300 600 1200
A83
Runoff
Terminator
U68
A67
U66
U65
U64
U63
U62
C61
C60
G59
G58
A57
C56
G55
C54
U53
A52
A51
U50
G49
G48
U47
C46
G45
U44
C43
C42
G41
G40
A39
C38
A37
A36
G35
C34
U33
A32
B
U A
G A
A U
U C
C=G
G=C
U=A
C=G
C=G
G=C
G=C
A=U
AAC UUUUAUUU
λ tR2
C
U A
A U
G=C
A=U
A=U
G=C
G=C
C=G
C=G
C=G
C=G
C=G
G=C
A=U
A=U
A=U
UGAG UUUUUU
hisT

Figure S2

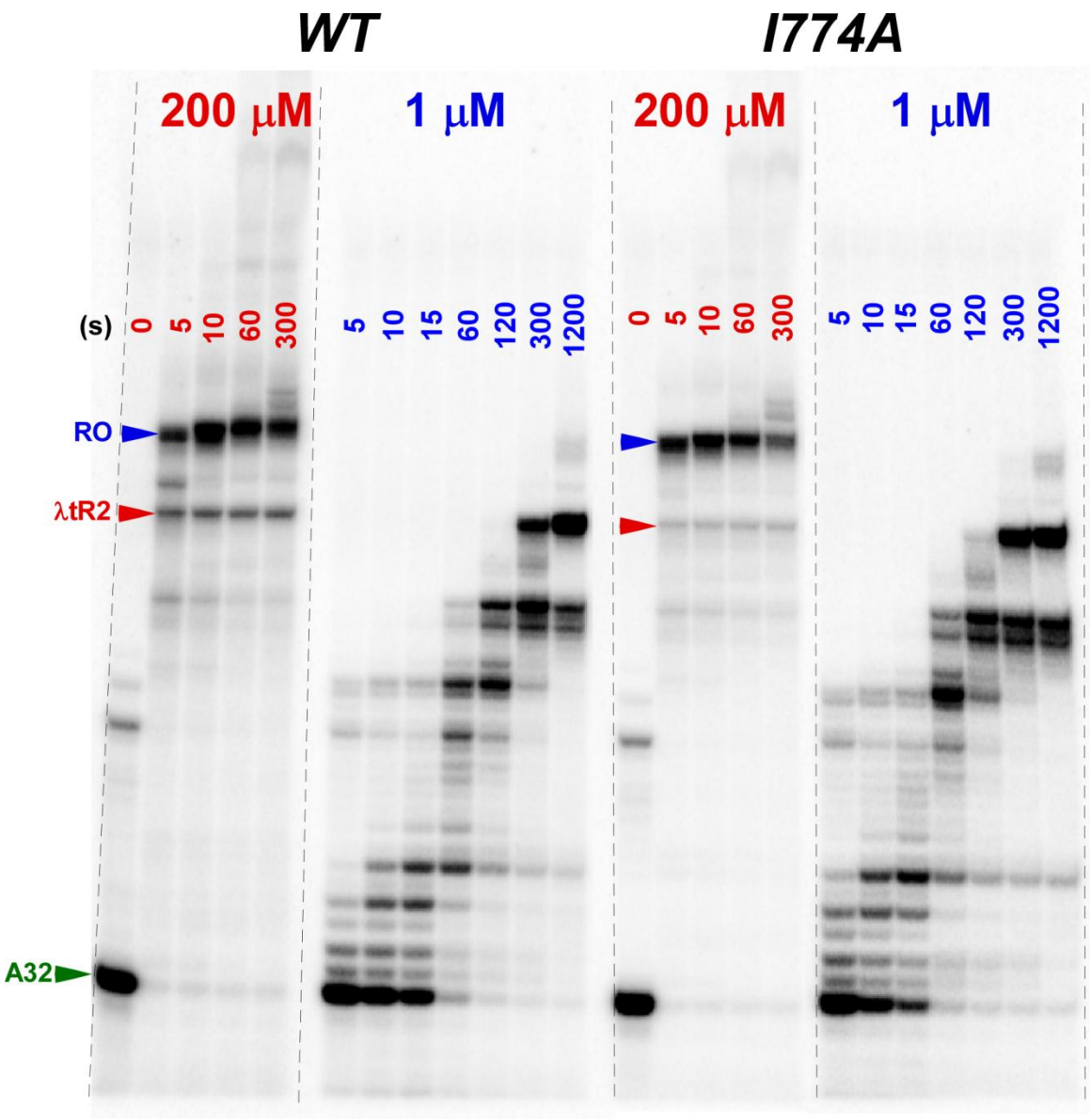

WT
I774A
200 μM
1 μM
200 μM
1 μM
(s)
0
5
10
60
300
5
10
15
60
120
300
1200
0
5
10
60
300
5
10
15
60
120
300
1200
RO
λtR2
A32

Figure S3

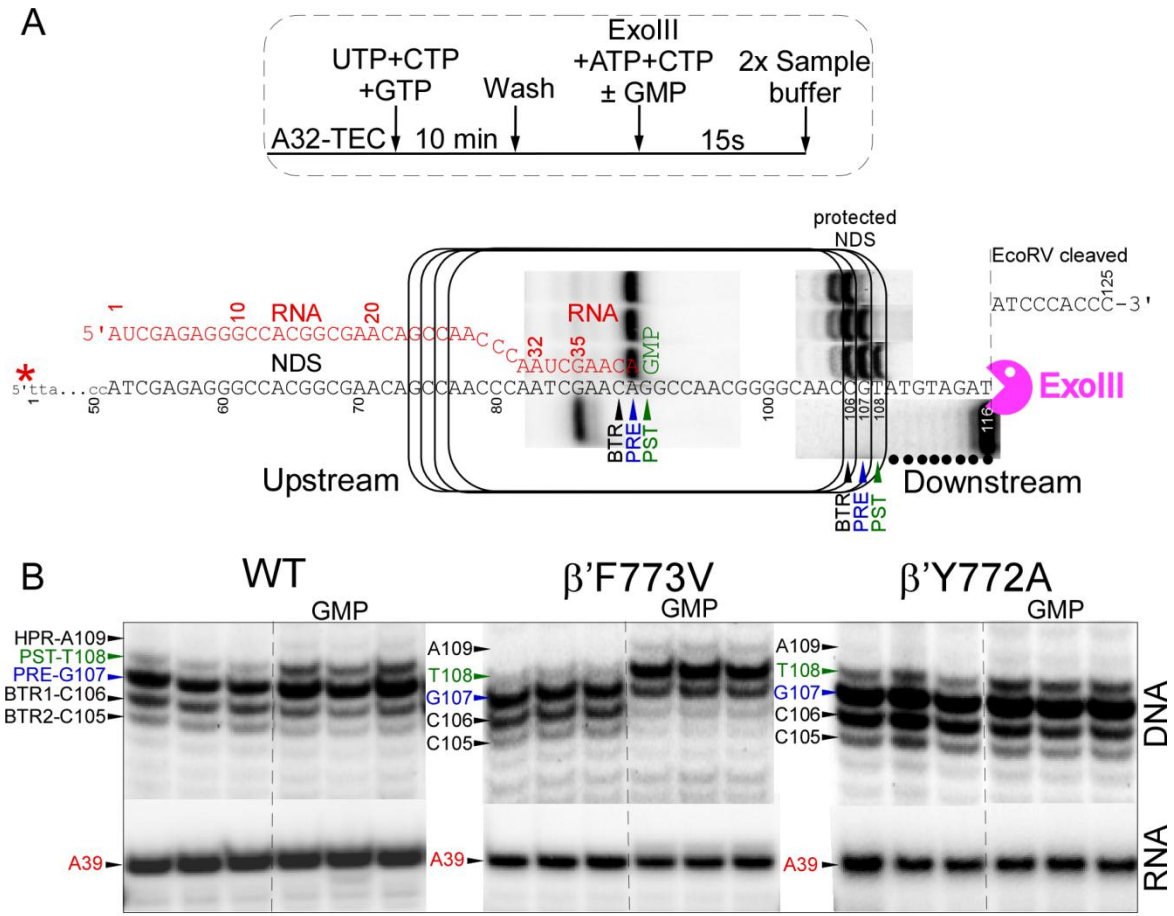
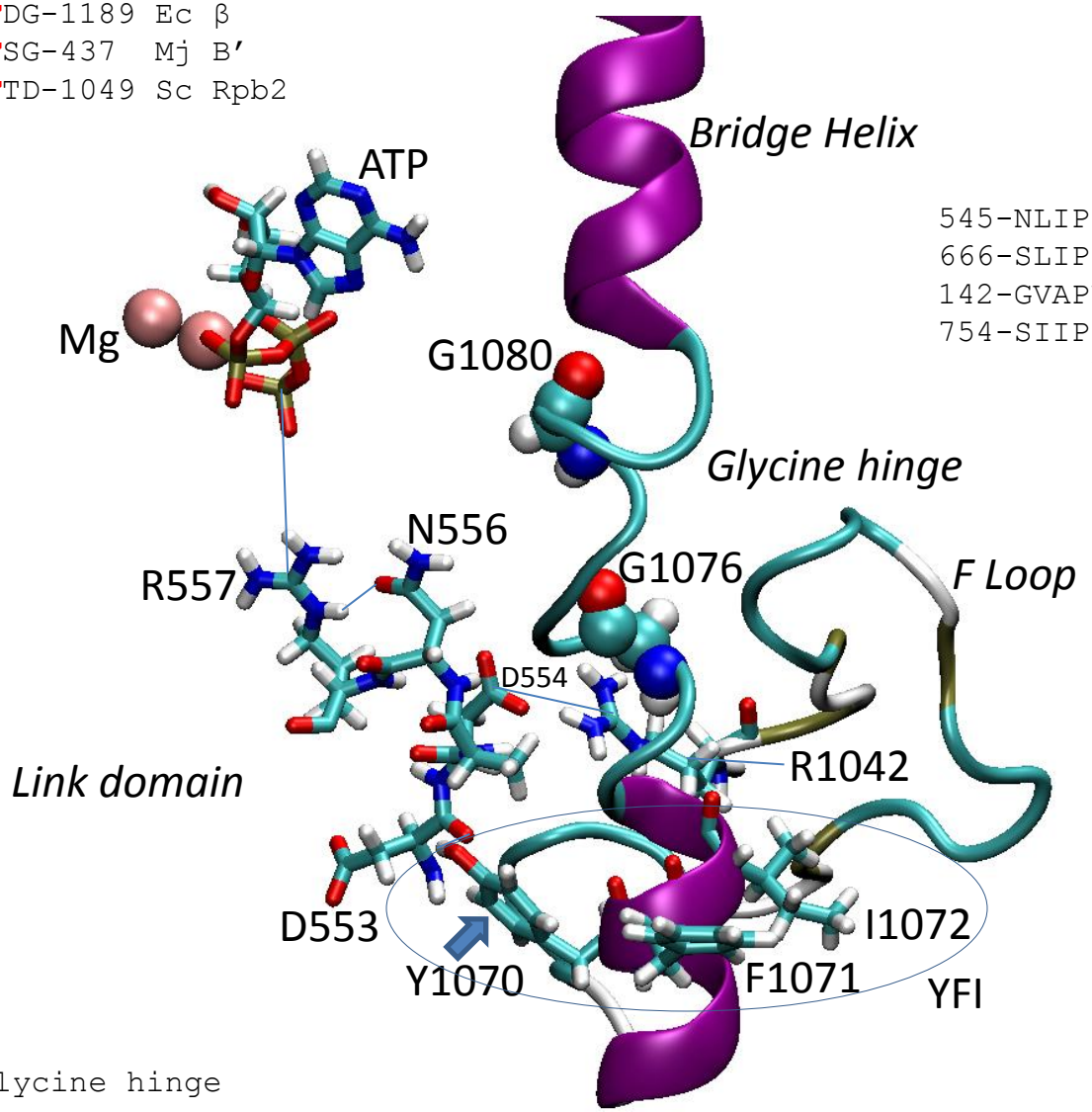

A
UTP+CTP
+GTP
Wash
ExoIII
+ATP+CTP
± GMP
2x Sample
buffer
A32-TEC
10 min
15s
protected
NDS
EcoRV cleaved
ATCCCACCC-3'
RNA
NDS
GMP
ExoIII
Upstream
Downstream
BTR
PRE
PST
B
WT
β'F773V
β'Y772A
GMP
HPR-A109
PST-T108
PRE-G107
BTR1-C106
BTR2-C105
A109
T108
G107
C106
C105
A39
DNA
RNA

Figure S4

Figure S5

```
 890-RYISPIFDG-908  Tt β
1181-PIATPVFDG-1189 Ec β
 429-RIDGTIFSG-437  Mj B'
1041-EGDASPFTD-1049 Sc Rpb2
```



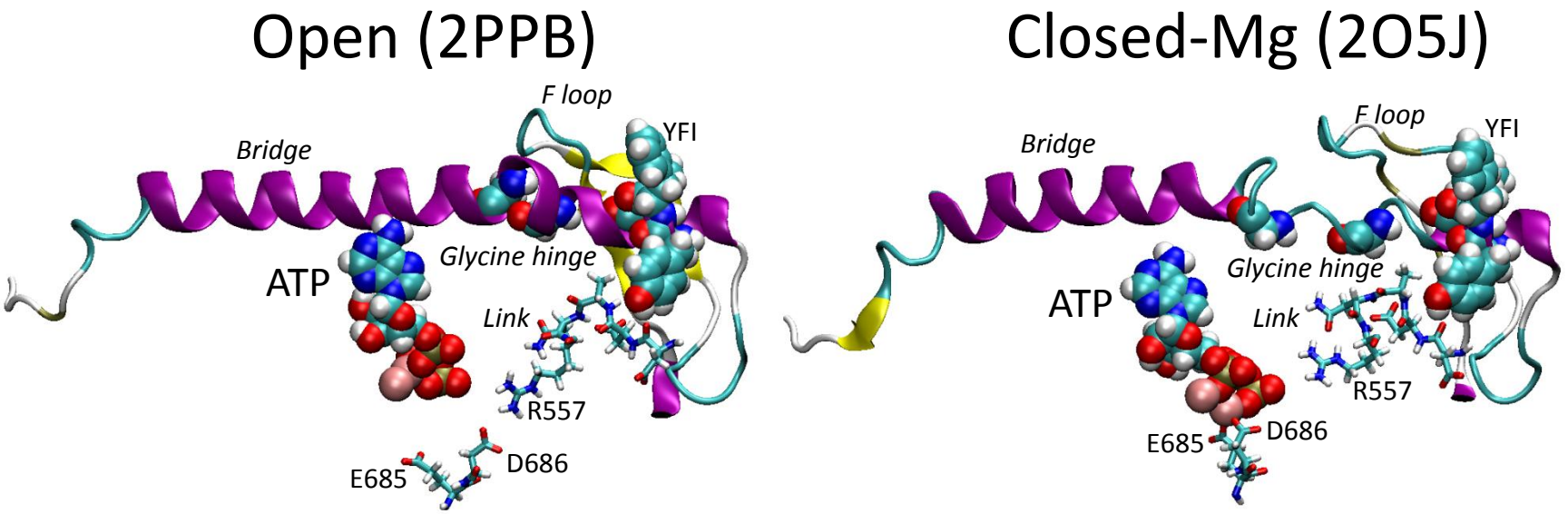

```
545-NLIPFLEHDDANRALMGSNM-564 Tt β
666-SLIPFLEHDDANRALMGANM-685 Ec β
142-GVAPYPEHNSAPRITMAAAM-161 Mj B"
754-SIIPFPDHNQSPRNTYQSAM-773 Sc Rpb2
        Link domain
```

glycine hinge

YFI

```
1067-VLEYFISSHGARKGGADTALRTADSGYLTRKLVDVTHE-1104 Tt β'   1034-QIRQLCGLRGLMQKPSGETFE-1054 Tt β'
 769-VLQYFISTHGARKGLADTALKTANSGYLTRRLVDVAQD-806  Ec β'    736-QIRQLAGMRGLMAKPDGSIIE-756  Ec β'
 800-PTEFFFHAMGGREGLVDQAVRTAQSGYMQRRLINALQD-837  Mj A'    747-NLTQMAACLGQQSVRGKRIFR-767  Mj A'
 810-PQEFFFHAMGGREGLIDTAVKTAETGYIQRRLVKALED-847  Sc Rpb1  757-NIAQMSACVGQQSVEGKRIAF-777  Sc Rpb1
                Bridge helix                                        F loop
```

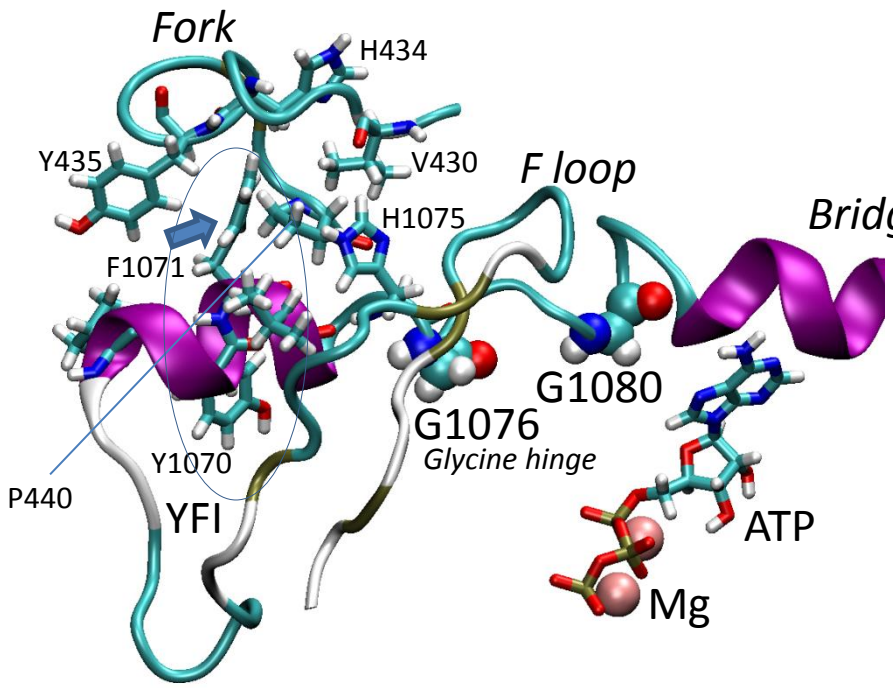
Figure S6

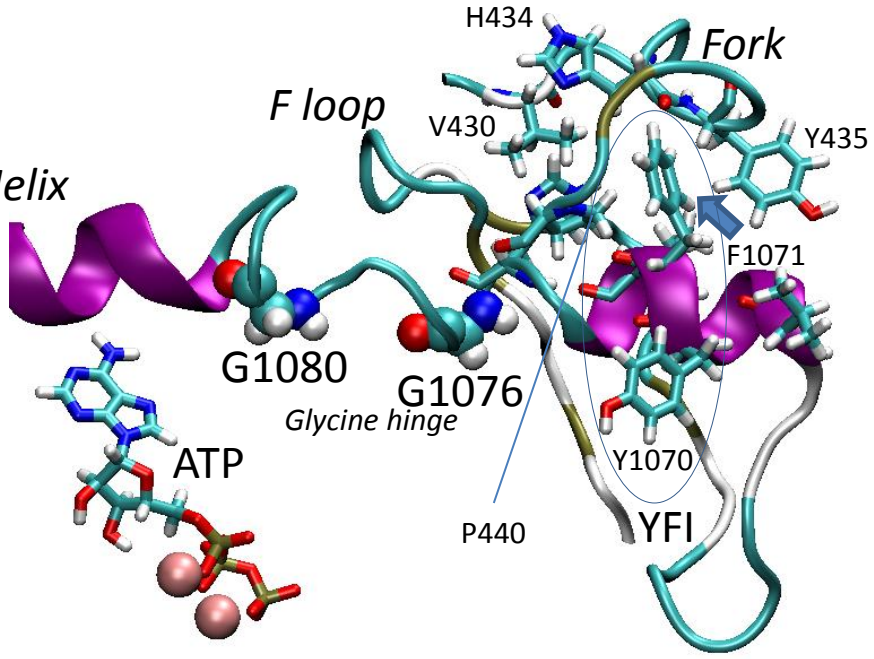



```
678-PFDGYNFEDAIVI-690 Tt β
807-PWNGYNFEDSILV-819 Ec β
217-SYEGYNMEDAIVF-229 Mj B′
829-CYSGYNQEDSMIM-841 Sc Rpb2
    Mg-II binding
```

```
          glycine hinge
        YFI
1067-VLEYFISSHGARKGGADTALRTADSGYLTRKLVDVTHE-1104 Tt β′
 769-VLQYFISTHGARKGLADTALKTANSGYLTRRLVDVAQD-806  Ec β′
 800-PTEFFFHAMGGREGLVDQAVRTAQSGYMQRRLINALQD-837  Mj A′
 810-PQEFFFHAMGGREGLIDTAVKTAETGYIQRRLVKALED-847  Sc Rpb1
                  Bridge helix
```

```
545-NLIPFLEHDDANRALMGSNM-564 Tt β
666-SLIPFLEHDDANRALMGANM-685 Ec β
142-GVAPYPEHNSAPRITMAAAM-161 Mj B″
754-SIIPFPDHNQSPRNTYQSAM-773 Sc Rpb2
        Link domain
```






```
420-RERAGFDVRDVHRTHYGRICPVETPE-445 Tt β
540-RERAGFEVRDVHRTHYGRVCPIETPE-565 Ec β
436-RSQPHFEARELHGTHWGKICPSETPE-461 Mj B"
504-RDGKLAKPRQLHNTHWGLVCPAETPE-529 Sc Rpb2
                Fork
```

```
                glycine hinge
          YFI
1067-VLEYFISSHGARKGGADTALRTADSGYLTRKLVDVTHE-1104 Tt β'
 769-VLQYFISTHGARKGLADTALKTANSGYLTRRLVDVAQD-806  Ec β'
 800-PTEFFFHAMGGREGLVDQAVRTAQSGYMQRRLINALQD-837  Mj A'
 810-PQEFFFHAMGGREGLIDTAVKTAETGYIQRRLVKALED-847  Sc Rpb1
                    Bridge helix
```

Figure S7