Structures of the RNA polymerase-σ^{54} reveal new and conserved regulatory strategies

Yun Yang[1,4,+], Vidya C. Darbari[1,2,+], Nan Zhang[3], Duo Lu[1,5], Robert Glyde[1,2], Yiping Wang[4], Jared Winkelman[6], Richard L. Gourse[6], Katsuhiko S. Murakami[7], Martin Buck[3], Xiaodong Zhang[1,2,*]

[1]Centre for Structural Biology, [2]Department of Medicine, [3]Department of Life Sciences, Imperial College London, South Kensington, SW7 2AZ, UK

[4]State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, China

[5]current address: State Key Laboratory of Bioactive Substance and Function of Natural Medicine, Institute of Material Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, China

[6]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI53706, USA

[7]Department of Biochemistry and Molecular Biology, Center for RNA Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA

[+]These authors contributed equally to this work

[*]Correspondence should be addressed to: Xiaodong.zhang@imperial.ac.uk, +44 (0)207 5943151

Abstract

Transcription by RNA polymerase in bacteria requires specific promoter recognition by σ factors. The major variant σ factor (σ^{54}) initially forms a transcriptionally silent complex requiring specialised ATP-dependent activators for initiation. Our crystal structure of the 450 kDa RNAP-σ^{54} holoenzyme at 3.8 Å reveals molecular details of σ^{54} and its interactions with RNAP. The structure explains how σ^{54} targets different regions in RNAP to exert its inhibitory function. Although σ^{54} and the major σ factor, σ^{70}, have similar functional domains and contact similar regions of RNAP, unanticipated differences are observed in their domain arrangement and interactions

with RNAP, explaining their distinct properties. Furthermore, we observe evolutionarily conserved regulatory hotspots in RNAPs that can be targeted by a diverse range of mechanisms to fine tune transcription.

Gene transcription is a tightly regulated event. A range of factors are required to maintain RNAP in a signal responsive, but inhibited state and to target it to specific genes (*1*). Bacterial sigma (σ) factors, eukaryotic TFIIB and other general transcription factors, are the primary promoter recruitment factors. The major variant σ^{54} (also called σ^{N}) is utilised in transcribing genes for numerous stress responses (*2*). Unlike the major σ^{70} class, which recognises the -35 and -10 promoter DNA elements, the σ^{54} class directs RNAP to promoter sites through the -24 and -12 DNA elements and forms a stable closed promoter complex that is unable to spontaneously melt DNA and initiate transcripts (*3*). Instead, the initiation process requires ATP-dependent activator proteins bound to upstream enhancer sites to actively remodel the RNAP-σ^{54}-DNA complex (*4*).

Our crystal structure of RNAP-σ^{54} reveals that σ^{54} contains four structural domains connected by long coils and loops that span a large area of the RNAP core enzyme, which consists of two α, β, β' and ω subunits (*5, 6*) (Fig. 1A-C, Fig. S1A-B, Table S1). σ^{54} Region I (RI, residues 1-56) forms a hook comprised of two α-helices. Region II (RII, residues 57-120) also contains two α-helices (residues 57-85) in addition to loops that are buried inside the RNAP (Fig. 1A, 1C). The RNAP core-binding domain (CBD) of σ^{54} extends as a structural fold to residue 250 and consists of two α-helical subdomains. Following on from the CBD, the backbone extends back to connect to a loop region, before an extra long α-helix (residues 315-353, hereafter called ELH, spanning 50 Å) followed by the HTH domain (residues 365 – 385), involved in interaction with the -12 promoter elements (*3*). The domain containing the RpoN box (RpoN domain), responsible for recognizing the -24 promoter elements, consists of a three-helical bundle (residues 415 – 477) (Fig. 1A) (*3*). The σ^{54} polypeptide chain snakes back and forth through its loop regions embedded in the RNAP. We carried out crosslinking experiments between σ^{54} and RNAP in solution using a *p*-benzoyl-L-phenylalanine-incorporated RNAP library. The amino acids that are crosslinked to σ^{54} agree well with the RNAP-σ^{54} crystal structure (Fig. 1D yellow, Fig. S1C).

We crystallized RNAP-σ^{54} in complex with a 28 bp DNA containing the -24 and -12 promoter elements. The crystals diffracted to 8 Å resolution, with the electron density for DNA being unambiguous (Fig. 1E), and the molecular envelope for the holoenzyme clear. We thus constructed a structural model of RNAP-σ^{54}-DNA (Fig. 1E, Fig. S1D).

All σ factors contain a major RNAP core-binding domain (CBD) and a major DNA binding domain, recognising either the -35 or -24 elements. The CBD of σ^{54} (Fig. 1A) binds towards the upstream face of the RNAP (Fig. 1C, based on promoter DNA binding orientation) making extensive interactions with many functional modules within the RNAP including the β-flap (residues 835-937), the C-terminus of the β subunit (residues 1267 – 1320), the β' zipper/Zn-binding domain (residues 35-100), the β' dock domain (residues 370-420) as well as the α subunit carboxyl terminal domain (α-CTD) (Fig. 2A). Using FeBABE cleavage assays, residue 198 of σ^{54} (within the CBD) can indeed be mapped to a region within the β-flap (*7*).

The RpoN domain of σ^{54} is the most conserved domain among σ^{54} from different organisms (Fig. S2). This domain extends from the main body of RNAP and does not contact other parts of σ^{54} nor core enzyme (Fig. 1A, 1C). Instead, it interacts with an adjacent RNAP molecule in the crystal, suggesting that its location is flexible in solution. In the RNAP-σ^{54}-DNA complex model, the RpoN domain is indeed moved relative to the RNAP-σ^{54} structure (Fig. S1D). Unlike the flexible RpoN domain, the RIII-HTH domain of σ^{54}, which binds to the -12 promoter region, stably associates with the RNAP core through RI-RIII (Fig. 1E, Fig. S1D) (*8-10*). Our data thus suggest that -12 binding is a major functional determinant for σ^{54} holoenzyme promoter recognition as well as stable closed complex formation.

The σ^{54} RI plays an inhibitory role and contains contact sites for its cognate activator proteins (*11, 12*). RI interacts with RIII, forming a structural module that lies along the cleft between β and β' (Fig. 1C and Fig. 2B), where template strand DNA enters the active site cleft (*13*). Our structure clearly suggests that the entry of promoter DNA into RNAP is blocked by RI and RIII of σ^{54}. Fitting of the RNAP-σ^{54} crystal structure into the cryo-EM map of activator bound RNAP-σ^{54} holoenzyme (*14*) positions the RI helices inside the connecting density leading to the activator protein,

indicating that RI in the RNAP-σ^{54} structure is presented favourably to contact the activator (Fig. S3A).

RII penetrates deeply into the DNA binding channel (Fig. 1, Fig. 2C-D). RII can be divided into three sub-regions based on their locations in the holoenzyme. RII.1 (residues 57-85) occupies the downstream DNA-binding cleft of RNAP, just above the bridge helix, thus playing an inhibitory role (Fig. 1 and Fig. 2C-D, Fig. S3A right). RII.2 (residues 86-105) occupies the space of the DNA template strand (Fig. 2C-D), indicating that RII.2 has to relocate to permit template-strand DNA access into the RNAP active site and for transcription initiation. RII.3 (residues 105-120) extends along the path that is occupied by RNA in the transcribing complex (Fig. 2C-D), suggesting that it will also need to relocate upon RNA synthesis.

Comparisons between σ^{54} and σ^{70} holoenzyme structures reveal that overall they contact similar regions of the RNAP core enzyme (Fig. 3A-B), in agreement with the idea that specific factors unrelated by structure and sequence can be functionally similar (*15*). However, the relative location of functional domains is completely different. For example, the σ^{54} CBD is located upstream, blocking the RNA exit tunnel (Fig. 3C) while σ_2, the major core binding domain of σ^{70}, contacts the downstream β' region (Fig. 3A-B). The structural differences suggest that σ^{54} CBD, and hence σ^{54}, will have to dissociate or relocate to allow progression from transcription initiation to elongation while σ^{70} can loosely associate with the RNAP core even during transcript elongation, in agreement with earlier biochemical data (*16, 17*). σ^{70} region 4 (σ_4), which recognises the -35 promoter element, occupies the upstream surface blocking the RNA exit channel instead (Fig. 3B).

The inhibitory RI of σ^{54} interacts with RIII across the RNAP cleft, blocking the template strand from entering. On the other hand, $\sigma_{1.1}$ of σ^{70}, which also has an inhibitory role (*18*), is located in the downstream DNA channel, overlapping with the region occupied by σ^{54} RII.1 (Fig. 3), suggesting that $\sigma_{1.1}$ and RII.1 might play related roles in transcription regulation. However, $\sigma_{1.1}$ connects with σ_2 by only a single flexible linker, allowing the $\sigma_{1.1}$ to move readily from the RNAP cleft (Fig. 1B, Fig. 3B). σ^{54} RII.1 is connected to the RI and hence the RIII-ELH as well as the CBD and the linker that connects RII.1 with the CBD is embedded in the DNA/RNA channel of

RNAP (Fig. 2C-D). Due to its topological restraint, σ^{54} RII.1 cannot relocate easily without affecting the conformation of other parts of σ^{54}, RII.1 will thus block downstream DNA entry.

Notable conformational differences exist between the σ^{70} and σ^{54} holoenzymes in many key modules within the RNAP when superposed on the bridge helix, chosen since the catalytic centre must be conserved (Fig. S3B). The β' coiled-coil, and the whole β' subunit, is rotated, narrowing the downstream DNA channel by as much as 5 Å (Fig. S3B). However, the clamp, which consists of β and β' domains forming the walls of the downstream DNA channel, was shown to adopt multiple conformations in solution (*19*). The β' coiled-coil forms helical bundles with σ_2 of σ^{70} and helps to stabilise the unpaired non-template DNA strand, and hence the transcription bubble (*20*) (Fig. S4A). B-linker of TFIIB also interacts with β' coiled-coil (Fig. S4B). In the σ^{54} holoenzyme structure, the β' coiled-coil does not form a helical bundle with σ^{54} (Fig. S4C). However, it is possible that en-route to transcription initiation, this interaction is established, in agreement with our biochemical data (Fig. S4D).

σ^{54} occupies multiple locations along the DNA and RNA path within the RNAP core. These locations are also targeted by other regulatory factors and in other RNAPs. For example, RI and RIII sit across the RNAP cleft, blocking the DNA template strand entry (Fig. 4A). At similar locations, domains 2 and 3 (σ_2 and σ_3) of σ^{70} form a V-shaped wedge that acts as a gate to allow the template strand to enter the active cleft (Fig. 4B) (*13*, *21*). TFIIB also uses two sub-domains, the core binding domain and the B-linker, to occupy similar areas in Pol II (Fig. 4C). This area is thus occupied by σ^{54} to inhibit while by σ^{70} and TFIIB to guide the template strand delivery. Furthermore, σ_2 of σ^{70} contains a number of aromatic residues that are shown to interact with non-template strand, therefore facilitating DNA melting and transcription bubble formation (*20*). In σ^{54}, there are few aromatic residues in this region (RI and RIII-ELH) that could facilitate DNA melting, in agreement with data showing that although deletion of σ^{54} RI can bypass the requirement of activator proteins, it fails on double-stranded DNA (*11*). Activator proteins may therefore be directly required for transcription bubble opening. The downstream DNA channel is occupied by both RII.1 of σ^{54} and $\sigma_{1.1}$ of σ^{70} (Fig. 4D). A negatively charged region of TFIIF is also located in the downstream DNA channel of Pol II (*22*). RII.2 of σ^{54} is located in the

channel that is occupied by the template strand DNA in the transcribing RNAP (Fig. 2D). σ^{70} and TFIIB lack equivalent structural features. This area, however, is occupied in Pol I by part of the expander (Fig. 4E), a unique insertion in Pol I compared to Pol II and shown to stabilize an expanded active centre cleft (*23, 24*) (Fig. 4E). Similar to RII.3, both region 3.2 in σ^{70} (σ finger) and the B-reader in TFIIB are shown to overlap with the space that would be occupied by RNA (Fig. 4D-F) (*25*). Indeed, RII.3 shares sequence homology with σ finger, especially the highly conserved DDE motif (Fig. S2). Acidic residues in these elements are proposed to facilitate template DNA loading and RNA separation from DNA, guiding it towards the exit channel (*25-28*). It is possible that RII.3 also performs similar roles.

Our structure uncovers a diverse range of regulatory strategies utilised by σ^{54} in tightly controlling transcription initiation. These include blocking the template strand from entering the RNAP active cleft by RI and RIII, occupying the downstream DNA channel by RII.1, interfering with the template strand and synthesized RNA by RII.2-RII.3. The position of RI and RIII plays a vital role in these regulatory functions. Activator proteins interact with RI and could relocate RI, RIII-ELH as well as RII, releasing the inhibition posed by some of these structural elements (Fig. S3A) (*29, 30*). Precisely how activators overcome these multiple modes of inhibition, and the role of ATP in this process, remains to be determined. Furthermore, we show that although σ^{54}, σ^{70}, TFIIB, TFIIF and Pol I subunits have no sequence or structural similarities, they target the same elements within their respective RNAPs. Our structure and comparisons thus provide clear evidence that regulatory hotspots within RNAPs are functionally conserved. Different transcription factors, irrespective of their structures and sequences, are utilised to target these hotspots in order to exert a fine-tuned transcription regulation strategy (Fig. S5). It will be interesting to see the extent to which additional multi-subunit RNAPs employ these regulatory strategies.

Legends

Figure. 1. Structure of complete RNAP-σ^{54}. (A) Structure and schematic diagram of σ^{54}. (**B)** Structure of σ^{70} (PDB 4IGC) with similar functional domains colored accordingly. (**C)** Holoenzyme in different orientations, RNAP core in cylinders. β-cyan, β'-salmon pink, α-grey, ω-pale blue, σ^{54} in ribbon and colored as in (**A**), **DS** represents downstream, **US** represents upstream. (**D)** Residues of β and β' subunits

that are crosslinked to σ^{54} are mapped onto the structure and colored in yellow (**E)** Electron density map and structural model of RNAP-σ^{54}-DNA.

Figure 2. Functional domains of σ^{54} and interactions with RNAP core (**A)** Detailed view of the CBD and RNAP interactions. **(B)** RI and RIII, **(C)** DNA-RNA channels, DNA-orange, RNA-magenta and **(D)** Overlay with RII.1, RII.2 and RII.3.

Figure 3. Comparisons of σ^{54} and σ^{70} holoenzymes. (**A**) RNAP-σ^{54} with RNAP displayed as spheres and σ^{54} as surface. (**B)** RNAP-σ^{70}, same view as in (**A)**. (**C)** Elongation complex (PDB code 2O5J).

Figure 4. Regulatory elements of σ^{54} and comparison with σ^{70}, TFIIB and Pol I. **(A)** σ^{54} **(B)** σ^{70} and (**C)** TFIIB in relation to template strand (gold) as defined in the complex with a full transcription bubble (4YLN). (**D-F)** Comparison of σ^{54} with σ^{70}, Pol I (4C2M) and TFIIB (4BBS) in the RNAP DNA-RNA channel; structures are superposed on the bridge helix. σ^{54} represented as spheres while σ^{70}, TFIIB and Pol I expander as ribbons, RNAPs as cylinders.

References and Notes

1. R. A. Mooney, S. A. Darst, R. Landick, Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell* **20**, 335-345 (2005).
2. A. Feklistov, B. D. Sharon, S. A. Darst, C. A. Gross, Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**, 357-376 (2014).
3. M. Buck, M. T. Gallegos, D. J. Studholme, Y. Guo, J. D. Gralla, The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *J Bacteriol* **182**, 4129-4136 (2000).
4. M. Rappas, D. Bose, X. Zhang, Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. *Curr Opin Struct Biol* **17**, 110-116 (2007).
5. P. Cramer, Multisubunit RNA polymerases. *Curr Opin Struct Biol* **12**, 89-97 (2002).
6. F. Werner, D. Grohmann, Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**, 85-98 (2011).
7. S. R. Wigneshweraraj, P. C. Burrows, S. Nechaev, N. Zenkin, K. Severinov, M. Buck, Regulated communication between the upstream face of RNA polymerase and the beta' subunit jaw domain. *EMBO J* **23**, 4264-4274 (2004).
8. Y. Guo, J. D. Gralla, Promoter opening via a DNA fork junction binding activity. *Proc Natl Acad Sci U S A* **95**, 11655-11660 (1998).
9. Y. Guo, C. M. Lew, J. D. Gralla, Promoter opening by sigma(54) and sigma(70) RNA polymerases: sigma factor-directed alterations in the mechanism and tightness of control. *Genes Dev* **14**, 2242-2255 (2000).

10. P. C. Burrows, S. R. Wigneshweraraj, M. Buck, Protein-DNA interactions that govern AAA+ activator-dependent bacterial transcription initiation. *J Mol Biol* **375**, 43-58 (2008).
11. J. T. Wang, A. Syed, J. D. Gralla, Multiple pathways to bypass the enhancer requirement of sigma 54 RNA polymerase: roles for DNA and protein determinants. *Proc Natl Acad Sci U S A* **94**, 9538-9543 (1997).
12. M. Chaney, M. Buck, The sigma 54 DNA-binding domain includes a determinant of enhancer responsiveness. *Mol Microbiol* **33**, 1200-1209 (1999).
13. Y. Zuo, T. A. Steitz, Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Mol Cell* **58**, 534-540 (2015).
14. D. Bose *et al.*, Organization of an activator-bound RNA polymerase holoenzyme. *Mol Cell* **32**, 337-346 (2008).
15. K. B. Decker, D. M. Hinton, Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol* **67**, 113-139 (2013).
16. G. Bar-Nahum, E. Nudler, Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli. *Cell* **106**, 443-451 (2001).
17. J. Mukhopadhyay, A. N. Kapanidis, V. Mekler, E. Kortkhonjia, Y. W. Ebright, R. H. Ebright, Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **106**, 453-463 (2001).
18. B. Bae, E. Davis, D. Brown, E. A. Campbell, S. Wigneshweraraj, S. A. Darst, Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of sigma70 domain 1.1. *Proc Natl Acad Sci U S A* **110**, 19772-19777 (2013).
19. A. Chakraborty *et al.*, Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**, 591-595 (2012).
20. A. Feklistov, S. A. Darst, Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* **147**, 1257-1269 (2011).
21. K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-1290 (2002).
22. Z. A. Chen *et al.*, Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* **29**, 717-726 (2010).
23. C. Engel, S. Sainsbury, A. C. Cheung, D. Kostrewa, P. Cramer, RNA polymerase I structure and transcription regulation. *Nature* **502**, 650-655 (2013).
24. C. Fernandez-Tornero *et al.*, Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**, 644-649 (2013).
25. D. Kostrewa *et al.*, RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323-330 (2009).
26. S. Sainsbury, J. Niesser, P. Cramer, Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437-440 (2013).
27. X. Liu, D. A. Bushnell, D. Wang, G. Calero, R. D. Kornberg, Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206-209 (2010).
28. R. S. Basu *et al.*, Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem* **289**, 24549-24559 (2014).

29. A. Sharma *et al*., Domain movements of the enhancer-dependent sigma factor drive DNA delivery into the RNA polymerase active site: insights from single molecule studies. *Nucleic Acids Res* **42**, 5177-5190 (2014).
30. L. J. Friedman, J. Gelles, Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell* **148**, 679-689 (2012).
31. Y. Aiba, Y. Katayama, T. Hishinuma, H. Murakami-Kuroda, L. Cui, K. Hiramatsu, Mutation of RNA polymerase beta-subunit gene promotes heterogeneous-to-homogeneous conversion of beta-lactam resistance in methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother* **57**, 4861-4871 (2013).
32. M. J. Merrick, J. R. Gibbins, The nucleotide sequence of the nitrogen-regulation gene ntrA of Klebsiella pneumoniae and comparison with conserved features in bacterial RNA polymerase sigma factors. *Nucleic Acids Res* **13**, 7607-7620 (1985).
33. N. Joly, N. Zhang, M. Buck, ATPase site architecture is required for self-assembly and remodeling activity of a hexameric AAA+ transcriptional activator. *Molecular cell* **47**, 484-490 (2012).
34. L. Morris, W. Cannon, F. Claverie-Martin, S. Austin, M. Buck, DNA distortion and nucleation of local DNA unwinding within sigma-54 (sigma N) holoenzyme closed promoter complexes. *J Biol Chem* **269**, 11563-11571 (1994).
35. W. V. Cannon, M. T. Gallegos, M. Buck, Isomerization of a binary sigma-promoter DNA complex by transcription activators. *Nat Struct Biol* **7**, 594-601 (2000).
36. W. Kabsch, Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132 (2010).
37. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).
38. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica* **66**, 213-221 (2010).
39. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica* **66**, 486-501 (2010).
40. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica* **67**, 355-367 (2011).
41. E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea, G. Bricogne, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta crystallographica* **60**, 2210-2221 (2004).
42. M. Chlenov, S. Masuda, K. S. Murakami, V. Nikiforov, S. A. Darst, A. Mustaev, Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase beta' subunit. *J Mol Biol* **353**, 138-154 (2005).
43. B. A. Knutson, J. Luo, J. Ranish, S. Hahn, Architecture of the Saccharomyces cerevisiae RNA polymerase I Core Factor complex. *Nat Struct Mol Biol* **21**, 810-816 (2014).

Acknowledgements

We thank Martin Michael, Daniel Bose for their earlier contributions to this project, Andreas Forster, Jiwei Liu and beamline scientists at the Diamond Light Source for their help with data collection, members in XZ and MB's labs for helpful discussion. We thank Dale Wigley, Ray Dixon, Robert Weinzierl and Ramesh Wigneshweraraj for critically reading the manuscript. YY was funded by the Chinese National Science Foundation and the China Scholarship Council. The majority of this work is funded by the UK Biotechnology and Biological Sciences Research Council to XZ and MB. KSM is supported by National Institutes of Health Grant GM087350. YPW is funded by 973 National Key Basic Research Programme (2015CB755700) in China. The atomic coordinate has been deposited in the Protein Data Bank with accession code XXXX.

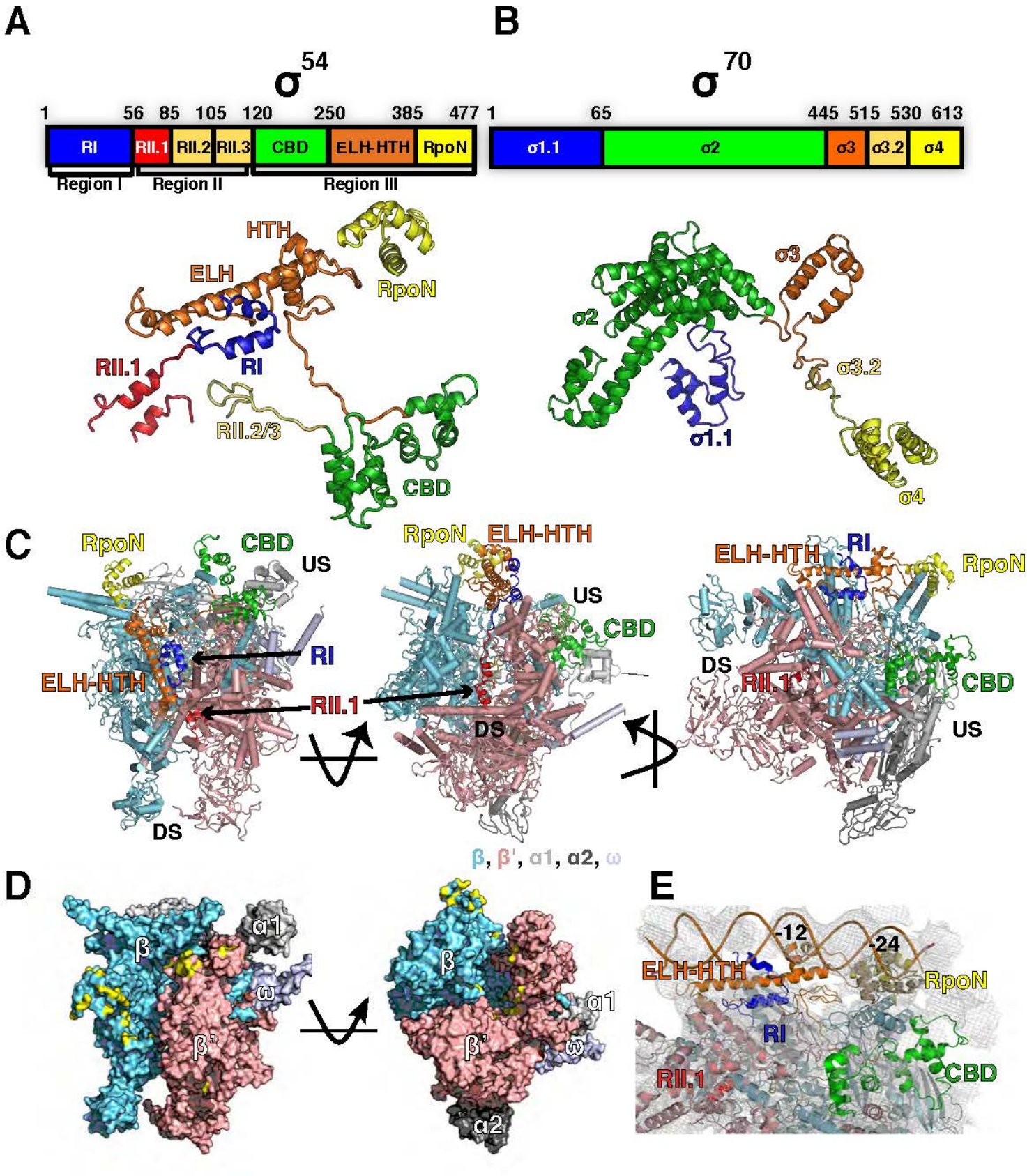
Supplementary Materials

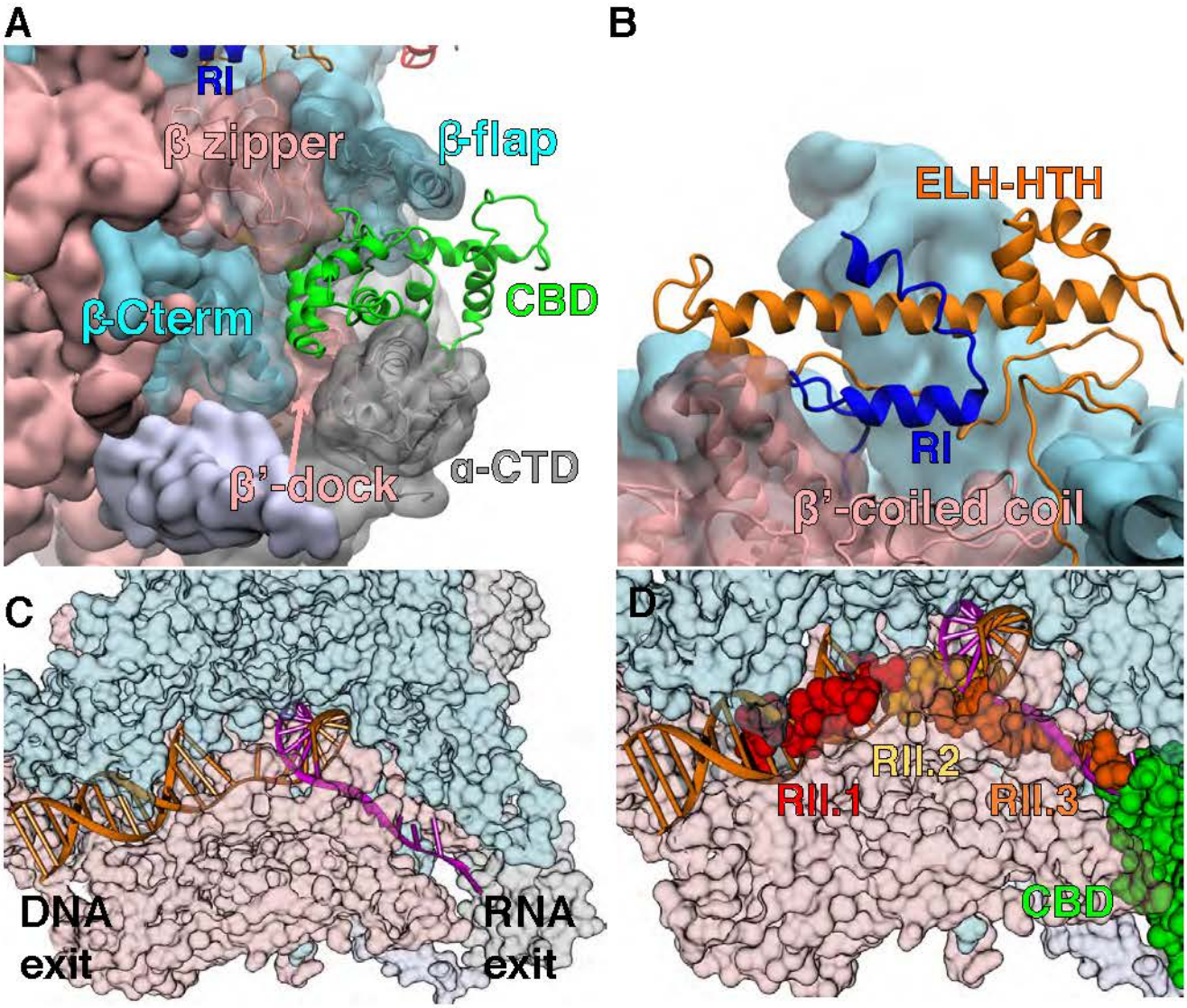
Materials and Methods

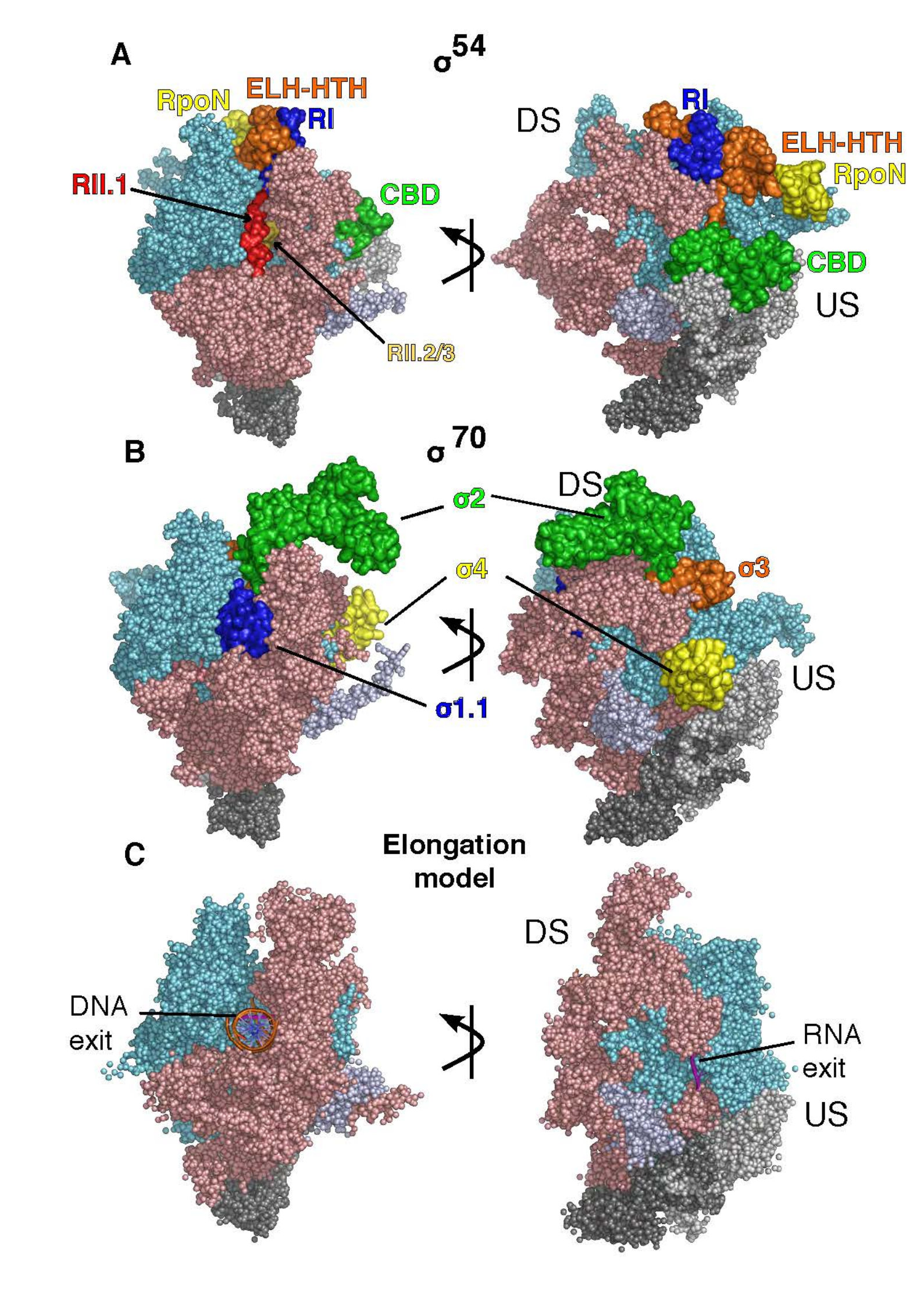
Figs. S1 to S5

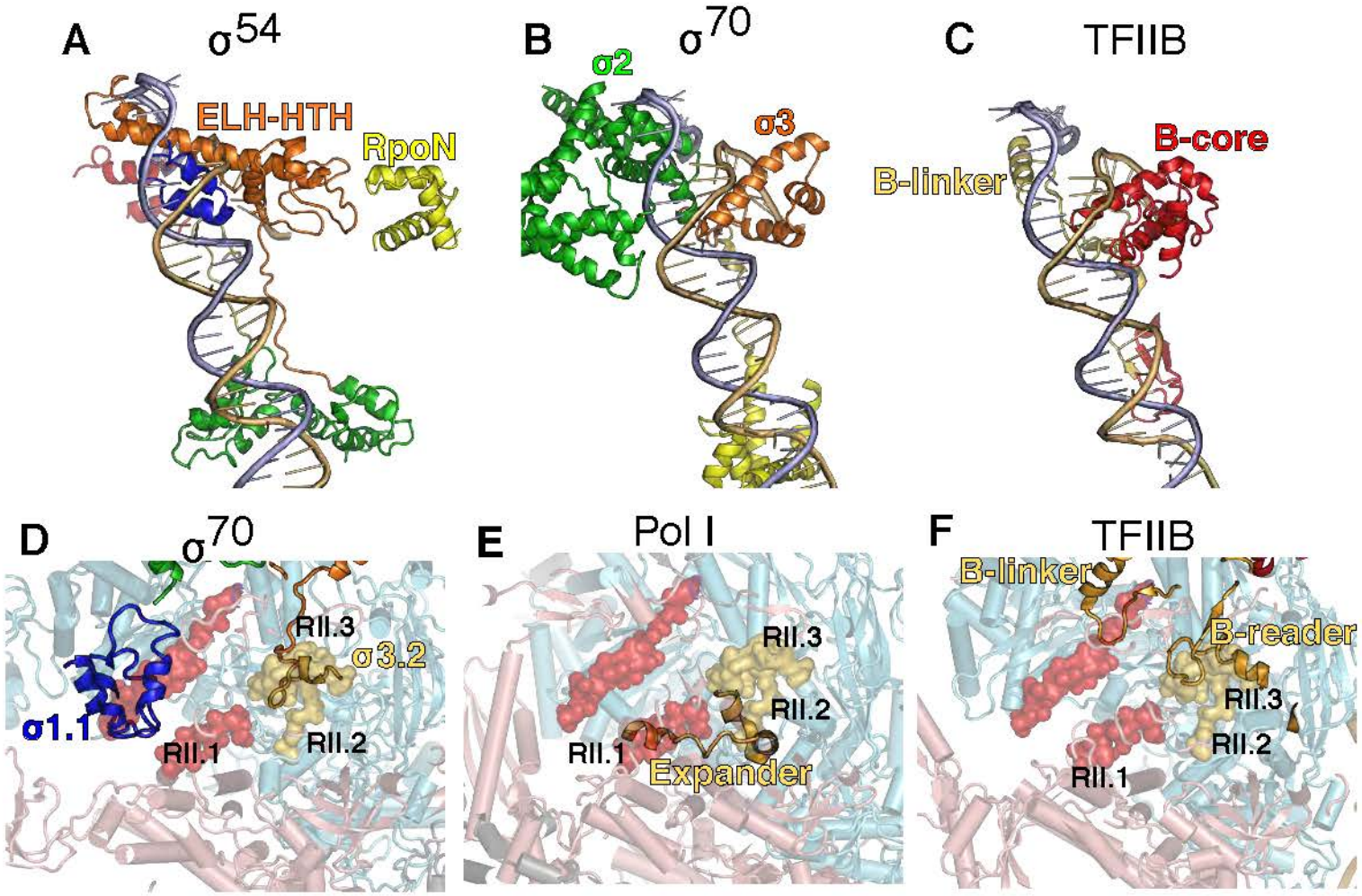
Tables S1

References (31-43)


A
σ54
1
56
85
105
120
250
385
477
RI
RII.1
RII.2
RII.3
CBD
ELH-HTH
RpoN
Region I
Region II
Region III
HTH
ELH
RpoN
RII.1
RI
RII.2/3
CBD
B
σ70
1
65
445
515
530
613
σ1.1
σ2
σ3
σ3.2
σ4
σ2
σ3
σ3.2
σ1.1
σ4
C
RpoN
CBD
US
RI
ELH-HTH
RII.1
DS
RpoN
ELH-HTH
US
CBD
DS
ELH-HTH
RI
RpoN
DS
RII.1
CBD
US
β, β', α1, α2, ω
D
α1
β
ω
β'
β
α1
β'
ω
α2
E
-12
-24
ELH-HTH
RpoN
RI
RII.1
CBD


A
RI
β zipper
β-flap
β-Cterm
CBD
β'-dock
α-CTD
B
ELH-HTH
RI
β'-coiled coil
C
DNA
exit
RNA
exit
D
RII.2
RII.1
RII.3
CBD


A
σ54
RpoN
ELH-HTH
RI
RII.1
CBD
RII.2/3
DS
RI
ELH-HTH
RpoN
CBD
US
B
σ70
σ2
DS
σ4
σ3
σ1.1
US
C
Elongation model
DNA exit
DS
RNA exit
US


A
σ54
ELH-HTH
RpoN
B
σ70
σ2
σ3
C
TFIIB
B-core
B-linker
D
σ70
RII.3
σ3.2
σ1.1
RII.1
RII.2
E
Pol I
RII.3
RII.2
RII.1
Expander
F
TFIIB
B-linker
B-reader
RII.3
RII.2
RII.1

Supplementary Materials for

Structures of the RNA polymerase-σ^{54} reveal new and conserved regulatory strategies

Yun Yang[1,4,+], Vidya C. Darbari[1,2,+] , Nan Zhang[3], Duo Lu[1,5], Robert Glyde[1,2], Yiping Wang[4], Jared Winkelman[6], Richard L. Gourse[6], Katsuhiko S. Murakami[7], Martin Buck[3], Xiaodong Zhang[1,2,*]

correspondence to: xiaodong.zhang@imperial.ac.uk

This PDF file includes:

Materials and Methods
Figs. S1 to S5
Table S1
References (31-43)

Materials and Methods

RNAP core purification

The Plasmid pGEMABC (addgene #45398) encoding *E. coli rpoA*, *rpoB* and *rpoC* (*31*) and plasmid pACYCDuet-omega (encoding *rpoZ*) were co-expressed in BL21(DE3) gold cells. The cells were grown at 25 ^{o}C to OD_{600} = 0.8 and then induced with 1mM IPTG for 3 hrs at 25 ^{o}C. Cells were harvested via centrifugation and re-suspended in lysis buffer (50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 5 mM DTT) supplemented with 1mM PMSF and protease inhibitor tablet (Sigma) and lysed through sonication. The soluble supernatant was collected and precipitated with 0.4% (final concentration) Polymin P (pH 7.9), followed by centrifugation at 18,000 g for 15 min. The precipitate was then washed with TGED buffer (10 mM Tris-HCl (pH 8.0), 5% glycerol, 0.1 mM EDTA, 1 mM DTT) with 500 mM NaCl. The sample was then eluted with TGED buffer + 1 M NaCl and precipitated with ammonium sulfate (with a final concentration of 50%). The precipitates were collected and suspended with TGED buffer and dialysed against TGED buffer + 100 mM NaCl. The sample was further purified with BioRex70 (BioRad) and HiTrap Q column (GE Healthcare). Purified protein was concentrated and stored at -80 ^{o}C.

σ^{54} purification

σ^{54} from *Klebsiella pneumoniae* M5A1(*32*) was expressed in BL21 (DE3) gold cells using pET28b-σ^{54} (*33*). *K. pneumonia* σ^{54} was used for consistency as it has been extensively studied in combination with *E. coli* core RNAP and is 90% identical (97% similar) to that of *E. coli*. Cells were grown at 37 ^{o}C until OD_{600} reached 0.6 and were induced at 22 ^{o}C with 1 mM IPTG. Cells were harvested through centrifugation and lysed by sonication. The soluble supernatant was collected and further purified through Ni and Heparin affinity chromatography (GE Healthcare). Selenomethionine substituted σ^{54} was expressed in B834 cells (methionine auxotrophic) and was purified using the same protocol as wild-type protein.

RNAP-σ^{54} holoenzyme purification and crystallization

The RNAP-σ^{54} was formed by incubating RNAP with σ^{54} at 4 ^{o}C for an hour with a

molar ratio of 1:4 and was further purified by size exclusion chromatography. The RNAP–σ^{54} complex was concentrated to 10 mg/ml. Crystals were grown at 18 ^{o}C in 48-well sitting-drop crystallization plates by mixing 0.5 μl protein with 0.5 μl crystallization solution (~9% PEG8000, ~18% Ethylene glycol, 100 mM Sodium HEPES/MOPS buffer (pH 7.9), 0.03M NaF, 0.03M NaBr, 0.03M NaI, 0-10mM DTT). The crystals were incubated in cryo-buffer (12% PEG8000, 25% Ethylene glycol, 100 mM Sodium HEPES/MOPS buffer (pH 7.9), 0.03 M NaF, 0.03 M NaBr, 0.03 M NaI, 10 mM $MgCl_2$) for 30 min and flash frozen in liquid nitrogen. The 28 bp DNA contains -30 to -3 *nifH* promoter region with a mismatch at -12/-11 to mimic the base unstacking observed in the closed complex DNA (*34, 35*).

Data collection, structural determination and refinement

Diffraction datasets were collected for native and Selenomethionine substituted crystals under cryogenic conditions (100K) at beam lines I03 & I04, the Diamond Light Source, UK. All the datasets were processed using XDS (*36*). Phases were obtained with molecular replacement using the *E. coli* RNA polymerase core structure (PDB code: 4IGC[12]) as the search model in PHASER (*37*). Density modification with multi-crystal averaging and feature-enhanced maps were calculated to improve the electron density map using the PHENIX software suite (*38*). The maps improved to reasonable quality thereby aiding *ab-initio* model building of σ^{54} in Coot (*39*). The core binding domain and C-terminal RpoN domain could be identified in the initial model using known structures for these domains from *A. aeolicus* (PDB code: 2O8K for RpoN domain and PDB code: 2KRM for core binding domain). Anomalous map calculated using phases obtained from molecular replacement was used to locate 8 of the 11 selenium sites that further aided in assigning the sequence for σ^{54} model. β' G/G' insertion domain was manually placed in the difference density map using the crystal structure 2AUK. Initial stages of refinement were carried out in REFMAC5 (*40*) using secondary structure restraints, restraints for core RNAP to known structure (PDB code: 4IGC) and jelly-body restraints. Later stages of refinement were carried out in BUSTER-TNT (*41*) using target restraints to known structures. All data processing and refinement statistics are summarised in Table S1.

pBpa-based UV crosslinking assays

We utilized a non-natural, UV-crosslinkable amino acid Bpa incorporated into RNAP subunits (developed in the Grouse lab) to map interactions between RNAP and σ^{54}. An equimolar ratio of *p*Bpa-RNAP and HMK tagged ^{32}P - σ^{54} (holoenzyme 0.3 μM) was incubated at 37 °C for 10 min, irradiated at 365 nm on ice for 30 min and analysed on a 4-12% SDS gradient gel. The crosslinked products were scanned by Typhoon PhosphorImager and analysed by Aida software.

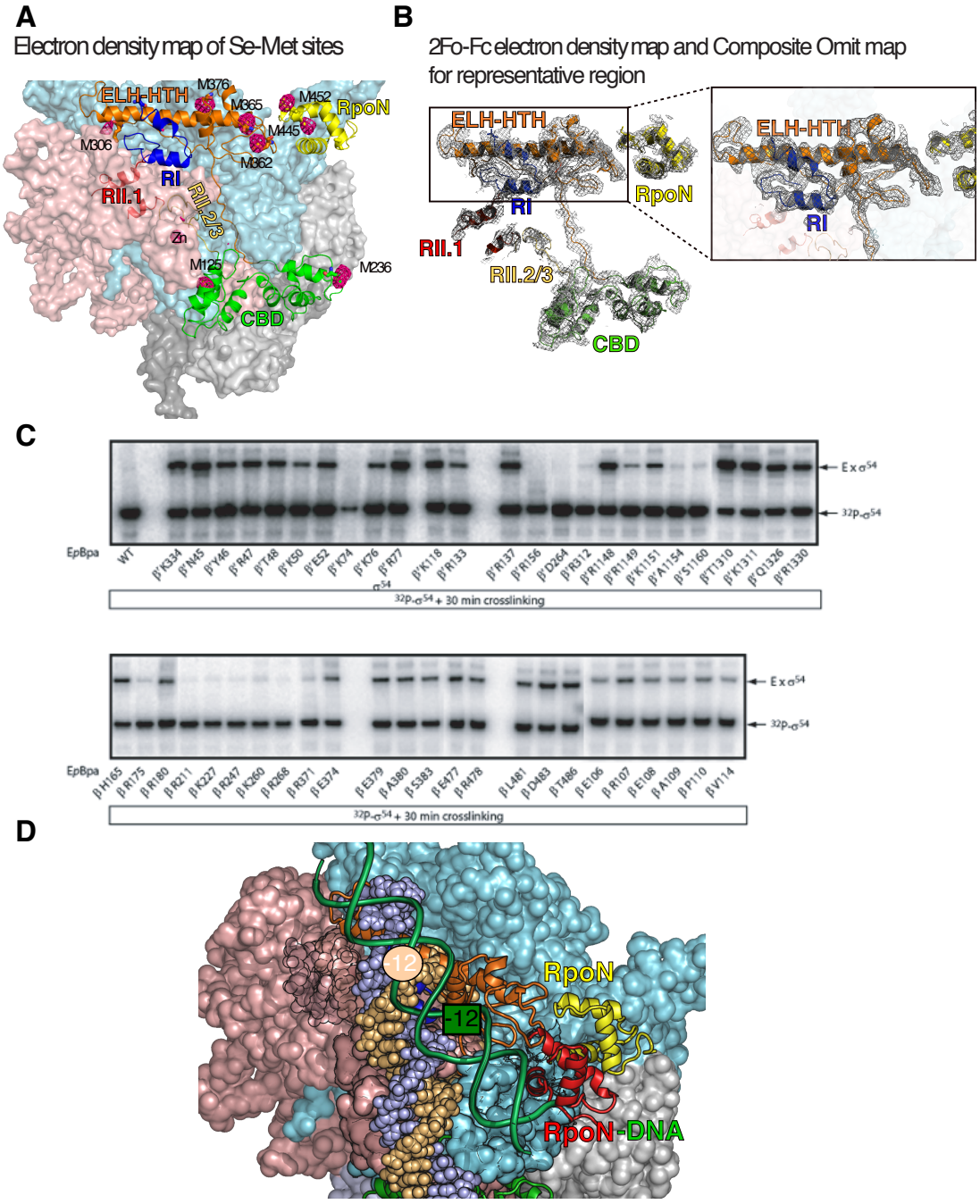


Fig. S1. Electron density and crosslinking data to determine the structure of σ^{54}-RNAP holoenzyme . (**A**) Anomalous electron density map (4σ, magenta mesh) of RNAP-σ^{54} holoenzyme containing Se-Met derivitized σ^{54} showing 8 out of 11 Se-Met sites (labeled) in σ^{54}, (**B**), 2Fo-Fc electron density map and composite omit map (1.5σ) of representative σ^{54} region, (**C**), RNAP to ^{32}P-σ^{54} crosslinking data. (**D**) Overlay of RNAP-σ^{54} and RNAP-σ^{54}-DNA models showing RpoN positions. -12 promoter element in the structure (green) is too far upstream compared to the -12 promoter element in the initiation complex (4YLN).

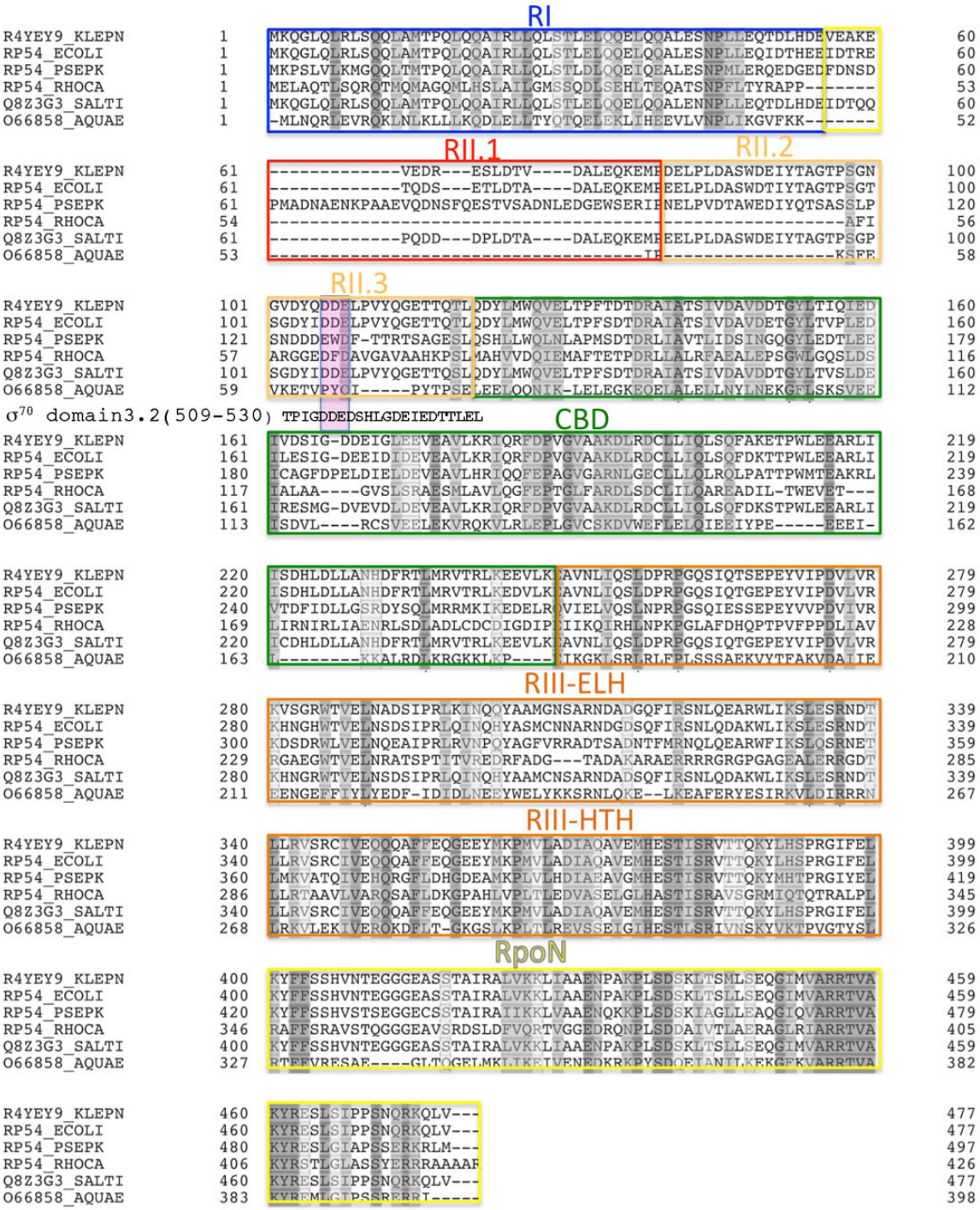


Fig. S2. Sequence alignment of σ^{54} from different species. Functional domains are boxed and colored coded as in Fig. 1. Sequences that are identical or similar are shaded in dark or light grey respectively. CBD and RpoN are highly conserved among σ^{54} from different species. KLEPN – *K. pneumonia,* ECOLI – *E.coli,* PSEPK – *P. putida,* RHOCA – *R. capsulatus*, SALTS - *S. typhimurium,* AQUAE – *A. aeolicus*

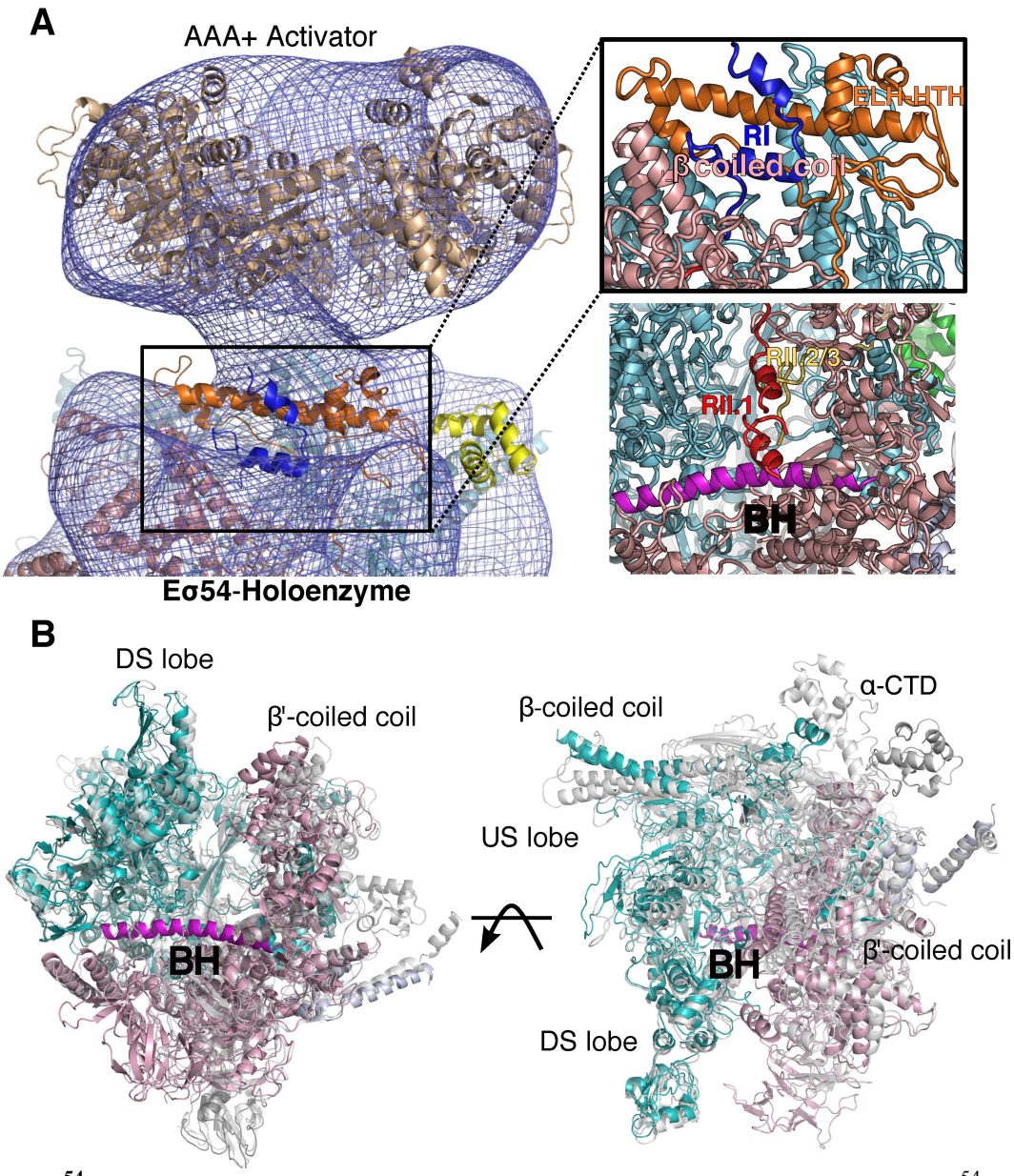


Fig. S3 σ^{54} key regions and comparison of holoenzymes. (**A**) Fitting of RNAP-σ^{54} holoenzyme crystal structure into the cryo-EM reconstruction of AAA+ activator bound holoenzyme. Right panels show enlarged view of region within the connecting density. Also shown is RII.1 in relation to the bridge helix (BH). (**B**) Overlay of RNAP core enzymes in σ^{70}-holoenzyme (grey) and σ^{54}-holoenzyme (color coded as in Fig. 1).

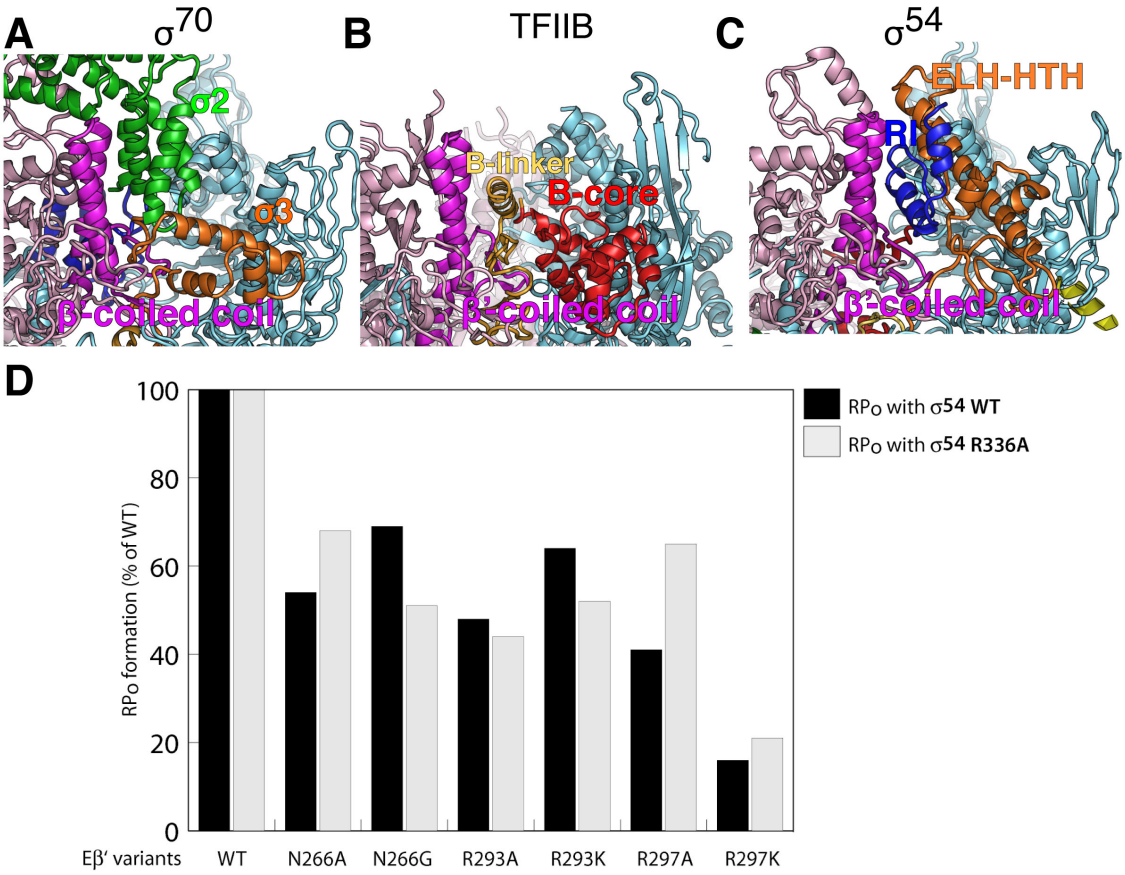


Fig. S4 Comparison of β'-coiled coil interactions. (**A**), β' coiled-coil forms a helical bundle with σ_2 of σ^{70} and B-linker of TFIIB (**B**), (**C**) β'-coiled coil does not form a helical bundle with σ^{54} RI-RIII-ELH. Instead it only interacts weakly at the tip of ELH. (**D**) R336A mutant within the ELH has different effects on open complex formation compared to mutations in β' coiled-coil, in agreement with the suggestion that interactions between RI and RIII-ELH which involves R336, and between RIII-ELH and β'-coiled coil, which does not involve R336, might form in different functional states, during open complex formation.

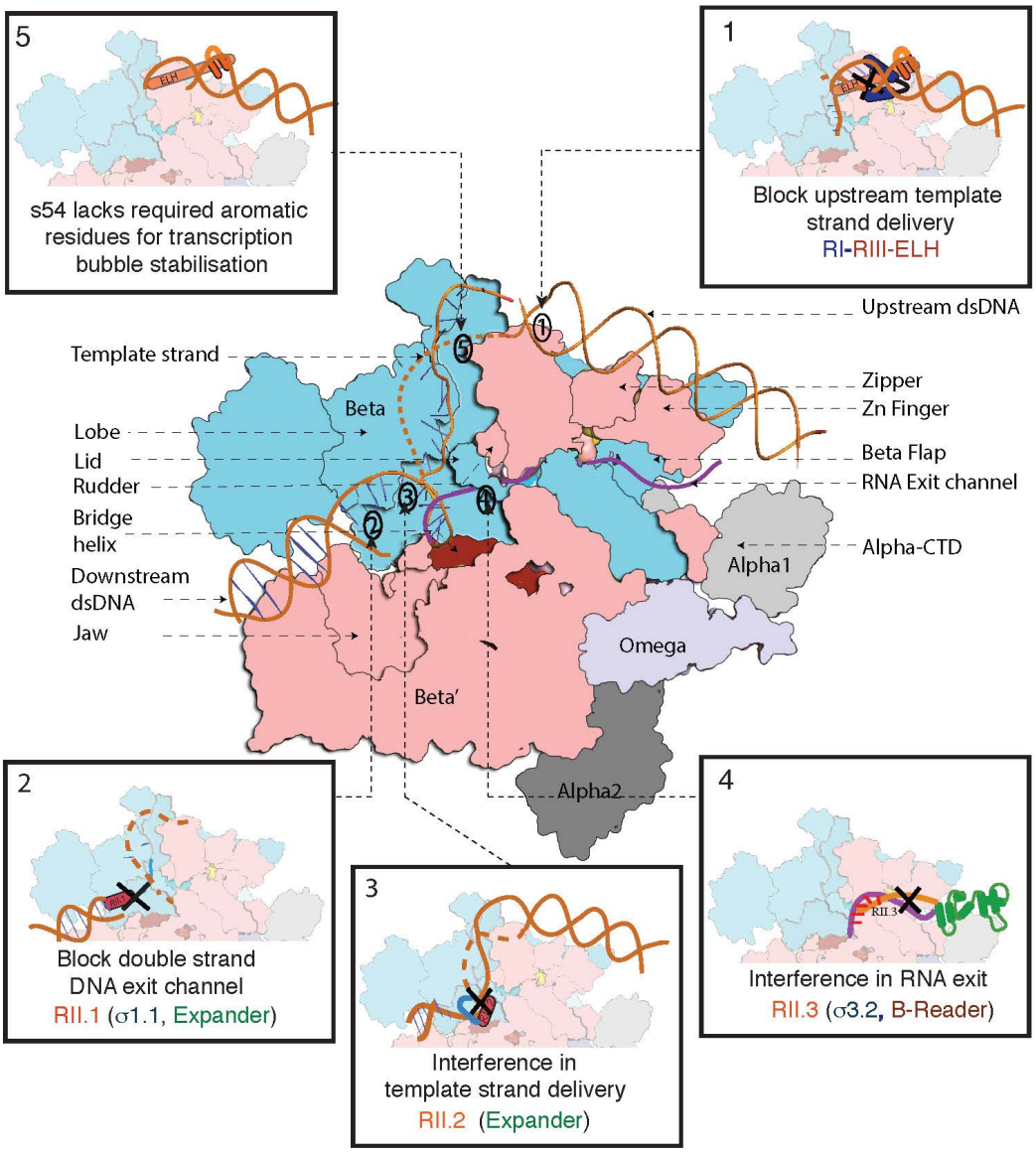


Fig. S5. Common regulatory hotspots. σ^{54} colour coded according to Fig. 1. Each box details the regulatory elements in σ^{54} and those shared by others are in brackets (dark blue: σ^{70}, green: Pol I, brown: TFIIB). There are at least five such hot spots uncovered in σ^{54}-RNAP holoenzyme. The RNAP cleft is capped by the lid formed by RI and RIII ELH-HTH of σ^{54} (box 1). The downstream DNA channel is another highly regulated site (Fig. box 2). RII.1 of σ^{54}, σ1.1 of σ^{70} and the expander in Pol I all occupy this space to regulate. The template strand path inside the RNAP cleft is interfered by RII.2 of σ^{54} and the expander in Pol I (box 3). A highly conserved regulatory spot is the RNA exit tunnel. RII.3 of σ^{54}, σ3.2 of σ^{70}, B-reader of TFIIB (*43*) (box 4). Furthermore, the number of aromatic residues along the non-template strand path could also fine-tune the ability to stabilise the transcription bubble (box 5), with σ^{70} acting independently of additional factors and σ^{54} requiring its ATP-driven activator.

Table S1
Summary of X-ray data collection and refinement statistics

	Native Crystal	Semet Crystal
Data collection		
Space group	$P2_12_12$	$P2_12_12$
Cell dimensions		
a, *b*, *c* (Å)	208.48, 151.52, 195.28	208.11, 152.15, 194.38
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	29.8 – 3.76 (3.89-3.76)*	91.74 – 3.98 (4.08 – 3.98)
R_{merge}	0.082 (0.66)	0.071 (0.54)
R_{meas}	0.087 (0.75)	0.087 (0.66)
R_{pim}	0.042 (0.45)	0.05 (0.38)
$CC_{1/2}$ (%)	99.9 (84.5)	96.6 (84.7)
$I/\sigma I$	14.1 (1.9)	13.3 (2.8)
Completeness (%)	98.4 (91.2)	99.8 (99.8)
Redundancy	8.3 (4.7)	5.3 (5.5)
Refinement		
Resolution (Å)	29.64 – 3.76	
No. unique reflections	Above 0 : 62409 Above 1σ: 60662 Above 3σ: 39123	
No. reflections	62409	
No. test reflections	3077	
R_{work}/ R_{free}	0.33/0.34	
No. atoms		
Protein	22647	
Ligand/ion	1	
Water	-	
Wilson B-factor (Å^2)	119	
Average B-factors (Å^2)		
Protein	200	
Ligand/ion	99.5	
Water	-	
R.m.s deviations		
Bond lengths (Å)	0.007	
Bond angles (°)	0.99	
MolProbity Statistics		
All atom clash score	0.3	
Rotamer outliers (%)	3.27	
Cβ deviations>0.25 Å	0	
Ramachandran Plot		
Favoured region (%)	85.5%	
Outliers (%)	3.2%	

*Highest resolution shell is shown in parenthesis.

References:

31. Y. Aiba, Y. Katayama, T. Hishinuma, H. Murakami-Kuroda, L. Cui, K. Hiramatsu, Mutation of RNA polymerase beta-subunit gene promotes heterogeneous-to-homogeneous conversion of beta-lactam resistance in methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother* **57**, 4861-4871 (2013).
32. M. J. Merrick, J. R. Gibbins, The nucleotide sequence of the nitrogen-regulation gene ntrA of Klebsiella pneumoniae and comparison with conserved features in bacterial RNA polymerase sigma factors. *Nucleic Acids Res* **13**, 7607-7620 (1985).
33. N. Joly, N. Zhang, M. Buck, ATPase site architecture is required for self-assembly and remodeling activity of a hexameric AAA+ transcriptional activator. *Molecular cell* **47**, 484-490 (2012).
34. L. Morris, W. Cannon, F. Claverie-Martin, S. Austin, M. Buck, DNA distortion and nucleation of local DNA unwinding within sigma-54 (sigma N) holoenzyme closed promoter complexes. *J Biol Chem* **269**, 11563-11571 (1994).
35. W. V. Cannon, M. T. Gallegos, M. Buck, Isomerization of a binary sigma-promoter DNA complex by transcription activators. *Nat Struct Biol* **7**, 594-601 (2000).
36. W. Kabsch, Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132 (2010).
37. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).
38. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica* **66**, 213-221 (2010).
39. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica* **66**, 486-501 (2010).
40. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica* **67**, 355-367 (2011).
41. E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea, G. Bricogne, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta crystallographica* **60**, 2210-2221 (2004).
42. M. Chlenov, S. Masuda, K. S. Murakami, V. Nikiforov, S. A. Darst, A. Mustaev, Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase beta' subunit. *J Mol Biol* **353**, 138-154 (2005).
43. B. A. Knutson, J. Luo, J. Ranish, S. Hahn, Architecture of the Saccharomyces cerevisiae RNA polymerase I Core Factor complex. *Nat Struct Mol Biol* **21**, 810-816 (2014).