

Imperial College of Science, Technology and Medicine

Department of Medicine

Structural studies of σ^{54} -mediated transcription initiation

Forson Gao

March 2023

Submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy of Imperial College London and

the Diploma of Imperial College London, March 2023

Statement of Originality

The work presented in this thesis is my own unless stated otherwise. The contributions of others are noted where relevant

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Transcription is an essential step in gene expression in all organisms and involves the conversion of sequential information from genomic DNA to single-stranded RNA by the enzyme RNA polymerase. Transcription initiation is the most regulated step in transcription and is essential in ensuring an organism can adapt and respond to changes in its environment and maintain homeostasis. In bacteria, transcription initiation is mediated by σ factors, which are classified as housekeeping σ^{70} family or major variant σ^{54} . σ factors recruit RNA polymerase to upstream promoter regions of genes and regulate the ability for RNA polymerase to melt and load DNA into its active site to form a competent open complex.

σ^{54} promoters are associated with nutrient depletion and stress response, and σ^{54} is unique as it recruits RNA polymerase to form a stable closed complex. σ^{54} initially inhibits open complex formation by blocking DNA loading, as well as occupying the active site and sitting atop of the RNA exit channel. Open complex formation therefore requires ATP hydrolysis by a specialised family of clade 6 AAA+ (ATPase associated with diverse cellular activity) proteins called bacterial enhancer binding proteins. Furthermore, σ^{54} must undergo a set of conformational changes to enable promoter escape and a transition to processive elongation.

We used biochemical approaches coupled with cryo-electron microscopy to capture multiple AAA+-engaged intermediate states to elucidate the mechanism of early DNA melting during open complex formation. Furthermore, we have

determined high-resolution structures capturing the initial stages of transcription, gaining insights into how σ^{54} is displaced during promoter escape.

The work presented here deepens our understanding of σ^{54} mediated transcription initiation, demonstrates conserved mechanisms of promoter escape in all bacteria and demonstrates a unique mechanism of transcription activation by a specialised family of AAA+ protein.

Acknowledgements

I have thoroughly enjoyed this PhD project, and it has been a personally challenging but deeply rewarding experience. I am extremely grateful to all the people that have given me encouragement, advice and support. First of all, I would like to thank Xiaodong Zhang and Martin Buck for giving me the opportunity to work on this amazing project and being a great source of guidance, encouragement and insight. I love this project and have learned so much, and I am forever grateful for this opportunity. I would also like to thank my progress review panel Harry Low, Avinash Shenoy and Ramesh Wigneshweraraj, who gave me great advice during my PhD. I would like to thank all current and former members of the Zhang and Buck labs, the structural and synthetic biology section and collaborators for their friendship and comradery, especially Fuzhou, who introduced me to the lab. Luke, Bilge, Lucas, Jie, Steph, Alex, Bowen, Pengtao, Meghan, Katie, Josh, Rhys, Amy, Qiao, Xiaojiao, Min, Monica, Rafa, Thomas, Milija, Ioly, George, Erik, Soo Mei, Liz, Edina, and Ciaran are all awesome people who are lifesavers in the lab, phenomenal scientists to bounce ideas off, have a laugh with and occasionally play badminton with.

I am super, super lucky to have learned from and worked with so many kind people during my time at Imperial and Nanjing. In chronological order: Huang Cheng, who gave me my first lab experience, and showed me the excitement and joy of working in the lab. Kurt Drickamer and Maureen Taylor, who instilled in me the principles of scientific reproducibility and rigour. I really want to express my gratitude to Morgan Beeby and Josie Ferreira. I am extremely grateful they took me in, taught me the theory behind cryo-EM and gave me hands on experience. They encouraged me to

apply for a PhD despite many rejections, and are extremely kind, patient and understanding mentors and my good friends.

I also want to thank Jeroen Claus and the team at Phospho Biomedical animation, who taught me molecular animation in SideFX Houdini, and let me work in their offices during the pandemic. I still have our small subunit processome cover submission as my desktop background, and I am still in awe of all the images and videos – they are some of the most beautiful artwork I have ever seen.

Many thanks to all the staff in the cryo-EM and wider scientific community - the LonCEM and CBEM microscope facilities, chemistry stores, technicians, cleaners, programmers, ICT and eBIC staff, in particular Nora, Paul, Yuriy, Alistair, Dan, Vinod, Suhail, Sjors, Danny, Andrea and many more who work incredibly hard to support the community. Cryo-EM, computing and science would be impossible without them, and I am very grateful that they responded to all my requests and took the time to help me make the most out of my samples and data.

To my family – Magson, who gave me his laptop which I used throughout my PhD, gave me his bike so I could commute to work, gave me a phone to communicate with and gave me a coat so I can stay warm during winter – thank you. Most of all, thank you for being there at every step in my life, so that I was never alone. Dad, for paying off all my student loan debt and financially supporting me, giving me the luxury to never have to worry about debt or finances. Jovia, for giving me a sofa to sleep on, strongly encouraging me to take breaks, providing free medical advice and being a rock of emotional support during every tough moment. This thesis is dedicated to Mum, who I owe everything to and gives everything to nurture and support me. I love you all.

Contents

Statement of Originality	i
Copyright declaration	i
Abstract.....	ii
Acknowledgements.....	iv
Contents.....	vi
List of Figures	xii
Abbreviations	xxi
List of Tables	xxiv
Introduction	25
1.1 Transcription and the central dogma of molecular biology.....	25
1.2 The stages of transcription	30
1.3 Bacterial RNA polymerase	33
1.3.1 α subunit	34
1.3.2 ω subunit	35
1.3.3 β and β' subunits	35
1.4 Comparison with Archaeal and Eukaryotic RNA polymerases	40
1.5 Conserved functions of RNA polymerase	43
1.5.1 Nucleotide addition.....	43
1.5.2 Pausing.....	45
1.6 σ factors	55

1.6.1 σ^{70} family	56
1.6.2 σ^{54} factor	64
1.7 ATPases associated with various cellular activities (AAA+)	68
1.7.1 Bacterial Enhancer binding proteins	72
1.7.2 Conserved AAA+ motifs in the C domain.....	74
1.8 σ^{54}-mediated open complex formation.....	76
1.9 Mechanisms of promoter escape	78
1.9.1 RNA 5' displacement.....	80
1.9.2 DNA scrunching	81
1.10 Project Aims	83
Cryo-electron Microscopy	84
2.1 Background	84
2.2 Resolution limits	85
2.3 Amplitudes, phases and Fourier slice theorem	88
2.4 Electron microscope anatomy.....	89
2.4.1 Vacuum source	90
2.4.2 Electron sources and coherence.....	90
2.4.3 Electron lenses	91
2.4.4 Cryobox and sample holders.....	92
2.4.5 Energy filters.....	92
2.4.6 Detectors and Direct Quantum Efficiency (DQE)	93
2.5 Image formation	95
2.5.1 Scattering and contrast	95
2.5.2 Spherical and Chromatic aberration.....	97
2.5.3 Contrast transfer function estimation and correction	98
2.6 Cryo-EM sample preparation	100

2.7 Data collection	102
2.7.1 Pixel size and Nyquist frequency	102
2.7.2 Nominal Defocus	103
2.7.3 Data acquisition rate and exposure time	103
2.8 Image Processing	105
2.8.1 Motion Correction	105
2.8.2 Particle picking	105
2.8.3 Alignment.....	107
2.8.4 Maximum likelihood approaches for refinement.....	107
2.8.5 2D and 3D classification.....	109
2.8.6 Fourier shell correlation	110
2.8.7 Model building	111
Materials & Methods	113
3.1 Plasmid transformation	113
3.2 Cell culture and induction	115
3.2.1 σ^{54} and PspF ₁₋₂₇₅ induction and expression	115
3.2.2 RNAP core ($\alpha_2\beta\beta'\omega$) induction and expression	116
3.2.3 Harvesting	116
3.3 Cell lysis	116
3.4 Protein purification	117
3.4.1 σ^{54} wildtype and R336A mutant purification	117
3.4.2 PspF ₁₋₂₇₅ purification.....	118
3.4.3 RNAP core ($\alpha_2\beta\beta'\omega$).....	119
3.4.4 RNAP- σ^{54} formation	120
3.5 Complex formation assays	121
3.5.1 Nucleic acid scaffold preparation	121
3.5.2 Native-PAGE	121

3.6 Complex formation	122
3.6.1 RP(-12/-11), RP(-11/-8) and RP(-11/-3) complexes.....	122
3.6.2 RP(-10/-1) complexes	123
3.6.3 RP5nt, 6nt, 7nt, 8nt and 9nt complexes	123
3.7 Cryo-EM sample preparation	123
3.7.1 Sample preparation of RP(-11/-8) and RP(-11/-3) complexes.....	124
3.7.2 Sample preparation of RP(-10/-1), RP5nt, 6nt, 7nt, 8nt and 9nt complexes	124
3.7.3 Data collection	124
3.7.4 Image processing	125
3.7.5 Model Building and refinement.....	125
3.8 In vitro transcription assays	125
Structural basis of σ^{54}-mediated open complex formation.....	127
4.1 Introduction	127
4.2 High resolution structure of closed and PspF₁₋₂₇₅-engaged intermediate complexes	128
4.3 Analysis of RPi(-12/-11) PspF₁₋₂₇₅ protomer coordination and interactions	133
4.4 Purification of intermediate complex protein components	142
4.4.1 RNA polymerase	142
4.4.2 σ^{54}	143
4.4.3 RNAP- σ^{54} holoenzyme	143
4.4.4 PspF ₁₋₂₇₅	144
4.5 Optimisation of complex assembly	145
4.5.1 Choice of DNA scaffolds	145
4.5.2 Native-PAGE and longer transcript gels of RNAP- σ^{54} -DNA-PspF complexes	145
4.6 RP(-11/-8) complexes.....	148
4.6.1 Sample Preparation of the RP(-11/-8) complex	148

4.6.2 Image processing of RP(-11/-8) complexes	149
4.6.3 Structural analysis of RP(-11/-8)	157
4.7 RP(-10/-1) complex	164
4.7.1 Sample preparation	164
4.7.2 Image processing of RP(-10/-1)	165
4.7.3 Structural analysis of RP(-10/-1)	168
4.8 Summary	172
Structural basis of σ^{54} displacement in Bacterial Promoter Escape	173
5.1 Introduction	173
5.2 RP5-9nt complexes	174
5.2.1 5nt complexes reveal a novel passage of region II through the RNAP core	176
5.2.2 Longer transcripts causes scrunching of the region II finger	182
5.2.3 Transcription extension drives increased perturbation of transcription bubble and extra-long helix densities.	185
5.3 Summary	187
Discussion	190
6.1 Mechanisms of σ^{54} mediated, PspF₁₋₂₇₅-driven open complex formation	190
6.2 σ^{70} mediated open complex formation comparison	192
6.3 Mechanism of bEBP hydrolysis and coordination	196
6.4 Comparison of bEBPs with other AAA+ proteins	198
6.5 σ^{54}-mediated promoter escape and comparison with σ^{70}	199
6.6 Eukaryotic perspectives	202
6.7 Summary and Future Perspectives	205
Bibliography	208

Appendix 246

List of Figures

Figure 1.1.1: An overview of the central dogma of molecular biology.	26
Figure 1.1.2: a) Chemical structure of nucleoside triphosphates.	28
Figure 1.3.1: Overview of the RNAP. Left – Downstream face of RNAP.	34
Figure 1.3.2: a) Location of the DPBB domains in the RNAP claw in <i>E.coli</i> RNAP (PDB ID: 6ALH).	37
Figure 1.3.3: a) Structure of RNAP elongation complex showing transcription bubble and DNA-RNA hybrid interactions with the RNAP core showing the β' (left) and β (right) interactions (Adapted from PDB ID 6XDR).	38
Figure 1.3.4: a) Switch regions in elongation complex.	40
Figure 1.4.1: Comparison of bacterial, archaeal, and eukaryotic RNAPs (<i>S. cerevisiae</i> used as an example). Key is shown in Table 1.4.1.	42
Figure 1.5.1: a) Schematic diagram of the nucleotide addition cycle. Adapted from Nudler (2009). b) Mechanism of nucleophilic attack from incoming rNTP during nucleotide addition. Adapted from Roston, Demapan & Cui (2019).	43
Figure 1.5.2: A comparison of pre-translocated, paused and post-translocated intermediates.	48
Figure 1.5.3: Comparison of <i>E.coli</i> RfaH and canonical NusG structures.	50
Figure 1.5.4: a) An overview of DSIF bound to RNAPII in eukaryotes.	52

Figure 1.5.5: An overview of TFIIIS (a) and GreB (b) bound paused structures, showing insertion of acidic tip residues into the active site via the secondary channel to promote cleavage.....	55
Figure 1.6.1: Overview of σ^{70} family members and binding sites. Disc=discriminator element, σF = σ finger.	56
Figure 1.6.2: a) Domain map with residue range for subdomains of <i>E. coli</i> σ^{70}	57
Figure 1.6.3: σ^{70} mediated open complex formation. a) an overview of the stepwise bubble opening in σ^{70} structures (PDB IDs: 6PSQ, 6PSR, 6PSS, 6PST, 6PSU, 6PSV, 6PSW).	59
Figure 1.6.4: Sequence logo of σ^{54} promoters taken from iPro54 database (Lin et al., 2014).	65
Figure 1.6.5: a) Domain map of σ^{54} . b-d) Structure of <i>E. coli</i> σ^{54} closed complex (5NSR).	67
Figure 1.7.1: a) structure of AAA+ interprotomer interface (PDB ID: 3M0E) showing <i>in trans</i> activity of R-fingers from adjacent subunit and ATP binding site within the α -lid and $\alpha\beta\alpha$ sandwich.....	69
Figure 1.7.2: Classical mechanism of AAA+ driven substrate translocation using a sequential hand over hand mechanism.	71
Figure 1.7.3: a) Schematic diagram of clade 6 AAA+ protein secondary structure. b) Model of PspF ₁₋₂₇₅ hexamer showing ATP binding sites. Model was created by fitting in ATP molecules from the PspF ₁₋₂₇₅ structure (2C96) onto the intermediate complex structure (5NSS). c) PspF ₁₋₂₇₅ hexamer contacting DNA and σ^{54} -RI via its L1 loops. d)	

NtrC1 monomer crystal structure (PDB ID: 3M0E) showing the conserved domains found in AAA+ proteins, as well as the L1/L2 loops. Area highlighted is the active site, as enlarged in e). e) The active site of NtrC1. The E108 residue was modelled using PspF₁₋₂₇₅ crystal structure (2C96). The residues are numbered as in *E. coli* PspF₁₋₂₇₅. Figure adapted from Gao et al. (2020). 74

Figure 1.8.1: Structures in the open complex formation pathway. a) Overall structures of the different RPc (5NSR), RPi (5NSS), RPip (6GH6), RPo (6GH5), RPitc (6GFW) and RPe (taken from expressome structure 6XDR) complexes, viewed from the downstream face. 76

Figure 1.9.1: Schematic mechanism of RNA displacement. 80

Figure 1.9.2: Schematic mechanism of DNA scrunching driven σ displacement. 81

Figure 1.10.1: Summary of the structural transitions being studied in this project. . 83

Figure 2.1.1: Schematic of 3D reconstruction method, demonstrating the principles of Fourier slice theorem - Fourier transforms of 2D projection images represent slices of a 3D Fourier transform. 85

Figure 2.2.1: Overview of wave equation. All waves contain amplitude, phase and wavelength. 86

Figure 2.2.2: a) Incident wave hitting an array of evenly spaced points. 87

Figure 2.3.1: Schematic representation of a double Fourier system. 88

Figure 2.4.1: Anatomy of the cryo-electron electron microscope. 89

Figure 2.4.2: a) Overall structure of CCD chip. Incoming electrons must first enter a scintillator, which causes the electrons to scatter and emit photons. 94

Figure 2.5.1: a) Schematic of scattering in an microscope.	98
Figure 2.5.2: a) Spherical and b) chromatic aberrations create an envelope that dampens the CTF.....	99
Figure 2.6.1: Schematic diagram of a holey grid.....	100
Figure 2.8.1: a) Single particle refinement involves aligning a set of noisy extracted particles with projection images from a reference.b i) Angular assignments are determined by calculating the probability of angles being correctly assigned. b ii) An alternative method of alignment is the maximum cross-correlation (CC) approach, which is suboptimal for noisy images. c) The aligned particles are then back-projected onto a low-pass filtered reference to produce the refined model, which can then be re-projected to produce improved reference projections, and the cycle is iterated until the accuracy in angular assignments converge.....	108
Figure 2.8.2: Fourier shell correlation measures the cross-correlation of the Fourier components of 2 independent halves of the refinement map. The correlation between each spatial frequency is then plotted and the nominal resolution is determined by finding the resolution at the correlation cutoff of 0.143.	111
Figure 4.1.1: Schematic of σ^{54} mediated open complex formation.	128
Figure 4.2.1: a) Domain map of σ^{54} . b) Model and c) composite map of RPi(-12/-11). d) Model and e) map of RPc(WT)	130
Figure 4.2.2: σ^{54} interaction with the <i>nifH</i> promoter DNA reveals nucleation of the transcription bubble by region I and ELH, stabilised by major groove interactions via region III-RpoN and HTH.....	131

Figure 4.2.3: a) PspF ₁₋₂₇₅ ring highlighting arrangement of L1 loops contacting RNAP- σ^{54} -DNA.	132
--	-----

Figure 4.3.1: a) overview of the PspF ₁₋₂₇₅ hexameric ring side and top views b) secondary structure schematic of PspF ₁₋₂₇₅	135
---	-----

Figure 4.3.2: a) overview nucleotide binding state in the hexamer. b-e) alignment of the protomers by the $\alpha\beta$ sandwich subdomain.	137
--	-----

Figure 4.3.3: The proposed conformational changes within the hexameric ring....	138
---	-----

Figure 4.3.4: Proposed mechanism of ATP hydrolysis and substrate translocation. a) Cartoon schematic of movement of protomer 1 during the power stroke leads to dissociation from protomer 2 and shifting of protomer 6 towards protomer 5, facilitating protomer 2 L1 loop opening, and nucleotide release by protomer 6. b) Structural schematic showing direction of translocation facilitated by ATP hydrolysis. c) Cartoon schematic showing the movement of protomers down the central axis of PspF ₁₋₂₇₅ during ATP hydrolysis.	140
--	-----

Figure 4.4.1. a) Elution profile of RNA polymerase from final anion exchange chromatography step.	142
--	-----

Figure 4.4.2. a) Elution profile of σ^{54} from final gel filtration step. Shaded region in the elution profile represent pooled fractions, as shown in bold in the SDS-PAGE profile in b).	143
---	-----

Figure 4.4.3. a) Elution profile of σ^{54} from final gel filtration step.	144
--	-----

Figure 4.4.4. a) Elution profile of PspF ₁₋₂₇₅ from final gel filtration step. Shaded region in the elution profile represent pooled fractions, as shown in bold in the SDS-PAGE profile in b).	145
---	-----

Figure 4.5.1. Native-PAGE gel of RNAP- σ^{54} -PspF ₁₋₂₇₅ complexes trapped with ADP.AIFx and varying DNA mismatch lengths (lanes 4-10)	146
Figure 4.5.2. a) Design of <i>nifH</i> DNA sequences.	147
Figure 4.6.1: Gel filtration profile of RP(-11/-8) and SDS-PAGE analysis of fractions.	148
Figure 4.6.2: Overall processing flowchart for RP(-11/-8) complexes.....	149
Figure 4.6.3: a) Pre-processing and 1st round of classification and refinement showed features of RNAP, σ^{54} , DNA and PspF ₁₋₂₇₅	150
Figure 4.6.4: a) Flowchart of PspF ₁₋₂₇₅ -DNA- σ^{54} region I subtraction processing pipeline.....	152
Figure 4.6.5: Angular distribution map of the global (i) RNAP- σ^{54} -DNA (ii) and PspF ₁₋₂₇₅ -DNA-region I (iii)-focused refinement maps of conformations 1-6 (a-f).	153
Figure 4.6.6: FSC curves of the global (i), RNAP- σ^{54} -DNA (ii), and PspF ₁₋₂₇₅ -DNA-region I (iii) focused refinement maps of conformations 1-6 (a-f).....	154
Figure 4.6.7: a) Flowchart of RNAP- σ^{54} -DNA subtraction processing pipeline.....	155
Figure 4.6.8: a) Flowchart of PspF ₁₋₂₇₅ subtraction processing pipeline.....	156
Figure 4.6.9: a) Clamp opening across different composite maps of the 6 conformations.	157
Figure 4.6.10: A comparison of the a) PspF ₁₋₂₇₅ tilt angles and b) DNA bending downstream of the -10 position for RPi(-12/-11) and RP(-11/-8).	159

Figure 4.6.11: RPi(-12/-11) and RP(-11/-8) aligned by PspF ₁₋₂₇₅ ring reveals differences in protomer arrangement.....	160
Figure 4.6.12: Nucleotide occupancy changes between protomers 5 and 6. a-b) Overview of protomer arrangement and occupancy. Zoomed in views of protomers 5 and 6 are highlighted red and blue respectively. c) Zoomed in view of the active site of protomers 5 (top panels) and 6 (bottom panels), showing protomer 5 becoming nucleotide bound, whereas protomer 6 becomes <i>apo</i> . d-e) Alignment of protomers by the walker A motif shows that from -12/-11 → -11/-8 is associated with slight opening in protomer 5 (d) and scrunching and closing of the α lid and $\alpha\beta$ sandwich subdomains in protomer 6 (e).	161
Figure 4.6.13: a) View of the RPi(-12/-11) complex from the β face. Zoomed in area in b) is highlighted. b) In the RP(-11/-8) conformation 3 structure, the -11G base on the template strand flips out and faces the ELH.	162
Figure 4.6.14: Overview of potential RpoN and protomer 4 interactions in the RP(-11/-8) dataset. a) Zoomed out view of RPi(-12/-11).	163
Figure 4.7.1: Overall processing workflow for RP(-10/-1) dataset.	165
Figure 4.7.2: a) Pre-processing and 1st round of classification and refinement of the RP(-10/-1) dataset showed features of RNAP, σ^{54} , DNA and PspF ₁₋₂₇₅	166
Figure 4.7.3: i) Local resolution maps ii) angular distribution maps and iii) Fourier-shell correlation curves of a) open clamp conformation	167
Figure 4.7.4: a) The PspF ₁₋₂₇₅ tilt for RP(-10/-1) conformation is the same angle as RP(-11/-8) conformation 3. b) the DNA bending angle of the RP(-10/-1) is 73°, a sharper bend corresponding to a larger mismatch length and bubble size.	168

Figure 4.7.5: a) DNA melting corresponds with changes in the region I N-terminus position, which suggests a pulling and potential unfolding of the $\alpha 1$ helix (b) and protomer arrangement of the PspF ₁₋₂₇₅ ring (c).	170
Figure 4.7.6: Comparison of the RPo(WT) structure with the previously published RPip(R336A) and RPo(R336A) mutant structures.	171
Figure 4.8.1: Schematic of open complex formation	172
Figure 5.2.1: Re-assignment of RNA density in RPitc model.....	174
Figure 5.2.2: Design of the DNA-RNA scaffolds used in this investigation.....	175
Figure 5.2.3: Processing pipeline for the RP5nt dataset.....	176
Figure 5.2.4: Comparison of region II in RPo, RPitc and RP5nt (pre-translocated) complexes.....	177
Figure 5.2.5: a) Local resolution maps. b) Angular distribution maps.	179
Figure 5.2.6: FSC maps.	180
Figure 5.2.7:a) Left: overview of RP5nt structure downstream face.	181
Figure 5.2.8: Processing pipeline for the RP6nt dataset.....	182
Figure 5.2.9: Processing pipeline for the RP7nt dataset.....	182
Figure 5.2.10: Processing pipeline for the RP8nt dataset.....	183
Figure 5.2.11: Processing pipeline for RP9nt dataset.....	183
Figure 5.2.12: RNA, template DNA and RII densities and conformational changes at different transcript lengths.	184
Figure 5.2.13: Increased RNA leads to loss of ELH density.....	186

Figure 5.3.1: Proposed mechanism of σ^{54} displacement.....	187
Figure 6.2.1: Comparison of between σ^{70} (a, b) and σ^{54} (c, d) upstream DNA binding elements.....	192
Figure 6.2.2: Comparison of subdomains binding to the internal cleft in σ^{70} (a, PDB ID:6PSQ) and σ^{54} (b, PDB ID: 5NWT).....	193
Figure 6.2.3: <i>M. tuberculosis</i> open complex formation involves open of the clamp to separate the switch region 2 (Sw2, orange) and fork loop 2 (FL2, cyan) in the final stages of DNA loading to enable template strand access to the catalytic site.	195
Figure 6.3.1: a) Classical mechanism. T= ATP-bound, D= ADP-bound, E=Empty.	197
Figure 6.5.1: Comparison of σ^{54} and σ^{70} (σ^A) initially transcribing complexes demonstrates that both utilise a similar mechanism of displacement of subdomains blocking the RNA exit channel, in which the 5' end of the growing RNA displaces σ factor, causing it to fold backwards.....	201
Figure 6.6.1: Comparison of eukaryotic and σ^{54} mechanisms in bubble stabilisation reveals common themes in bubble maintenance.....	204

Abbreviations

α	Alpha subunit
α NTD	Alpha-N-terminal domain
β	Beta subunit
β'	Beta prime subunit
λ	Wavelength
σ 54	Sigma-54
σ 70	Sigma-70
φ	Phase
ω	Omega subunit
A	Amplitude
A	Adenosine
Å	Angstrom
AAA+	ATPases associated with diverse cellular activities
ADP	Adenosine diphosphate
ADP.AIF _x	Adenosine diphosphate-aluminium fluoride (where x = 3 or 4)
ADP.BeF ₃	Adenosine diphosphate-Beryllium fluoride
AMPPNP	Adenosine 5'-(β,γ -imido) triphosphate
ATP	Adenosine triphosphate
ATPyS	Adenosine 5'-(3-thio)triphosphate
bEBPs	Bacterial enhancer binding proteins
bp	Base pair
C	Cytosine
c	Speed of light
cAMP	Cyclic Adenosine 5' monophosphate
CC	Coiled-coil
Cryo-EM	Cryo-electron microscopy
CTF	Contrast transfer function
D	DNA binding domain
DNA	Deoxyribonucleic acid
DS	Downstream
E	Extinction coefficient
EDTA	Ethylenediaminetetraacetic acid
ELH-HTH	Extra long helix-helix-turn-helix

EM	Electron microscopy
E σ^{54}	RNAP- σ^{54} holoenzyme complex
FRET	Förster resonance energy transfer
FT	Flowthrough
G	Guanine
HTH	Helix-turn-helix
IHF	Integration host factor
IPTG	Isopropyl β -D-1-thiogalactopyranoside
kDa	Kilodaltons
LB	Lysogeny broth
M	PageRuler Plus Prestained 10 to 250 kDa molecular weight ladder
mRNA	Messenger RNA
MWCO	Molecular weight cutoff filter
NIH units	National institute of health units
NTP	Nucleoside triphosphate
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
pH	Potential hydrogen
pI	Isoelectric point
PspF	Phage shock protein F
R	Regulatory domain
RII.1-RII.3	Region 2.1-Region 2.3
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPc	RNAP closed promoter complex
RPi	RNAP intermediate promoter complex
RPip	RNAP partially loaded promoter complex
RPitc	RNAP initially transcribing promoter complex
RPo	RNAP open promoter complex
SDS	Sodium dodecyl-sulfate
srRNA	Short regulatory RNA
T	Thymine
TCEP	Tris(2-carboxyethyl)phosphine
TF	Transcription factor

TGET buffer	Tris-glycerol-EDTA-TCEP buffer
tRNA	Transfer RNA
TSS	Transcription start site
U	Uracil
UAS	Upstream activating sequence
UpG	Uridylyl-2'-5'-phospho-guanosine
US	Upstream

List of Tables

Table 1.3.1: Length and molecular weight of <i>E. coli</i> RNA polymerase subunits.....	34
Table 1.4.1: RNAP subunit conservation between eukaryotic and prokaryotic organisms. Colours of cells correspond to colours in Figure 1.4.1.....	42
Table 1.6.1: Overview of σ^{70} family members (Paget, 2015; Roberts et al., 2017)....	56
Table 1.7.1: Classification of AAA+ protein clades. PS1BH=part of the Pre-sensor 1 β hairpin superclade. Adapted from Puchades et al. (2020) and Khan et al. (2022).	68
Table 1.7.2: An overview of the different groups of bEBPs. Abbreviations: RR = Response regulator, PRD = Phosphotransferase System Regulation Domain, PAS = Per-ARNT-Sim domain, V4R = 4-vinyl reductase, ACT = aspartokinase, chorismate mutase, and TyrA, GAF = cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA. (Reproduced with permission from (Gao et al., 2020).....	73
Table 3.1.1: Proteins used in this investigation and statistics.....	114
Table 3.1.2: List of plasmids used in this study and their corresponding genes. The molecular weight, number of residues, isoelectric point (pI) and extinction coefficient (E) can be found in Supplementary Table 1	114
Table 3.1.3: Buffer recipes	115
Table 3.5.1: Recipe for native-PAGE gels and loading buffers	122

Chapter 1

Introduction

1.1 Transcription and the central dogma of molecular biology

The central dogma of molecular biology describes the replication, transfer, and utilisation of sequential information in a biological system. Sequential information governs the structure and function of non-coding RNA and proteins, with the latter performing most enzymatic functions within living organisms. Proteins are made up of sequences of amino acid residues that are covalently linked by peptide bonds, resulting in a polypeptide chain (Crick, 1970; Yao, Wang & Chen, 2019).

Protein structure and function is determined by the sequence of residues and inter- and intra-molecular interactions between amino acid residues, solvent and ligands in 3D space (Figure 1.1.1) (Crick, 1970).

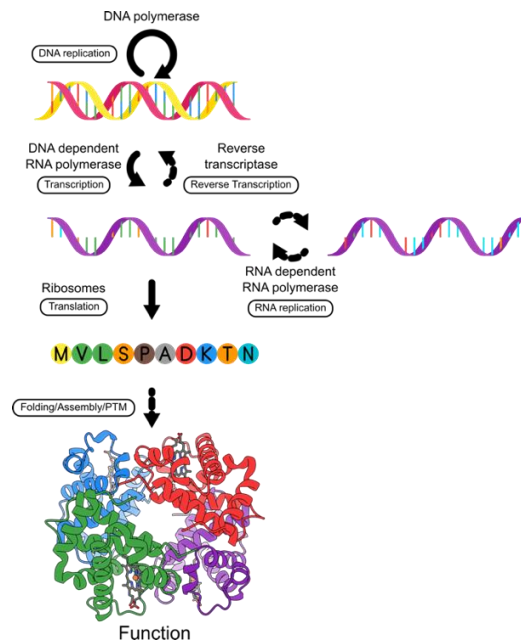


Figure 1.1.1: An overview of the central dogma of molecular biology. The central dogma describes the transfer of sequential information that enables the genome (genotype) to inform the function of proteins that catalyse reactions involved in biological processes within a cell or organism, which results in observable traits (phenotype).

The information required to make proteins is encoded in deoxyribonucleic acid (DNA) which is transcribed into ribonucleic acid (RNA) and subsequently translated into polypeptides that mature to form functional proteins. Transcription is the first step of gene expression and consequently plays a key role in regulating gene expression in response to environmental stimuli. Protein synthesis is one of the most energy intensive processes, and transcription regulation therefore plays an essential role in optimising energy usage (Li et al., 2014). Transcription initiation is the most regulated step in transcription and its regulation plays vital roles across all kingdoms of life (Malinen et al., 2022).

DNA is a right-handed double-stranded helix, with each strand consisting of a polymer of nucleotides connected by phosphodiester bonds. Each nucleotide

consists of the 2-deoxyribose-5-phosphate covalently bonded to a nitrogenous base via a glycosidic bond. There are 4 different nitrogenous bases in DNA; adenine (A), guanine (G) thymine (T) and cytosine (C). The structure of each type of base enables the formation of hydrogen bonds with their complementary base on the other strand, also known as a Watson-Crick base pair. A pairs with T, and C pairs with G (Franklin & Gosling, 1953; Watson & Crick, 1953; Wilkins, Stokes & Wilson, 1953).

Each DNA strand runs in antiparallel directions, designated $5' \rightarrow 3'$ based on the 5-phosphate and 3-hydroxyl group of the nucleotides (Figure 1.1.2). The DNA strand that is used as a template for ribonucleotides to base pair during transcription is called the template strand and the other DNA strand is the non-template strand (Riley, 1970; Watson & Crick, 1953).

Transcription is carried out by an enzyme called DNA dependent RNA polymerase (RNAP). RNAP is a claw-shaped protein complex that binds to DNA and helps facilitate DNA unwinding, often aided by a variety of transcription factors (Zhang et al., 1999). Transcription factors (TFs) are proteins that regulate RNAP activity by binding to upstream promoter regions, remote enhancer regions or to the RNAP complex (Stepanova et al., 2009; van Eeuwen et al., 2021). TFs form part of the transcription initiation complex machinery to either activate or repress transcription by modulating the recruitment of RNAP to promoter DNA, the accessibility of the template strand or NTPs to the RNAP catalytic site or modulating the ability of RNAP to recruit additional TFs or cofactors. Whilst some TFs are constitutively

active, most are regulated by a wide array of signals, including phosphorylation, ligand binding, as well as homo- or hetero-oligomerisation (Spitz & Furlong, 2012).

The RNAP catalytic site enables successive ribo-nucleotide triphosphates (rNTPs) to base pair with the complementary nucleotide on the exposed template strand. RNA bases differ in that uracil (U) replaces T. RNAP catalyses phosphodiester bond formation by facilitating a nucleophilic attack from the 3' OH of the nascent RNA strand to the α -phosphate of the rNTP to elongate the RNA chain (Brueckner et al., 2009).

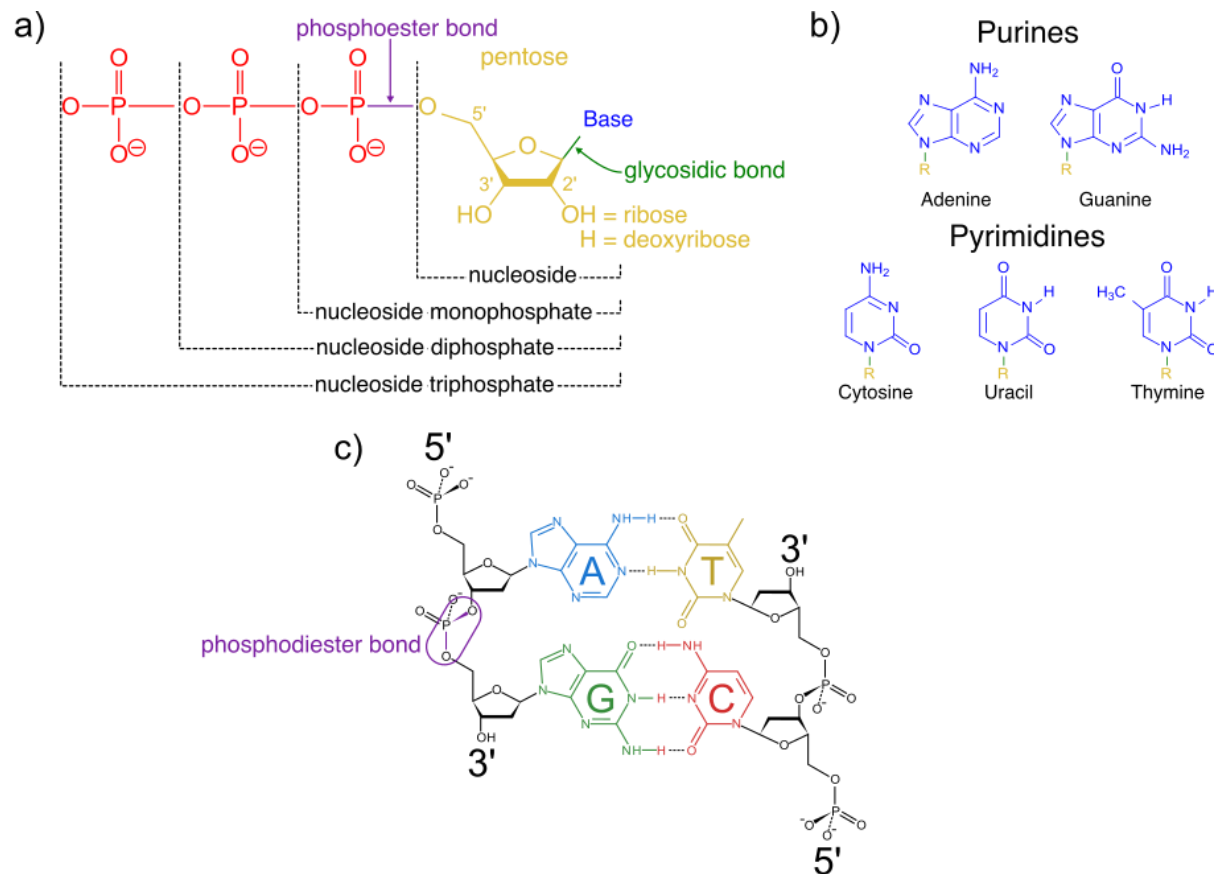


Figure 1.1.2: a) Chemical structure of nucleoside triphosphates. b) Types of nitrogenous base. c)

Chemical structure of DNA.

In eukaryotes, transcription is carried out by 3 different RNA polymerases (RNAPI-I, II, III), with plants having additional RNAPs that are closely related to RNAP II

(RNAPIV-V), as well as chloroplast and mitochondrial RNAPs (Borner et al., 2015; Ringel et al., 2011). RNAPI synthesises ribosomal RNA, RNAPII synthesises messenger RNA (mRNA) and small regulatory RNAs (srRNA), and RNAPIII transcribes transfer RNAs (tRNA), 5S rRNAs and other small RNAs. RNAPIV and V transcribes siRNAs (Cramer et al., 2008; Haag & Pikaard, 2011). Chloroplast RNAP resembles cyanobacterial multisubunit RNAP, whereas mitochondrial polymerase resembles single-subunit T7 viral RNAP (Baum & Baum, 2020). Bacteria and archaea have a single type of multisubunit RNA polymerase for all RNAs (Werner & Grohmann, 2011).

Following prokaryotic transcription, mRNA is engaged by ribosomes to catalyse the formation of a polypeptide. In eukaryotes, an additional splicing step is necessary after transcription to remove introns to produce the mature mRNA. The mRNA sequence is the template for the corresponding protein sequence. The ribosome recruits tRNAs, each with covalently attached to an amino acid by amino acyl tRNA synthetases via an ester bond between the 3' end of the RNA which corresponds to 20 different naturally occurring amino acids or a start or stop codon. For every non-overlapping triplet of RNA bases, also known as codon, forms base pairs with complementary anticodons on a tRNA. Each specific sequence in the codon can only encode a single amino acid, however, each amino acid can be encoded by multiple different three letter codon sequences. The ribosome catalyses the hydrolysis of the amino acid from the tRNA and formation of peptide bonds between adjacent amino acids to form the polypeptide. Polypeptides are then folded, assembled and often post-translationally modified to produce a functional

protein that performs functions within the cell, which gives rise to observable traits, also known as phenotype (Alberts et al., 2022) (Figure 1.1.1).

1.2 The stages of transcription

Transcription can be broken down into 3 major phases: initiation, elongation and termination (Mohamed, Vazquez Nunez & Vos, 2022).

RNAP is a processive complex that binds non-specifically to DNA. In order to transcribe from the beginning of the gene, during the initiation phase, RNAP is recruited to promoter regions located upstream (US) of the transcription start site (TSS, +1) by TFs known as σ factors in a process known as promoter search. There are multiple types of σ factor (discussed in more detail in section 1.6) that compete with each other to ensure specific gene expression occurs under the appropriate conditions. During promoter search, the RNAP is thought undergo 3D diffusion between DNA loci as well as collisions, which leads to eventual recruitment to the promoter region (Wang et al., 2013). The holoenzyme forms a transcription incompetent closed complex (RP_c) upon binding to the promoter, which during productive transcription undergoes open complex formation to form a transcription competent open complex. The closed complex is defined by the DNA being in a fully double-stranded state. During open complex formation, the DNA must be melted, and the template strand is loaded into the catalytic site of RNAP to form the 12-15 nucleotide transcription bubble (Glyde et al., 2018; Nudler, 2009; Zhang et al., 2012).

RNAP then undergoes rounds of abortive initiation. Abortive initiation plays a key role in determining expression levels and regulating NTP concentrations within the cell (Hsu, 2009). Abortive initiation occurs as a result of increased tension within the RNAP complex to accommodate the transcription bubble, resulting in an energetic barrier to elongation. RNAP transcribes 2-11 nt RNA (although some studies have shown up to 17 nt in rare cases), followed by DNA scrunching, pausing, backtracking and release of RNA (Hsu, Vo & Chamberlin, 1995; Imashimizu, Tanaka & Shimamoto, 2011). DNA scrunching is the process in which the holoenzyme remains bound to the promoter region, whilst pulling the downstream DNA into the active site, resulting in DNA unwinding.

This is followed by promoter clearance, where the TFs involved in promoter recruitment are displaced and RNA polymerase enters a processive elongation state, where the transcription bubble shrinks to 10 nt (Kang et al., 2017).

Elongation is the stage where the RNA is synthesised by the processive, active form of RNAP. During elongation, RNAP polymerises RNA at a rate of 20-100 nucleotides per second and translocates stepwise along the bases of the template strand in a 3'→5' direction, with the RNA being synthesised in a 5'→3' direction beginning from the 3' end of the template strand. RNAP supports a 7-9 nt DNA-RNA hybrid duplex within the transcription bubble, with the 5' of the single stranded RNA emerging from the RNA exit channel (Fuchs et al., 2014). The fidelity and processivity of elongation is often regulated further by additional TFs that alter the frequency and length of pausing and backtracking. RNAP is able to accurately and consecutively synthesise thousands of nucleotides (Svetlov & Nudler, 2009).

Termination is essential in ensuring the correct length of the RNA transcript and maintaining correct levels of RNA throughout the cell. Transcription termination in bacteria can be categorized into intrinsic termination, via the formation of a GC-rich RNA hairpin structure followed by a U-rich sequence on the RNA, or via a factor dependent mechanism using Rho helicase, which actively unwinds the DNA-RNA hybrid structure leading to RNA release from the RNAP catalytic site. After termination, bacterial RNAP undergoes a process of sliding, and sometimes dissociation, although often, transcription is re-initiated in the opposite direction, in a process known as secondary initiation (Harden et al., 2020).

In eukaryotes, the mechanism of termination varies across the RNAPI-III. RNAPIII is terminated by a factor-independent process which involves recognition of a 4-7 nt polyT sequence on the non-template DNA, resulting in a weaker A-U DNA-RNA base pairing, leading to dissociation (Girbig et al., 2022). RNAPI is also terminated upon encountering a polyT sequence on the non-template strand. The polyT sequence is followed by a specific sequence that is recognised by Nsi1 in *S. cerevisiae* (TTF-I in mice) that acts as a roadblock and causes RNAPI to stall. The pre-mRNA is then cleaved up to the Nsi1 binding site by endonuclease Rnt1. In *S. cerevisiae*, RNA exonuclease Rat1 or the Sen1 helicase is required to actively separate the remaining RNA-DNA hybrid and release RNAPI. In mice, this function involves binding of PTRF, in a passive process that is not fully understood, but is thought to involve RNAPI subunit Rpa12 (Xie, Libri & Porrua, 2023).

RNAPII transcribes the majority of pre-mRNA in the cell, and its termination is coordinated with 3'-polyadenylation and splicing. The mechanism of termination

involves recruitment of cleavage and polyadenylation specificity factor (CPSF) as well as cleavage factors I and II (CFI-II) to a consensus poly-A signal (AAUAAA) with the help the Ser2 phosphorylated Rpb1-CTD. Following recognition of the poly-A signal, the CPSF utilises its endonuclease activity to cleave downstream of the poly-A signal, and poly-A polymerase adds a polyA tail to the 3' end to enable nuclear export and translation. The Rat1 exonuclease then degrades the 5' end of the RNA until it collides RNAPII, causing dissociation (Park, Kang & Kim, 2015).

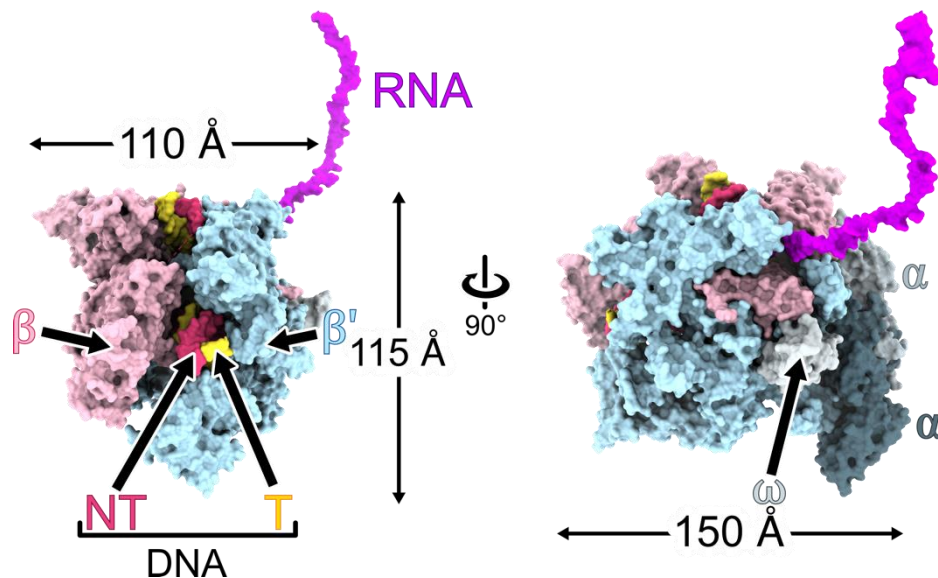
1.3 Bacterial RNA polymerase

DNA-dependent RNA polymerase was first discovered in nuclear fractions of rat liver in 1959 by Weiss and Gladstone (Weiss & Gladstone, 1959), before being found in bacteria and plant organisms shortly thereafter (Huang, Maheshwari & Bonner, 1960; Hurwitz, Bresler & Diring, 1960; Stevens, 1960). RNA polymerase was first purified in 1962 from *E.coli* by Anders and coworkers (Furth, Hurwitz & Anders, 1962), and the first structural studies used 2D electron crystallography to determine the first low resolution structures in 1988, before the high-resolution X-ray crystallographic structures were published in 1999 by Darst and coworkers (Darst, Kubalek & Kornberg, 1989; Darst et al., 1988; Zhang et al., 1999).

E. coli RNAP is a 386 kDa complex made of 5 subunits: $\alpha_2\beta\beta'\omega$ (Table 1.3.1). The catalytic site is located in the β and β' subunits, which form the RNAP claw. RNAP complex formation begins with formation of the α_2 homodimer, which binds to the β subunit. This is followed by β' subunit recruitment to form the characteristic claw, and finally the ω subunit binds the β' subunit (Figure 1.3.1) (Wang et al., 1997; Zhang et al., 1999).

Table 1.3.1: Length and molecular weight of *E. coli* RNA polymerase subunits

Subunit	Length (No. residues)	Molecular Weight (kDa)
α	329	37
β	1342	150
β'	1407	156
ω	91	11

Figure 1.3.1: Overview of the RNAP. Left – Downstream face of RNAP. Right – β' face of RNAP.

Structure adapted from PDB ID: 6XDR

1.3.1 α subunit

The α subunits play a key role in RNAP core assembly, serving as a platform for binding and assembling the $\beta\beta'$ claw, with each α subunit binding predominantly to the β or β' subunits. Each α subunit contains a N and C-terminal domain (α NTD and α CTD). The α CTD binds to US regions of the DNA, most notably with AT-rich “UP” elements located US of the -35 region, that is found in gene promoters associated with genes encoding ribosomal RNA and ribosomal proteins. The α CTD is essential for binding to activating region 1 on class I and class II catabolite activator proteins (CAPs), that play critical roles in detecting high cAMP levels under low glucose

conditions (Hao et al., 2022; Liu et al., 2017; Murayama et al., 2015). The α NTD is connected to the α CTD by a long flexible linker and contains conserved hydrophobic α helices that form the α_2 homodimerization interface (Wang et al., 1997). The α NTD also binds activating region 2 on class II CAPs in *E. coli* (Niu et al., 1996; Shi et al., 2020a; Sutherland & Murakami, 2018).

1.3.2 ω subunit

The ω subunit is the smallest subunit of RNAP and is not essential for *in vitro* transcription activity, although it is proposed to be involved in regulating the stability and folding of the β' subunit. ω is also known to play a role in regulating σ factor recruitment, DNA supercoiling, promoter selection and stringent response (Kurkela et al., 2021). More recent studies have suggested a role in mediating H-NS regulated prophage gene silencing, although the mechanism is unclear (Yamamoto et al., 2018).

1.3.3 β and β' subunits

The β and β' subunits form the claw of the RNAP that encloses the catalytic site at the base. The $\beta\beta'$ claw contains four structural components that enable transcription; the catalytic site, transcription bubble maintenance elements, claw opening elements, and DNA, RNA and secondary channels (Lane & Darst, 2010).

The RNAP catalytic site consists of two double- ϕ - β -barrel (DPBB) domains that utilises a network of interactions to select and position complementary rNTPs. The DPBB from the β' subunit contains the aspartate triad -NADFDGD- motif that coordinates the primary Mg^{2+} cation. The second DPBB domain comes from the β subunit that forms salt-bridge interactions with the γ -phosphate on the rNTP and

interacts with the secondary Mg^{2+} cation that is bound to the β - and γ - phosphates of the incoming rNTP. The β' subunit contains the trigger loop and bridge helix, which play essential roles in rNTP selection and catalysis (Figure 1.3.2) (see section 1.5.1).

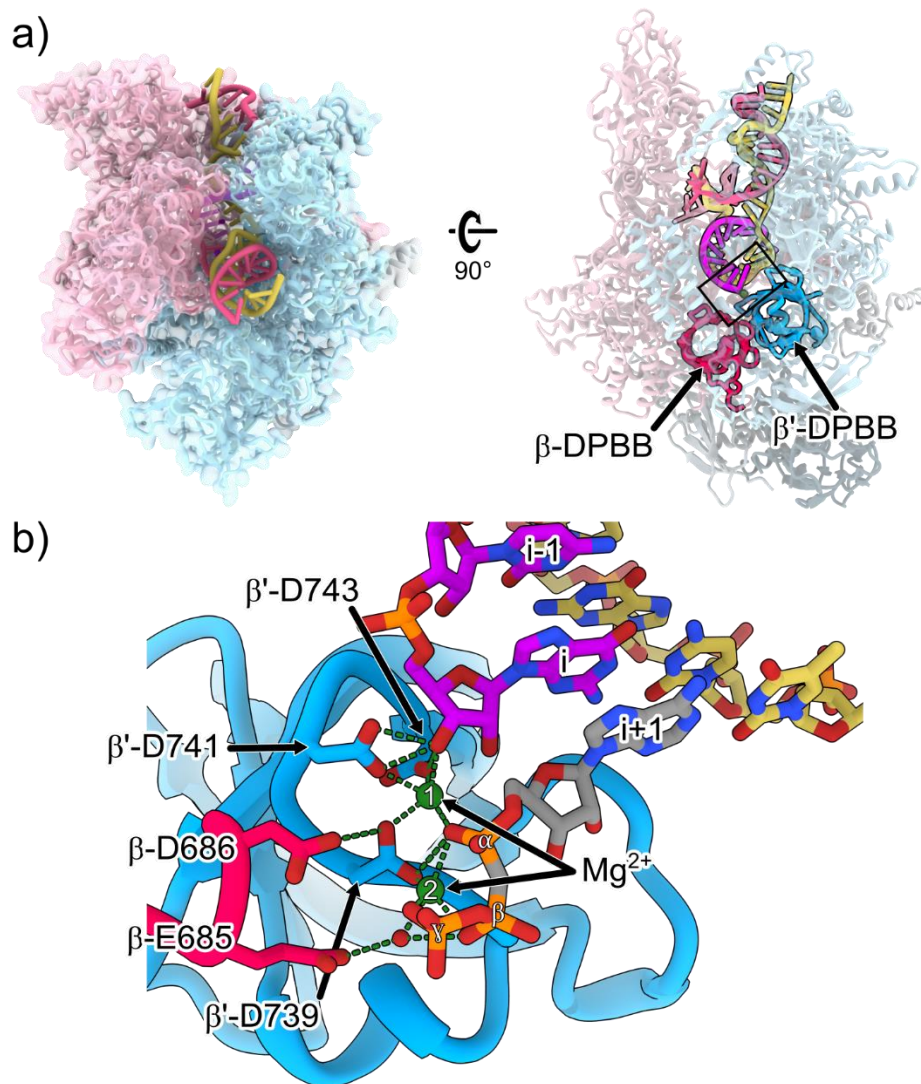


Figure 1.3.2: a) Location of the DPBB domains in the RNAP claw in *E.coli*/RNAP (PDB ID: 6ALH). Black square indicates the location of the zoomed in region shown in b). b) Zoom in of the *T.thermophilus* elongation complex (PDB ID: 2O5J) catalytic site highlighting the two Mg^{2+} cation coordinated by conserved aspartate triad and glutamate residues on the β' - and β -DPBB domains.

The 12-15nt transcription bubble is stabilised at the US end by the β' rudder and in the downstream (DS) end by β fork loop 2, which is supported by β fork loop 1. The 8-9nt DNA-RNA hybrid is enclosed by the bridge helix on the DS (3' RNA) end. On the US end, the β' lid forms a wedge that prevents extended DNA-RNA hybrid formation and is supported by the β' zipper, β flap and β switch 3 loops that form

part of the RNA exit channel (Naryshkina, Kuznedelov & Severinov, 2006; Vassylyev et al., 2007). The β' zipper separates the US dsDNA exit channel and RNA exit channel, forming part of the DNA exit channel with the β' coiled-coil, and β US lobe (Figure 1.3.3). The β flap also plays a key role in intrinsic termination, as the RNA hairpin is known to press against the β flap, resulting in internal stress that leads to RNA dissociation (Ray-Soni, Bellecourt & Landick, 2016).

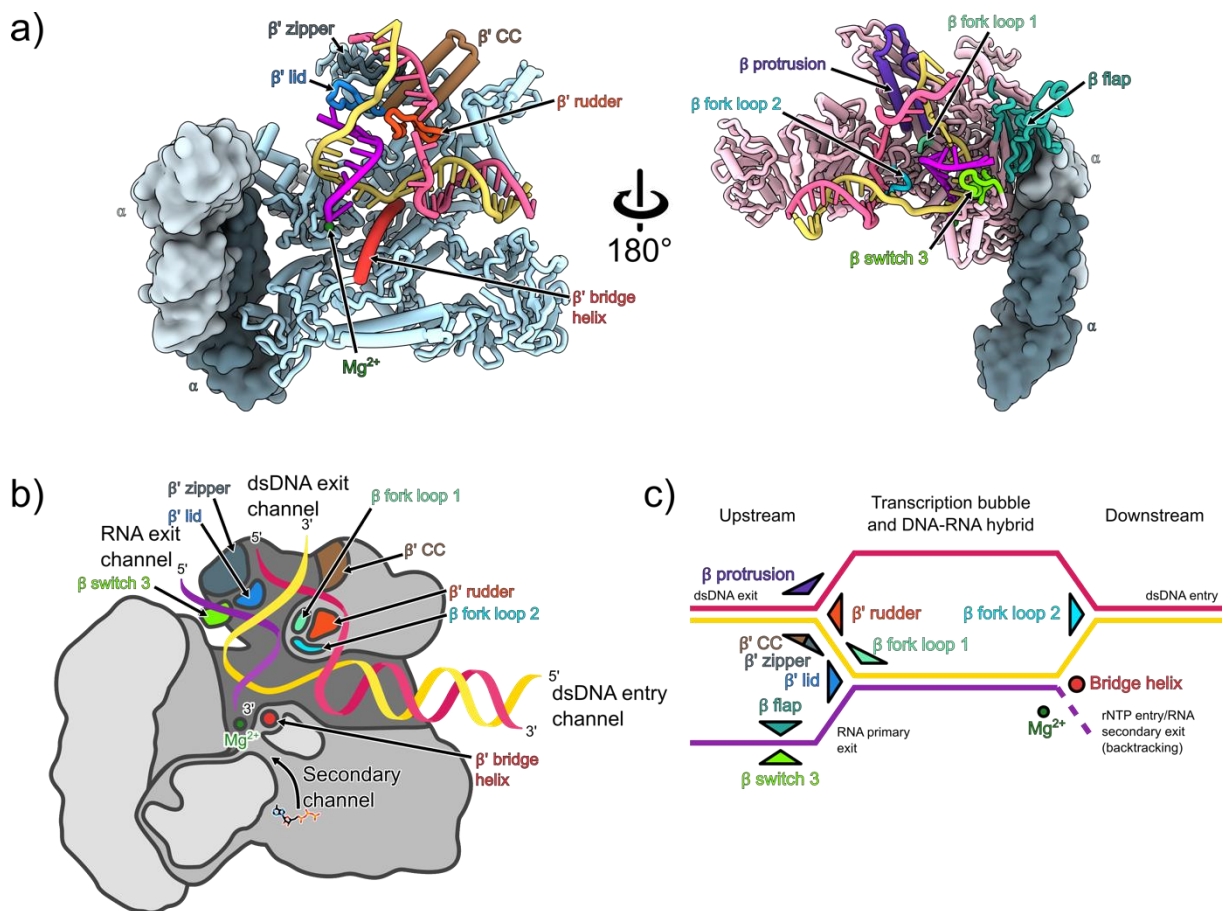


Figure 1.3.3: a) Structure of RNAP elongation complex showing transcription bubble and DNA-RNA

hybrid interactions with the RNAP core showing the β' (left) and β (right) interactions

(Adapted from PDB ID 6XDR). b) Simplified cartoon schematic of the catalytic cleft, showing

the channels within RNAP as well as the arrangement of the transcription bubble and DNA-

RNA hybrid. c) Linear schematic of a).

In addition, the base of the RNAP clamp is lined with 5 “switch” regions that transition from a disordered to folded state upon dsDNA binding and act as hinges to open and close the clamp as RNAP converts to the processive closed state (Figure 1.3.4a), and are also the target for multiple antibiotics, including myxopyronin and coralopyronin (Boyaci et al., 2019; Mukhopadhyay et al., 2008). The closing of the clamp prevents DNA escaping from the RNAP clamp during elongation (Cramer, 2002; Lane & Darst, 2010). Clamp opening and closing is also facilitated by the β lobe, β protrusion and β' clamp, which are located at the tip of the RNAP pincers (Figure 1.3.4b) (Chakraborty et al., 2012). The opening and closing of the pincers occurs during binding to σ factors, promoter DNA and during termination. Pincer opening and closing controls the accessibility of the catalytic site (Bellecourt et al., 2019; Geszvain & Landick, 2004).

The β' clamp, β' jaw (which includes the sequence insert 3 in *E. coli*), and ω subunit, along with the upstream side of β' form part of the swivel module, that plays important roles in backtracking and enlarging the secondary channel (discussed further in section 1.5.2.1) (Abdelkareem et al., 2019).

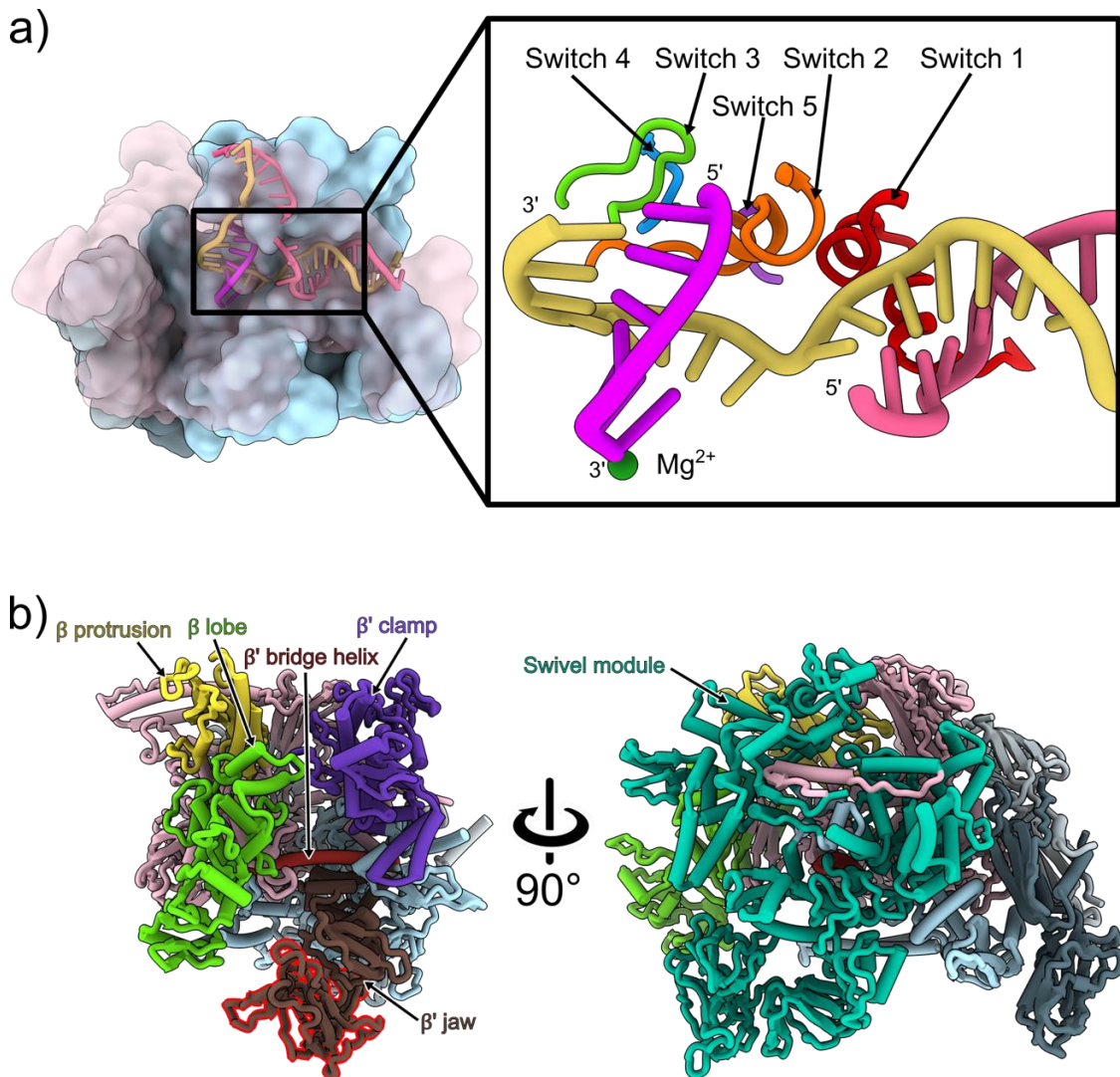


Figure 1.3.4: a) Switch regions in elongation complex. b) Mobile domains in RNAP. *E. coli* specific β' sequence insertion 3 is outlined in red. c) All structures are adapted from PDB 6XDR.

1.4 Comparison with Archaeal and Eukaryotic RNA polymerases

Both archaeal and eukaryotic polymerases share functional and structural homology with bacterial RNAPs, however, contain an increased number of subunits due to more diverse promoter elements, recruitment factors and pre-mRNA processing.

All eukaryotic polymerases contain a claw like complex made up of 2 subunits that are analogous to the β and β' subunits, whereas archaeal RNAP splits RPOA, which

is analogous to the β' subunit, into 2 subunits, termed RPOA and RPOA'' (Table 1.4.1, Figure 1.4.1). The eukaryotic polymerases I-III contain 7 to 12 additional subunits, whereas archaeal and plant-specific polymerases contain an additional 7 (Figure 1.4.1). Eukaryotic and archaeal RNAPs all contain a unique stalk domain made up of Rpb4 and Rpb7 in RNAPII, that binds to numerous general TFs and has been suggested to be important regulating clamp opening and processivity (Jun et al., 2014). Rpb10 and Rpb12 in RNAPII and their homologues form a complex with Rpb3 and are essential for the assembly of the Rpb1-Rpb2 claw (Turowski & Boguta, 2021).

Rpb5 modulates binding to elongation factors such as Spt5 to regulate the transition from initiation to elongation, downstream DNA and regulates backtracking (Martinez-Fernandez et al., 2018). Rpb8 was shown to be essential in a variety of deletion and mutagenesis studies and has been suggested to play a role in assembly of RNAPIII, but its exact function remains unknown (Abel et al., 2018; Voutsina et al., 1999). Rpb4 forms a heterodimer with Rpb7 and binds to single-stranded DNA and RNA to stabilise the DNA-RNA hybrid (Meka et al., 2005). Rpb9 binds to Rpb1 and stabilises the bridge helix to regulate processivity (Zaros et al., 2007).

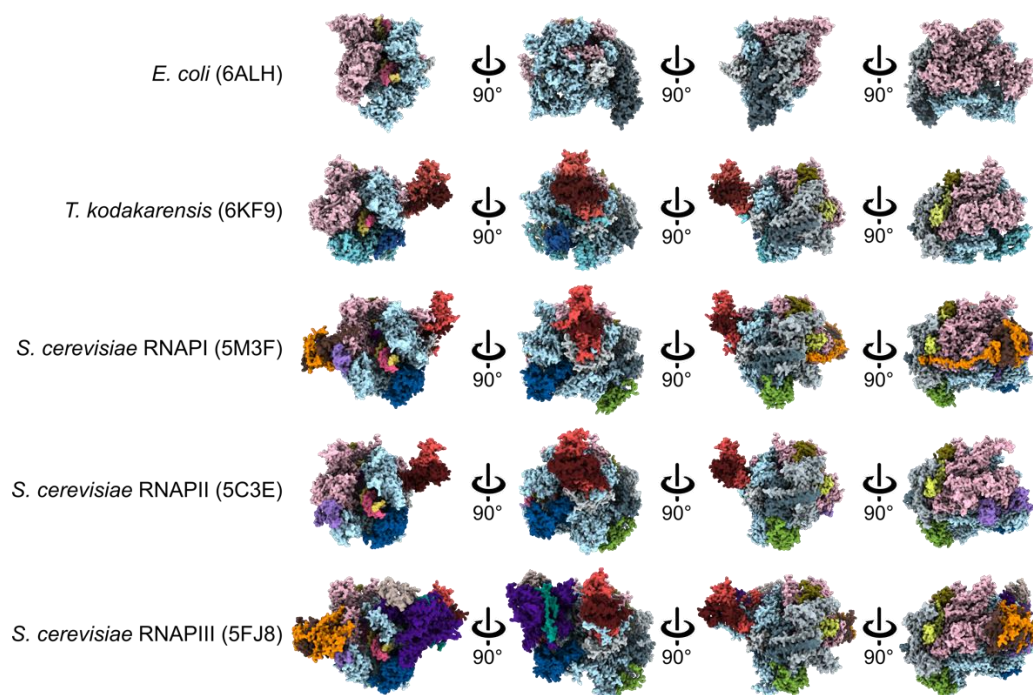


Figure 1.4.1: Comparison of bacterial, archaeal, and eukaryotic RNAPs (*S. cerevisiae* used as an example). Key is shown in Table 1.4.1.

Table 1.4.1: RNAP subunit conservation between eukaryotic and prokaryotic organisms. Colours of cells correspond to colours in Figure 1.4.1.

Bacterial RNAP	RNAPI		RNAPII		RNAPIII		RNAPIV (Plant only)	RNAPV (Plant only)	Archaea	
	Yeast	Human	Yeast	Human	Yeast	Human				
β'	Rpa190	RPA1	Rpb1	RPB1	Rpc160	RPC1	NRPD1	NRPE1	RPOA	RPOA''
β	Rpa135	RPA2	Rpb2	RPB2	Rpc128	RPC2	NRPD2	NRPE2	RPOB	
$\alpha 1$	Rpc40	RPAC1	Rpb3	RPB3	Rpc40	RPAC1	NRPD3	NRPE3a/3b	RPOD	
$\alpha 2$	Rpc19	RPAC2	Rpb11	RPB11a	Rpc19	RPAC2	NRPD11	NRPE11	RPOL	
Nonconserved	Rpb10	RPABC5	Rpb10	RPABC5	Rpb10	RPABC5	NRPD10	NRPE10	RPON	
Nonconserved	Rpb12	RPABC4	Rpb12	RPABC4	Rpb12	RPABC4	NRPD12	NRPE12	RPOP	
ω	Rpb6	RPABC2	Rpb6	RPABC2	Rpb6	RPABC2	NRPD6	NRPE6a/b	RPOK	
Nonconserved	Rpb5	RPABC1	Rpb5	RPABC1	Rpb5	RPABC1	NRPD5	NRPE5	RPOH	
Nonconserved	Rpb8	RPABC3	Rpb8	RPABC3	Rpb8	RPABC3	NRPD8a/8b	NRPE8a/8b	RPOG*	
Nonconserved	Rpa14	RPA14	Rpb4	RPB4	Rpa17	RPC9	NRPD4	NRPE4	RPOF	
Nonconserved	Rpa43	RPA43	Rpb7	RPB7	Rpc25	RPC8	NRPD7	NRPE7	RPOE	
Nonconserved	Rpa12	RPA12	Rpb9	RPB9	Rpc11	RPC10	NRPD9b	NRPE9a/b	RPOX/TFS	
Nonconserved	Rpa49	RPA49			Rpc53	RPC4				
Nonconserved	Rpa34	RPA34			Rpc37	RPC5				
Nonconserved					Rpc82	RPC3				
Nonconserved					Rpc34	RPC6				
Nonconserved					Rpc31	RPC7				

1.5 Conserved functions of RNA polymerase

1.5.1 Nucleotide addition

Nucleotide addition occurs in three steps: nucleotide selection, phosphodiester bond formation and translocation. rNTPs enter the catalytic site via the secondary channel (Figure 1.3.3b). The secondary channel is a negatively charged funnel that narrows to 12 Å at the entrance to the catalytic site, limiting access to a single rNTP diffusing in or out at any one time during the “pre-insertion” state. In the pre-insertion state, the trigger loop is in an “open” disordered state (Figure 1.5.1a) (Wang, Sexton & Feig, 2017).

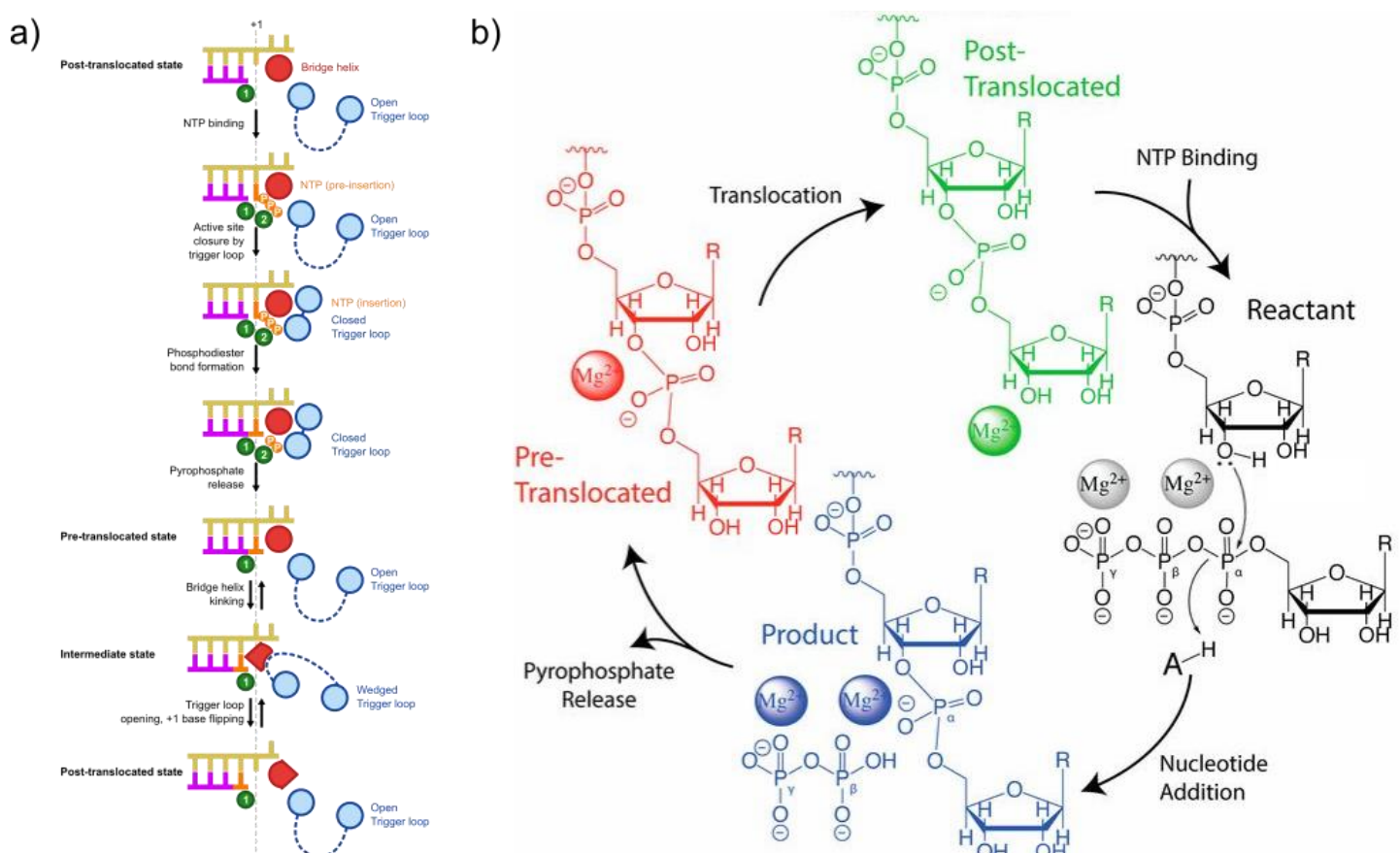


Figure 1.5.1: a) Schematic diagram of the nucleotide addition cycle. Adapted from Nudler (2009). b)

Mechanism of nucleophilic attack from incoming rNTP during nucleotide addition. Adapted from

Roston, Demapan & Cui (2019).

If the incoming nucleotide is complementary and the 2' OH group is detected, the trigger loop transitions to the “insertion” state; the trigger loop folds from the “open” state into a “closed” α -helical hairpin upon formation of a Watson-Crick base pair with the template DNA nucleotide at the $i+1$ site (Sydow & Cramer, 2009). The folded state of the trigger loop forms a helical bundle with the bridge helix, occluding the rNTP from exiting the catalytic site and orients the rNTP for nucleophilic attack (Figure 1.5.1b). The insertion state is stabilised by the aforementioned interactions between the aspartate triad residues with the primary and secondary Mg^{2+} cations. The primary Mg^{2+} cation positions the 3' OH group on the nascent RNA chain with the α -phosphate from incoming rNTP, whilst the secondary Mg^{2+} cation neutralises the negative charge from the phosphate groups (Cheung & Cramer, 2012; Steitz, 1998).

Formation of the phosphodiester bond results in the release of pyrophosphate, which exits via the secondary channel. The formation of hydrogen bonds between complementary bases and the solvation of the pyrophosphate is proposed to provide the chemical energy to facilitate the unfolding of the trigger loop and the mechanical movement of the bridge helix (Tan et al., 2008).

The bridge helix kinks to translocate the 3' end of the RNA hybrid 1 bp upstream, allowing the next template DNA base to enter the active site. The bridge helix is supported by the trigger loop, which forms a wedge between the bent region of the bridge helix and the wall of the RNAP catalytic site (Brueckner & Cramer, 2008) (Figure 1.5.1).

1.5.2 Pausing

Despite the processive nature of RNAP, ≥ 1 second pauses occur every 100-200 bases in both prokaryotes and eukaryotes. Pauses can be classified as short lived elemental pauses or longer duration backtracked, RNA-hairpin or regulator-mediated pauses. Pauses can be regulated *in trans* by a variety of factors, including NusA, NusG/RfaH (Spt4/Spt5 in eukaryotes) interactions (Saba et al., 2019).

Sequence-specific pause sites occur in *E.coli* at a RNA consensus sequence of $G_{-11}G_{-10}XXXXXXY_{-3}G_{-2}Y_{-1}G_{+1}$, where X represents any base, Y represents pyrimidine, the -1 position represents the RNA 3' end and +1 represents the incoming rNTP (Larson et al., 2014). In eukaryotes, the sequence is $G_{-10}XXXXXXY_{-2}G_{-1}Y_{+1}$ (Gajos et al., 2021). Pause sites in prokaryotes are enriched just before start codons (Larson et al., 2014) and eukaryotic pauses are enriched at the upstream edge (in promoter proximal pausing) and downstream edge of exons, as well as at roadblocks such as nucleosomes and CCCTC-binding factor (CTCF). This suggests that pausing helps coordinate transcription-translation in bacteria and mRNA splicing and nucleosome retention in eukaryotes (Noe Gonzalez, Blears & Svejstrup, 2021; Saba et al., 2019).

In this section, we will focus on factors conserved throughout prokaryotes, archaea and eukaryotes: intrinsic pauses, NusG/RfaH (Spt4/Spt5) and GreA/B (TFIIS). For an in-depth view of prokaryote- and eukaryote-specific *in trans* pausing and co-transcriptional processing, we refer to the following references (Blaha & Wade, 2022; Ehara et al., 2019; Ehara et al., 2022; Farnung et al., 2021; Filipovski et al., 2022; Kang et al., 2019; Landick, 2021; Noe Gonzalez, Blears & Svejstrup, 2021; Wang et al., 2022; Webster & Weixlbaumer, 2021; Zhang et al., 2021).

1.5.2.1 Elemental and hairpin-mediated pausing

Recent structural work by the Darst and Landick groups have determined the mechanisms of elemental and hairpin mediated pausing in prokaryotes using pause sequences with and without an RNA pause hairpin (Saba et al., 2019). The elemental pause is a non-productive state that occurs during elongation. As shown in Figure 1.5.2, translocation of the paused sequence results in a half-translocated step, where the RNA is translocated independently of the template DNA. This tilts the base stacking, such that the $i+1$ site does not have a free template strand base available, and a conserved arginine residue is proposed to hold the $+1$ template base in position. This was later corroborated by fluorescence translocation assays (Saba et al., 2019).

Additionally, the swivel module (Figure 1.3.4b) is rotated in such a way that it would prevent the proper folding of the trigger loop, in what is referred to as the “swivelled” state. The swivelled state is stabilised by the RNA pause hairpin, with a class being found in a half-translocated state but without the pause hairpin and the rotated swivel module, suggesting that swivelling occurs after elemental pauses as a result of RNA pause hairpin formation (Kang et al., 2018a). More recent unpublished studies suggest that there are a multitude of different elemental paused states with varying degrees of swivelling depending on the folded state of the trigger loop, insertion state of the F-loop, β' -rim coiled-coil, SI3 and the DNA sequence (Kang et al., 2022).

In bacteria, NusA NTD and S1 subdomains are involved in binding to the RNA pause hairpins as well as binding the RNAP β flap, α CTD and ω subunit (via the CTD) to promote intrinsic termination and pausing. NusA also stabilises the flap tip helix and widens the RNA exit channel that is suggested to make room for further RNA folding and stabilise the swivelled state (Guo et al., 2018).

It remains unclear whether this half-translocated step occurs exclusively during pausing or is also present in between steps of processive elongation (Landick, 2021). Similar half-translocated DNA-RNA hybrids were also observed in eukaryote arrested RNAPII (Cheung & Cramer, 2011) and paused RNAPII complexes bound to DSIF(Spt4-5)-NEF (Vos et al., 2018b).

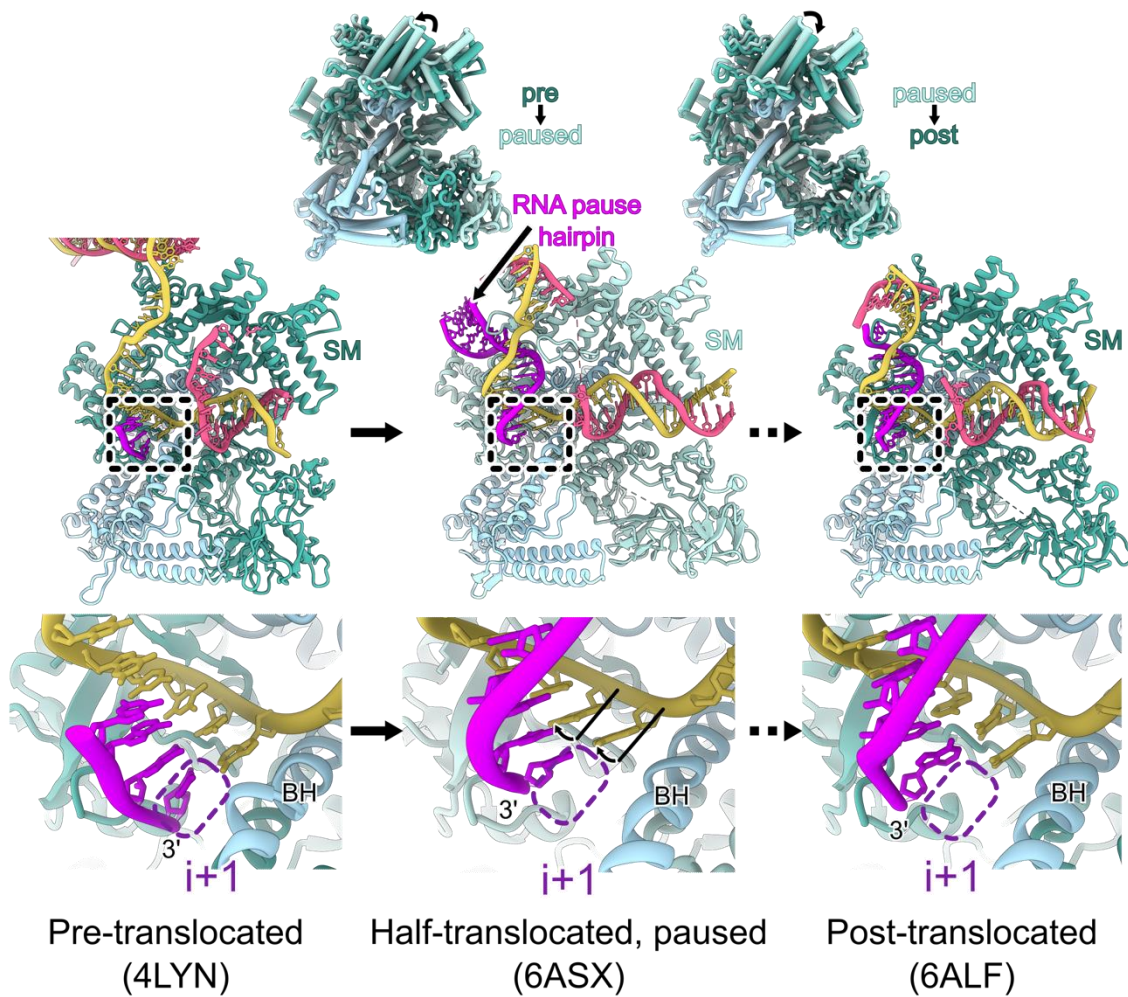


Figure 1.5.2: A comparison of pre-translocated, paused and post-translocated intermediates. Top row shows rotation of the swivel module from pre→paused and paused→post translocated states. The middle row shows a cross section of the RNAP core, with the RNA pause hairpin highlighted. The area is highlighted by dotted lines are zoomed in on bottom row. The bottom row shows the base stacking of the DNA-RNA hybrid at the i and $i+1$ (circled in purple) sites. The two black lines in the lower middle panel shows the rotation of the base stacking in the paused relative to the pre-translocated states, showing that the template base is not available in the $i+1$ site. BH=bridge helix, SM=swivel module.

1.5.2.2 Pause regulation by NusG/Spt5

NusG binds to RNAP via its conserved NusG N-terminal domain (NGN) and has been shown to bind to a wide variety of other complexes via its C-terminal KOW domain, including Rho, ribosomes and NusA. Eukaryotic and archaeal Spt5 also contains an NGN domain, which binds to conserved site on the tip of the β' (Rpb1) clamp of RNAP and bridges across to the tip of the β (Rpb2) lobe and protrusion (Kang et al., 2019).

Bacteria encode multiple NusG paralogs that can both stimulate and repress pausing depending on the interaction partner and organism. *E. coli* and most NusG homologs inhibit pausing and backtracking and are referred to as canonical NusG proteins. Canonical NusG is proposed to inhibit pausing via 3 mechanisms: firstly, by binding to upstream DNA and promoting annealing on the upstream edge of the transcription bubble. Secondly, NusG stabilises the non-template strand in a productive state, and lastly, NusG favours the catalytically active non-swivelled state. RfaH is a sequence specific NusG paralog that binds to the bases on non-template strand operon polarity suppressor, or *ops* loops on specific promoters (Kang et al., 2018b; Zhu et al., 2022). This is supported by biochemical evidence showing that NusG binds transiently *in vivo*, whereas RfaH binds with higher affinity and remains bound until termination (Figure 1.5.3a, b) (Wang & Artsimovitch, 2020).

In contrast to *E. coli*, mycobacterial and bacillial NusG promotes pausing. *B. subtilis* recognises TTNTTT motifs on the non-template DNA (where N=any base) (Landick, 2021; Yakhnin et al., 2020). Recent unpublished work by Landick and Murakami groups show that *M. tuberculosis* and *B. subtilis* NusG promote pausing

by flipping out and scrunching bases in the non-template strand to stabilise the non-productive swivelled state (Delbeau et al., 2022; Vishwakarma et al., 2022).

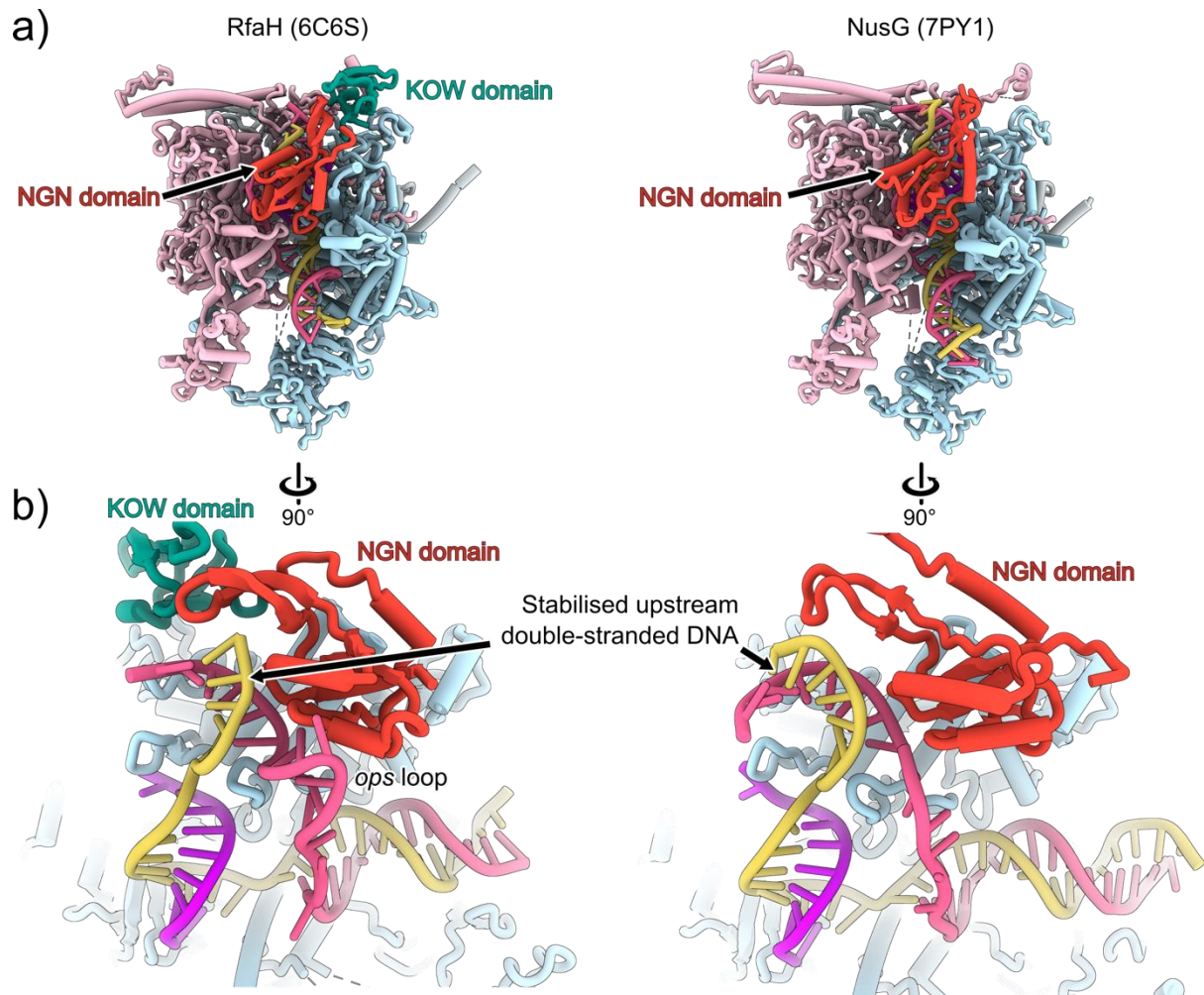


Figure 1.5.3: Comparison of *E.coli* RfaH and canonical NusG structures. a) overview of RfaH and NusG show that NGN domains bind conserved positions on RNAP. b) RfaH binds to a loop on the non-template strand *ops* loop, whereas canonical NusG stabilises the non-template strand backbone. Both stabilise the elongation state by stabilising the double-stranded DNA on the upstream edge of the transcription bubble.

Spt4 and Spt5 in eukaryotes forms a DSIF heterodimer, with Spt5 containing the NGN and multiple KOW motifs. The DSIF alone, DSIF-TFIIS-Elf1, DSIF-NELF and DSIF-PAF-SPT6-RTF1 complexes have been elucidated by the Cramer and Sekine

groups. Much like NusG, DSIF can both promote and inhibit pausing depending on the interaction partners. DSIF alone or with TFIIS forms a DNA-sequence and eukaryotic-specific RNA clamp that stabilises the closed state of RNAPII and prevents swivel module rotation and RNA hairpin formation, inhibiting pausing (Figure 1.5.4a, b) (Ehara et al., 2017). In contrast, DSIF-NELF complexes mediate promoter-proximal pausing. NELF inhibits TFIIS binding (discussed further below), trigger loop folding and restricts secondary channel entry (Figure 1.5.4c, d). NELF inhibition is alleviated by PAF and phosphorylation Spt5 by p-TEFb (Bernecky, Plitzko & Cramer, 2017; Vos et al., 2018a; Vos et al., 2020; Vos et al., 2018b).

1.5.2.3 Cleavage activity and Secondary channel factors

Despite the tightly controlled mechanism of nucleotide addition, errors occur at a rate of 1 error per 10^4 nucleotides, due to nucleotide misincorporation, transcription blockages, and DNA damage (Milo & Phillips, 2015).

The mechanism of mismatch repair varies depending on the type of mismatch, as well as the organism. Mismatched nucleotides result in distortions in the DNA-RNA hybrid and are usually excised by 3' RNA nuclease activity. Nuclease activity can be intrinsic or dependent on a variety of elongation factors that insert into the secondary channel and interact with the active site via 2 conserved acidic residues.

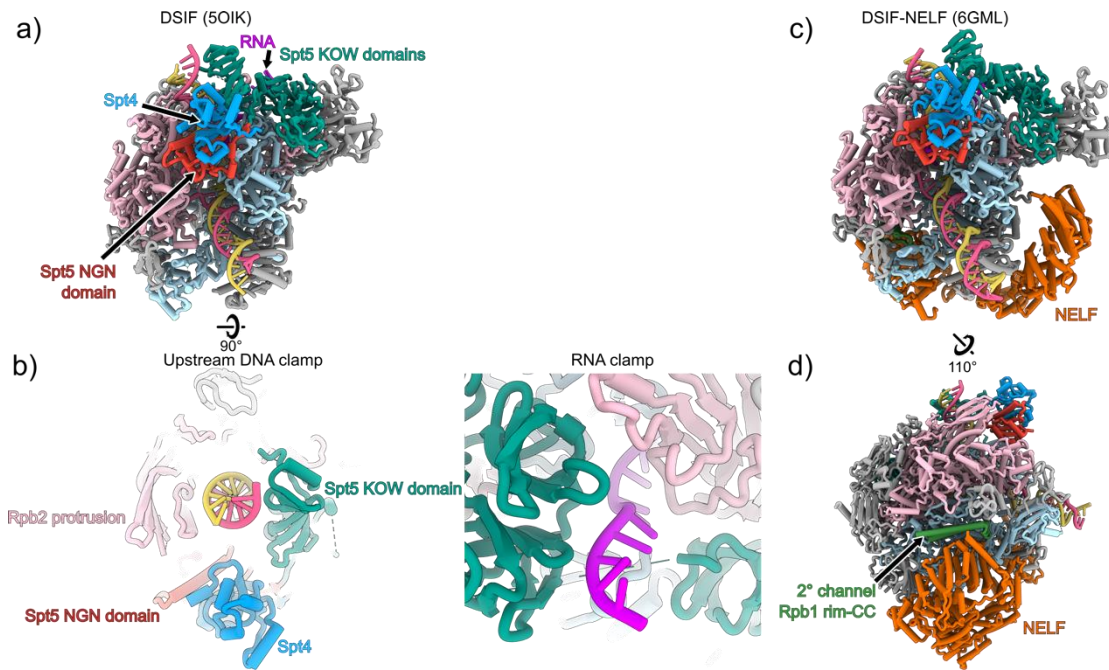


Figure 1.5.4: a) An overview of DSIF bound to RNAPII in eukaryotes. b) DSIF forms a DNA and RNA clamp that favours processive elongation. c) Upon recruitment of NELF, a pause-promoting complex is formed by NELF restricting the secondary channel and TFIIIS binding. Rpb1 and 2 are coloured as shown in Figure 1.4.1, remaining subunits coloured in grey.

In eukaryotes, cleavage is stimulated by TFIIIS, which binds the Rpb1 jaw and inserts its C-terminus into the catalytic site via the secondary channel (Kettenberger, Armache & Cramer, 2003). TFIIIS contains conserved acidic residues at its C terminus that can activate water molecules for nucleophilic substitution (Kettenberger, Armache & Cramer, 2004). However, RNA polymerase I-III does not require TFIIIS for cleavage *in vivo*, unless larger backtracking and arrest occurs (Cheung & Cramer, 2011; Sydow et al., 2009). Misincorporation leads to the deceleration of elongation, arrest and backtracking. Crystal structures of T-U mismatched Pol II showed that the T-U mismatch results in a wobble-base pair that shifts the 3'-OH to a suboptimal position for nucleophilic attack and causes dissociation of the primary Mg^{2+} cation.

Scaffolds with an additional DNA-RNA base pair proceeding the mismatch demonstrated that addition of 1 rNTP to the mismatched rUMP resulted in 3' RNA end fraying, with the base flipping away from the DNA and preventing the formation of the trigger helix, leading to arrest. The fork loop 2 also undergoes a conformational change to block translocation of downstream DNA into the active site. RNAP then backtracks to position the penultimate phosphodiester bond in the catalytic site for cleavage. RNA polymerase then catalyses hydrolysis of a dinucleotide from the 3' end of the RNA, catalysed by Mg^{2+} cations from the non-complementary rNTP itself or from TFIS, in a reverse of the addition reaction (Wang et al., 2009; Zenkin, Yuzenkova & Severinov, 2006).

Bacterial RNAP also has both intrinsic and factor-dependent cleavage, however, with the exception of cyanobacteria (which does not have Gre factor homologues), most cleavage events *in vivo* require specialised Gre elongation factors to cleave the misincorporated RNA in the backtracked complexes, regardless of the extent of backtracking (Abdelkareem et al., 2019; Toulme et al., 2000). This can in part be explained by the different conformation of the backtracked RNA base in bacteria compared to eukaryotic RNAP (Abdelkareem et al., 2019).

GreA cleaves 2-3 nucleotides from the 3' end, and GreB cleaves up to 9 nucleotides (Borukhov, Sagitov & Goldfarb, 1993). Gre factors are a family of secondary channel binding proteins, including DksA, TraR and Gfh1, that inhibit rRNA transcription initiation and transcript cleavage respectively (Lamour et al., 2006; Shin et al., 2021).

Gre factors contain a C-terminal domain that binds to a conserved β' -rim coiled-coil at the edge of the secondary channel, and is connected to its C-terminal coiled-coil domain by a flexible linker, which inserts into the catalytic site to facilitate RNA 3' hydrolysis and reactivation of the arrested RNAP (Figure 1.5.5) (Vassylyeva et al., 2007).

Recent unpublished studies have shown that GreA and GreB association requires different amounts of swivel module rotation. The angle or rotation of the swivel module is dictated by length of RNA extrusion from the secondary channel, and is suggested to therefore ensure recruitment of the correct Gre factor, in addition to contributing to the active site via their acidic tip residues. The interaction between the acidic tip and the active site causes a conformational change in the trigger loop that favours cleavage (Abdelkareem et al., 2019; Eduljee, 2021).

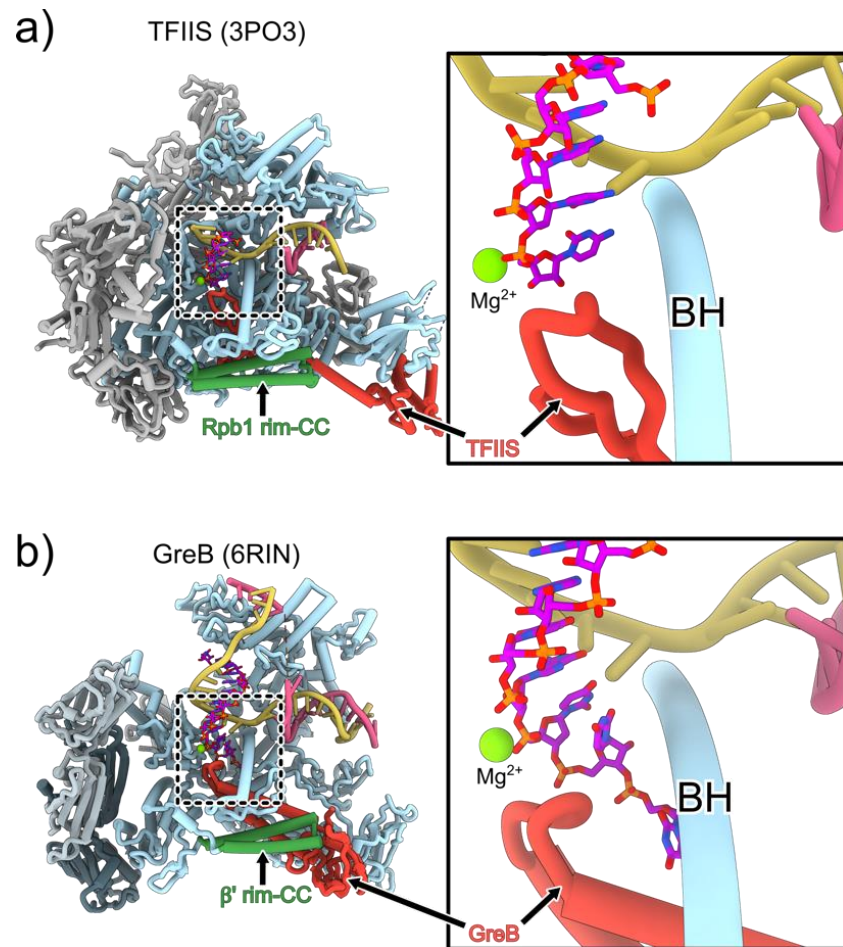


Figure 1.5.5: An overview of TFIIIS (a) and GreB (b) bound paused structures, showing insertion of acidic tip residues into the active site via the secondary channel to promote cleavage. β and Rpb2+9 are hidden for clarity

1.6 σ factors

σ factors were initially discovered by Burgess and coworkers as an essential co-factor for selective transcription *in vitro* (Burgess et al., 1969). In bacteria, σ factors are the transcription initiation factors that recruit RNAP to specific promoter regions and facilitate open complex formation. There are two families of σ factor: σ^{70} and σ^{54} . The majority of genes are transcribed by housekeeping σ^{70} factor, however, under stress conditions, alternative σ factors are activated and can be recruited to specific gene promoters in response to specific environmental stimuli.

1.6.1 σ^{70} family

The σ^{70} family can be divided into five distinct groups based on sequence conservation and function (Table 1.6.1, Figure 1.6.1) (Roberts et al., 2017).

Table 1.6.1: Overview of σ^{70} family members (Paget, 2015; Roberts et al., 2017).

Group	Function	Examples
1	Primary σ factor, for growth, cell division	σ^{70} ,
2	Dispensable for cell growth, stationary phase stress	σ^S/σ^{38} , σ^{32}
3	Sporulation, flagella biosynthesis, heat shock response	σ^{28}/FlhA , σ^D , σ^{WhiG} , <i>B. subtilis</i> σ^F , σ^E , σ^G , σ^K
4	Extracytoplasmic function (ECF) family, e.g., misfolded proteins in periplasm	<i>E. coli</i> σ^E/σ^{24} , <i>B. subtilis</i> σ^W , <i>E. coli</i> $\sigma^{\text{Fecl}}/\sigma^{19}$, <i>P. aeruginosa</i> σ^{PvdS} , <i>S. coelicolor</i> σ^R
5	Clostridium specific toxin and bacteriocin genes	TcdR, TetR, BotR, UviA

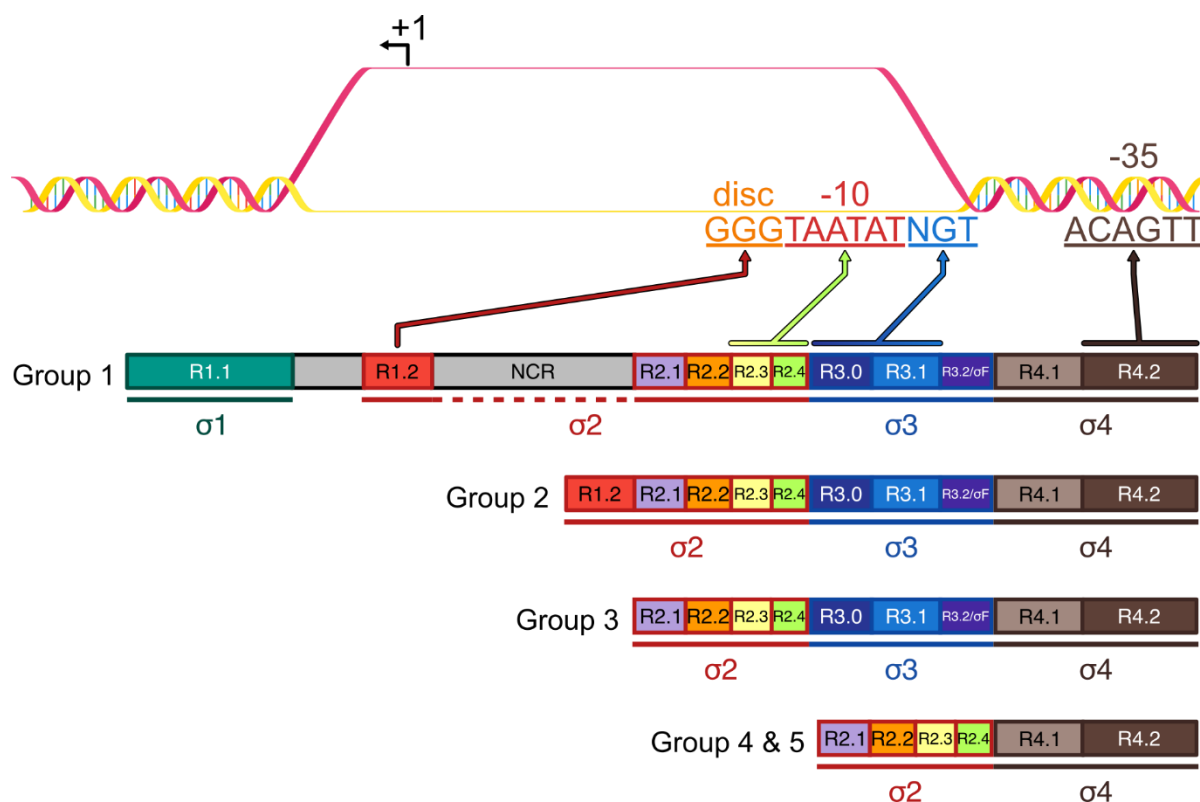


Figure 1.6.1: Overview of σ^{70} family members and binding sites. Disc=discriminator element, σF=σ finger.

Group 1 contains a single σ^{70} member, which has 4 structural subdomains connected by flexible linkers (Figure 1.6.2a). σ^{70} recruits RNAP to conserved -35 (TTGACA) and -10 (TATAAT) regions. In most cases (with the exception being *Mycobacteria* that requires RbpA and CarD), upon formation of a stable RNAP- σ^{70} -promoter DNA complex, σ^{70} is able to spontaneously isomerise to form an open complex.

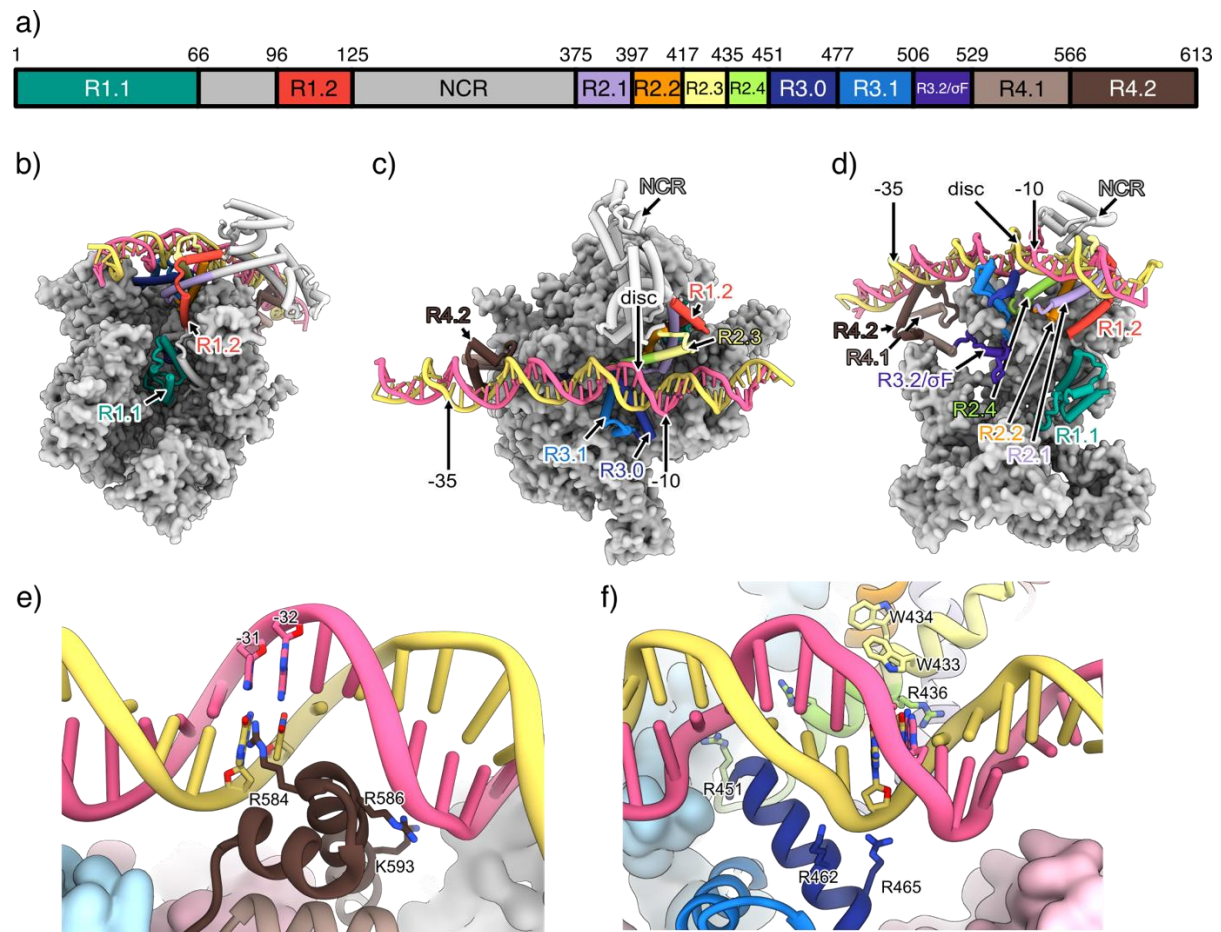


Figure 1.6.2: a) Domain map with residue range for subdomains of *E. coli* σ^{70} . b-d) Structure of *E. coli* σ^{70} closed complex (6PSQ). b) Downstream face of closed complex clamp. c) Top-down view looking into the catalytic site. d) cross section of claw. α and β subunits are hidden for clarity (Chen et al., 2020). e) R4.2 interaction with -35 consensus region and f) R2.3-3.0 interaction with the discriminator and -10 consensus region.

The structure of the σ^{70} closed complex is shown in Figure 1.6.2b-d. σ^{70} R1.1 binds to the downstream DNA channel in RNAP and plays a role in inhibiting spurious DNA binding in free σ^{70} . σ^{70} R1.2 interacts with the discriminator element (GGG) downstream of the -10 region. σ^{70} non-conserved region (NCR) projects away from the RNAP core and is implicated in promoter escape (Chen et al., 2020; Paget, 2015).

σ^{70} R2.1-2.2 binds to the β' pincer region of the RNAP core, whilst σ^{70} R2.3, 2.4 form an α -helix that binds to the consensus -10 region (TATAA) on the promoter DNA and is required for DNA melting in open complex formation. σ^{70} R3.0 binds to the major groove of the extended -10 region (TG) on promoter DNA. Critical for nucleation is the conserved tryptophan dyad residues in σ^{70} R2.3 that stabilises a flipped base at -11 on the promoter DNA, initiating transcription bubble formation (Figure 1.6.2f). σ^{70} R3.1 interacts with the β -flap and β' -rudder, and σ^{70} R3.2 (also known as the σ finger) interacts with the RNA exit channel, making them important in the process of σ^{70} displacement during promoter escape. σ^{70} R4.1 and 4.2 interacts with the β -flap tip helix, whilst σ^{70} 4.2 binds to major groove of the consensus -35 region of the promoter (TTGACA) (Figure 1.6.2e).

1.6.1.1 σ^{70} -mediated open complex formation

Recent work from the Darst group used inhibitory TF TraR to capture stepwise melted intermediates in σ^{70} -mediated open complex formation (Figure 1.6.3a). In the closed complex, σ^{70} R2.4 and σ^{70} R3.0 form a basic crevice that binds the phosphate backbone of the -10 region of the promoter DNA, whilst σ^{70} R4.1 and 4.2 form

sequence specific interactions with the -35 consensus region. The β protrusion binds to the +2 downstream DNA (Figure 1.6.2e, f).

The RNAP- σ^{70} holoenzyme contains up to five base specific pockets, with *E. coli* containing 4 pockets – two from σ^{70} , and two from the β subunit – that enable spontaneous open complex formation (Figure 1.6.3b, panels 1-3). In σ^{70} , there are tryptophan dyad residues that enable spontaneous nucleation of the transcription bubble at the -11 non-template strand position to form a DNA fork junction structure.

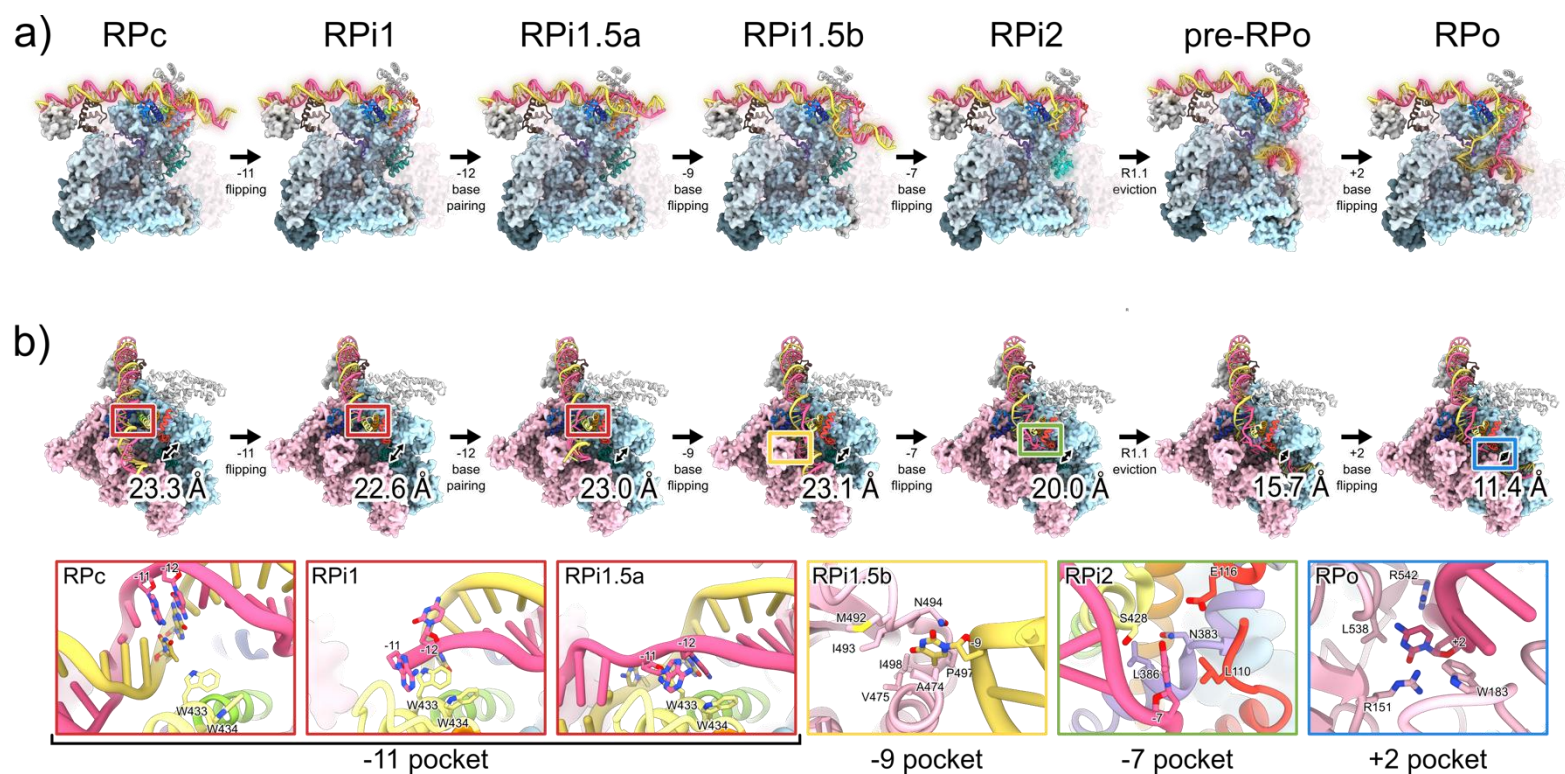


Figure 1.6.3: σ^{70} mediated open complex formation. a) an overview of the stepwise bubble opening in σ^{70} structures (PDB IDs: 6PSQ, 6PSR, 6PSS, 6PST, 6PSU, 6PSV, 6PSW). β subunit is transparent. b) Clamp opening widths (above) and base-specific pocket interactions during melting. Colour of boxes in the top panel represents the zoomed in positions in the corresponding box in the below panel.

The mismatch is further extended and stabilised by a conserved hydrophobic pocket in the β protrusion that stabilises a flipped-out base at the -9 position on the template strand (Figure 1.6.3b panel 4).

Upon further bubble opening, there is a pocket made up of hydrophobic and salt bridge interactions between σ^{70} R1.2, σ^{70} R2.1 and the -7 position on the non-template strand (Figure 1.6.3b panel 5). DNA opening is followed by ejection of σ^{70} R1.1 from the dsDNA binding cleft and transient clamp opening, enabling formation of the fully opened 13nt transcription bubble, with the +2 non-template strand base binding to a hydrophobic cleft in the β subunit, and β fork loop 2 positioned at the downstream fork of the transcription bubble (Figure 1.3.3, Figure 1.6.3b panel 6).

1.6.1.2 σ^{70} group 2 & 3 regulation

Both group 2-5 σ^{70} factors show reduced affinity for RNAP, and recruitment is therefore regulated by a variety of different factors that enhance retention of the σ factor to DNA or the RNAP core (Maeda, Fujita & Ishihama, 2000). Group 2 σ^{70} factors lack the 1.1 and NCR region and demonstrate lower binding affinity for RNAP compared to group 1 σ^{70} (Figure 1.6.1). The affinity of group 2 σ^S factor *in vivo* is bolstered by Crl, that tethers σ^{70} R1.2 to the β' clamp (Cartagena et al., 2019; Cavaliere et al., 2014; Liu, Zuo & Steitz, 2016). σ^S is further regulated by RssB, which binds to σ^{70} R3.0, and facilitates ClpXP mediated degradation during lag and exponential growth phases (Wang et al., 2021).

Group 3 lacks R1.2 in addition to 1.1 and NCR; consequently, the DNA is shifted away from the β' subunit and R4.2 helices; consequently, the affinity at the -35 region is reduced. σ^{28} is the best-studied of this group and is essential for flagella biosynthesis and sporulation. RNAP- σ^{28} -DNA complexes are stabilised by relocation of the zinc binding domain (ZBD) located at the N terminus of the β' subunit. The β' -ZBD binds to the -35 consensus promoter, improving the binding of R4.2 to enable stable open complex formation (Shi et al., 2020b).

1.6.1.3 σ^{70} group 4 (ECF)

ECF σ factors contain an altered linker region between $\sigma 2$ and $\sigma 4$ and has an altered $\sigma 4$ conformation to bind to the -35 consensus region. The lack of the extended -10 binding elements in $\sigma 3$ result in an altogether different mode of promoter binding and regulation – consequently, over 30% of ECF σ factors are regulated by anti- σ factors. ECF factors also lack the tryptophan diad at the -10 position and instead use a conserved N88 residue to wedge between the two DNA strands at the -12 region in comparison to -11 (Li et al., 2019).

ECF factors also contain pockets for reading 3 bases (-12,-11 and -7), in comparison to only two bases read by the group 1 σ^{70} factor (-11 and -7). This has been suggested to explain the lower efficiency of RPo formation with ECF σ factors (Li et al., 2019).

1.6.1.4 UP elements, α -CTD factors and class I and II activators of σ^{70}

In addition to -35 and -10 regions, σ^{70} contains “UP” elements between positions -36 and -60, which binds to the α -CTD. The UP elements contain A/T rich sequences and are important for transcription of rRNA and it contributes to a 30 fold increase in transcription activity. Recent cryo-EM structures have shown that both α -CTDs form a head-to-tail dimer that binds to the minor groove at the -54 to -48 region of the “UP” element in the RPc state, aiding in RNAP recruitment (Shin et al., 2021).

σ^{70} activators can be classified as class I and class II, depending on whether the activator can bind to sites upstream of the -35 region or overlap with the core promoter elements. One of the most well-studied example of a class I and II activator is catabolite activator protein, or CAP. CAP binds to over 100 different promoters and is essential for carbon metabolism under low glucose conditions. Upon binding cAMP, CAP is recruited to a specific 22 bp sequence; CAP consists of a homodimer that binds to a 22 bp sequence that creates a near $\sim 90^\circ$ bend in DNA.

A well-studied example of a class I promoter is the *lac* promoter. The CAP dimer binds to the α -CTD and the -61 region of the DNA via their helix-turn-helix motifs. Cryo-EM studies show that each CAP monomer (CAP I and CAP II) introduce a 33° and 55° kink respectively that is stabilised by the α -CTD binding to the minor groove via a conserved R265 residue. CAP acts as a stabilising partner with the α -CTD, improving RNAP retention and thus open complex formation (Busby & Ebright, 1999; Liu et al., 2017).

Recent advances in cryo-EM have also led to a range of new insights into class II CAP promoters and the Mer family of class II activators. Work from Liu and co-workers have found that in the *gal* promoter, CAP binds to the -41 region of the DNA. Most notably, they found that RNAP- σ^{70} -CAP-DNA complexes exists in a classical closed and widened RNAP clamp state. In the closed state, CAP interacts with the β flap, α -NTD, as well as σ^{70} R4.2. However, in the widened state, CAP interaction with the α -NTD and R4.2 is lost; instead, CAP interacts with the base of the β i9 coiled-coil, causing a repositioning of the β subunit. This suggests that class II CAP binding activates transcription and DNA loading via interactions with the β subunit to trigger clamp opening (Shi et al., 2020a).

Another type of class II activators are the MerR family, which includes CueR, EcmR and BmrR. The MerR family activates promoters with a lengthened 19-20bp spacers between the -10 and -35 elements. The MerR family exists as a winged helix-turn-helix domain connected by a α -helical coiled-coil motif that serves as the site of dimerization. Upon binding to ligand, MerR activators twist and kink the spacer DNA and engage with σ^{70} 4.2 and σ^{70} R3.1 to encapsulate the spacer DNA and promote transcription bubble melting (Fang et al., 2020; Fang et al., 2021; Yang et al., 2021).

1.6.1.5 *RbpA and CarD activators*

RbpA and CarD are Actinobacterial σ tethers. Work from Campbell and co-workers showed that RbpA enforces the interactions between RNAP core and σ^{70} by tethering the σ NCR and β' ZBD, which interacts with σ 4. RbpA also interacts with

the backbone DNA at the extended -13, -18 and -19 regions of the DNA, stabilising the nascent transcription bubble. CarD binds to the β lobe and interacts with the minor groove at the -10 region, forming salt bridge interactions with the phosphate backbone to prevent bubble collapse during transcription bubble melting. Structural models suggest CarD and RbpA work together to bend and stabilise the transcription bubble intermediates (Bae et al., 2015; Hubin et al., 2017).

1.6.2 σ^{54} factor

σ^{54} was first discovered by Kustu and Magasanik in promoters associated with nitrogen fixation under nitrogen depleted conditions (Hirschman et al., 1985; Hunt & Magasanik, 1985). It has since been identified in over 60% of prokaryotic genomes and is responsible for RNAP recruitment to gene promoters associated with stress response and nutrient depletion. σ^{54} shares no significant sequence or structural homology with σ^{70} , with the exception of the helix-turn helix (HTH) motif (Gao et al., 2020).

σ^{54} is mechanistically distinct from σ^{70} in that it forms an energetically favoured closed complex by binding to consensus -24 and -12 regions upstream of the TSS (Figure 1.6.4). RNAP- σ^{54} does not spontaneously isomerise to open; instead, the activation of σ^{54} -dependent promoters are activated by a group of clade 6 AAA+ proteins called bacterial enhancer binding proteins (bEBPs), that are regulated by a range of different signals. σ^{54} binds to promoters independent of RNAP, and of the 70 σ^{54} promoters in *E. coli*, there are thought to be 110 σ^{54} molecules *in vivo* during both exponential and stationary phases (Jishage et al., 1996; Zhao, Liu & Burgess, 2010).

This suggests that some σ^{54} is pre-bound to the strong promoters, enabling rapid activation upon receiving stress stimuli and recruitment of RNAP and bEBPs.

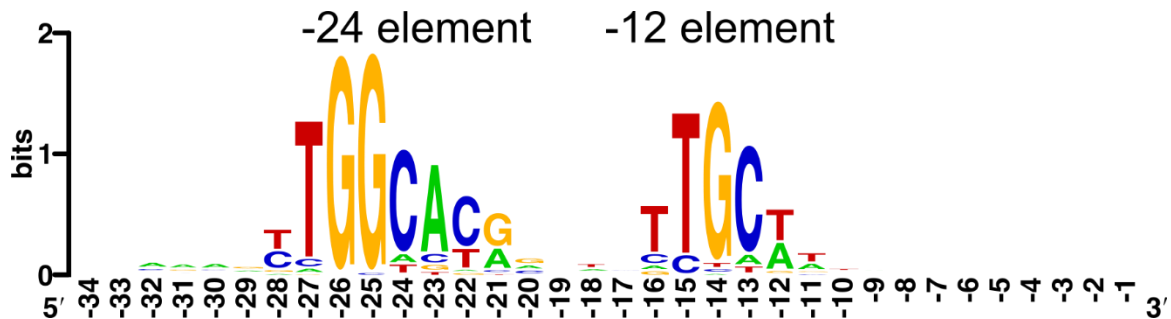


Figure 1.6.4: Sequence logo of σ^{54} promoters taken from iPro54 database (Lin et al., 2014). All sequence logos were generated using WebLogo (Crooks et al., 2004).

σ^{54} is made up 3 regions (RI-RIII) based on sequence conservation. RIII can be subdivided into 3 structural subdomains (Figure 1.6.5a, e, f). RI consists of a flexible N terminus, followed by 2 α helices, $\alpha 1$ and 2. The $\alpha 1$ helix interacts with the -11/-12 region of the DNA, $\alpha 2$ sits in a crevice formed by β' -CC. RI and RIII are connected by RII, which is a poorly conserved flexible acidic linker, that descends into the RNAP catalytic cleft and occupies the dsDNA and DNA-RNA hybrid binding sites. RII in *K. oxytoca* and *E. coli* are split into 3 sections (RII.1-3) corresponding to the downstream DNA, bridge helix and RNA exit channel locations within the catalytic cleft (Glyde et al., 2017; Yang et al., 2015).

Region III core binding domain (CBD) is located at the RNA exit channel and consists of an α -helical bundle, that interacts with the β -flap and β subunit (Wong, Tintut & Gralla, 1994; Yang et al., 2015).

Region III extra-long-helix helix-turn-helix (ELH-HTH) is the largest structural subdomain, and bridges across the β and β' subunits, as well as interacting with the -11/-12 region of the DNA (Figure 1.6.5b-d). ELH-HTH sits atop of parts of the β -flap and the base of the β i9 coiled-coil. The N-terminus of ELH interacts with the β' coiled-coil. Crucially, the ELH contains a conserved R336 residue, that is essential for open complex inhibition – R336A mutants are “bypass” mutants that can transcribe without activation when using preopened DNA, although the mechanism is unclear. The RpoN domain is a winged-helix-turn-helix motif that interacts with the -24 consensus region of the promoter, binding and stabilising the major groove (Glyde et al., 2018; Zhang & Buck, 2015).

In the closed complex, the RI and ELH form a barrier blocking DNA entry; consequently, bEBPs are required to remodel the closed complex to enable DNA melting and loading to produce the open complex (Danson et al., 2019).

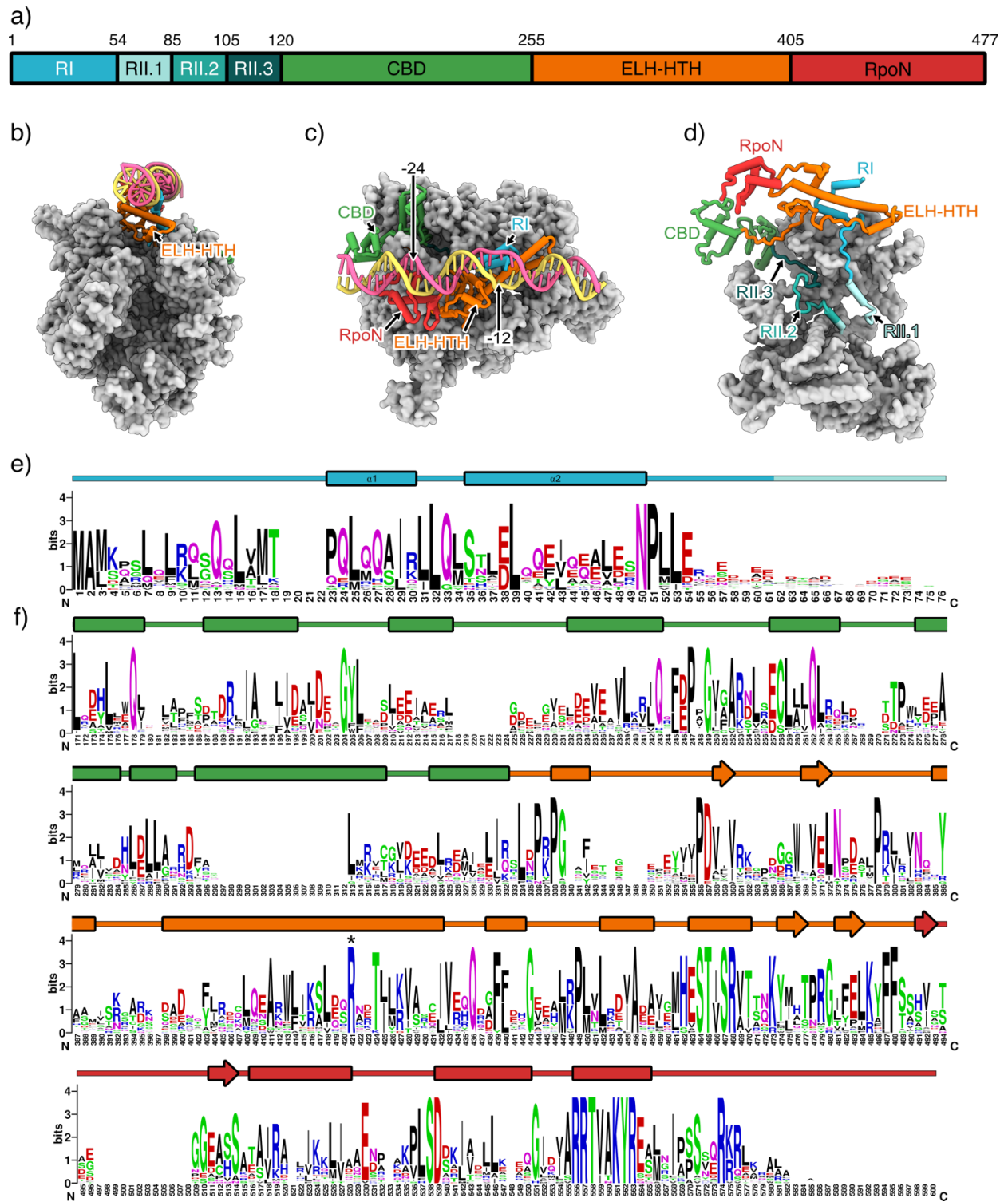


Figure 1.6.5: a) Domain map of σ^{54} . b-d) Structure of *E. coli* σ^{54} closed complex (5NSR). Region II was not resolved in this structure. b) Downstream face of closed complex clamp. c) Top-down view looking into the catalytic site. d) cross section of claw. α and β subunits are hidden for clarity. e) Sequence logo of region I. f) Sequence logo of region III. Conserved R336 indicated by an asterisk.

1.7 ATPases associated with various cellular activities (AAA+)

AAA+ are a superfamily of proteins that couple the energy release from ATP hydrolysis to the mechanical movement and remodelling a wide range of substrates. AAA+ proteins are used throughout the central dogma to translocate and melt nucleic acids to enable DNA replication, transcription and repair (Table 1.7.1) (Puchades, Sandate & Lander, 2020).

Table 1.7.1: Classification of AAA+ protein clades. PS1BH=part of the Pre-sensor 1 β hairpin superclade.

Adapted from Puchades, Sandate & Lander (2020) and Khan, White & Brunger (2022).

Clade	Name		Function	Examples
1	Clamp loader		DNA replication and repair	RFC3, T4 clamp loader, HolB
2	Initiator		DNA replication	DnaA, DnaC, ORC/Cdc6
3	Classic	Type I	Protein polymer disassembly	VPS4, Fidgetin, Bcs1
		Type II	Vesicle fusion and protein disaggregation	NSF, p97, Hsp104
		Type III (Protease)	Protein degradation	Yme1, AFG3L2, 26S protease
4 (PS1BH)	SFIII helicase		DNA replication	E1 helicase, SV40 initiation complex, REP68
5 (PS1BH)	HCLR (HslU/ClpAB/Lon/RuvB)		Protein folding, degradation, transport, helicases, etc.	ClpAB, Lon, RuvB
6 (PS1BH)	Helix-2-insert		Restriction enzymes, transcription factors	bEBPs, McrB
7 (PS1BH)	Pre-sensor 2-insert		DNA replication, protein translocation, ribosomal assembly etc.	MCM, Dynein, MoxR

AAA+ proteins are classified loosely into 7 clades based on sequence homology and phylogeny, all containing conserved motifs for ATP hydrolysis (see section 1.7.2). All AAA+ proteins consist of protomers containing a ATP binding pocket nestled between an $\alpha\beta\alpha$ sandwich lobe, a poorly conserved α -lid lobe, and the

arginine-finger (R-finger) motifs from the adjacent protomer in the oligomeric ring (Figure 1.7.1a) (Chen et al., 2010).

ATP hydrolysis results in a conformational shift between the $\alpha\beta\alpha$ sandwich and α -lid lobes, as well as disengagement of the R-fingers from the adjacent protomer. This enables detection of hydrolysis state between adjacent protomers and propagation of ATP hydrolysis-driven conformational changes throughout the ring, which is subsequently used to remodel the substrate (Jessop, Felix & Gutsche, 2021).

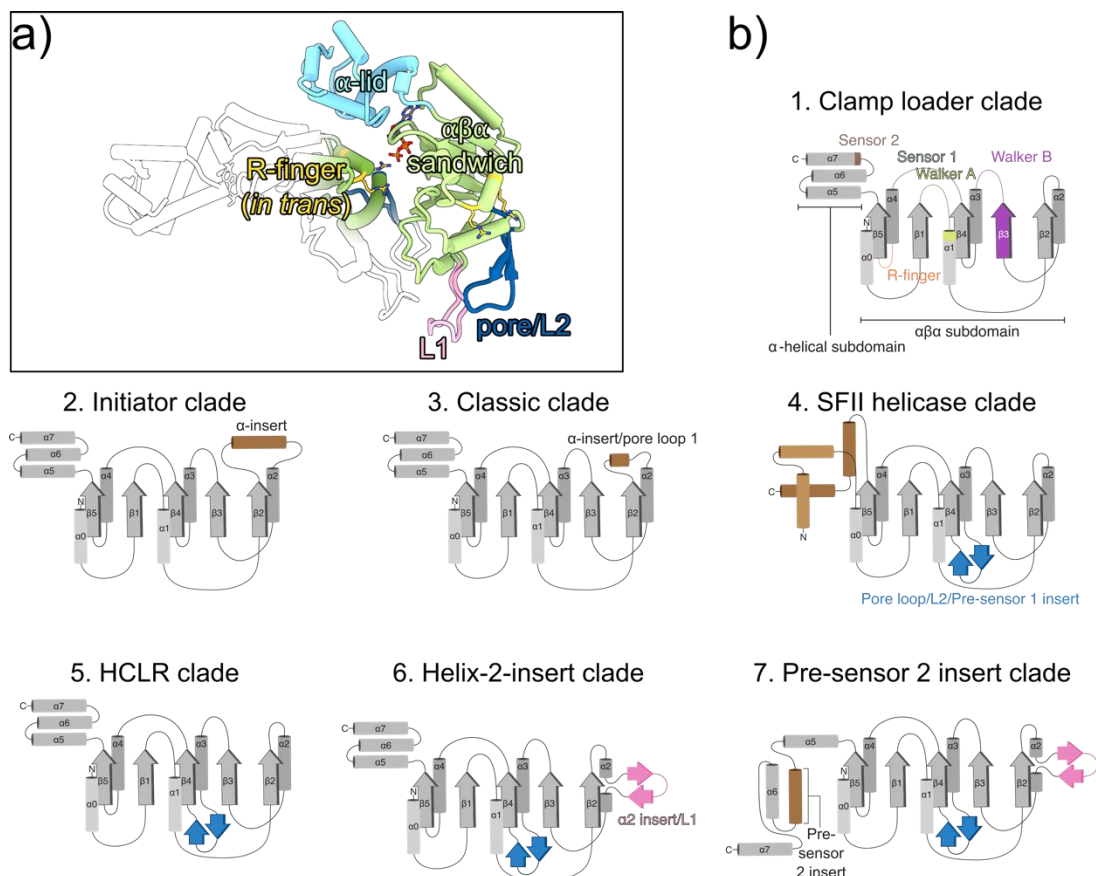


Figure 1.7.1: a) structure of AAA+ interprotomer interface (PDB ID: 3M0E) showing *in trans* activity of R-fingers from adjacent subunit and ATP binding site within the α -lid and $\alpha\beta\alpha$ sandwich. b) overview of AAA+ protein clades and unique insertions. Figure adapted from Seraphim & Houry (2020).

The structural motifs of different AAA clades are shown in Figure 1.7.1b. Clade 1 are the most minimal AAA+ proteins and are heteropentamers that loads DNA polymerase sliding clamp onto primer-template DNA junctions. The sliding clamp enables recruitment of replication and repair factors and is essential in maintaining processivity during DNA replication (Cvetkovic et al., 2022).

Clade 2 contain a α -helical insert between $\beta 2$ and $\alpha 2$ and consist of DNA replication initiation proteins responsible for assembly of the DNA replisome at the origin of replication. The two families in clade 2 are the ORC/Cdc6 family and DnaA/DnaC family. The ORC1-5 forms a heteropentamer around the origin DNA with Cdc6 and utilises ATP to enable the loading of the MCM helicases onto the replication origin (Feng et al., 2021). DnaA/C similarly binds to the origin of replication and recruits the DnaB helicase (Khan, White & Brunger, 2022).

Clade 3 is the largest and best characterised group of AAA+ proteins. Clade 3 AAA+ proteins are distinct due to the presence of a conserved pore loop 1, a flexible linker with a shortened α -helix. Clade 3 AAA+ proteins typically form a hexameric ring around their substrate and are involved in protein translocation, unfolding or degradation. The central pore of the ring is lined with pore loop 1, which consists of a conserved aromatic residue that is essential for non-specific translocation. Clade 3 is often further classified into type I-II (with some suggesting a type III for proteases) based on domain organisation and function. With the exception of Bcs1 which utilises a concerted mechanism (Kater et al., 2020), most AAA+ adopt a spiral staircase arrangement of pore 1 loops, and are suggested to utilise a classical sequential (also known as hand-over-hand) mechanism for substrate translocation,

where 1 ATP hydrolysis event results in the translocation of adjacent protomers down the hexamer axis, with the 6th protomer disengaging from the substrate in the *apo* or ADP bound state, before re-engaging upon nucleotide exchange (Figure 1.7.2) (Puchades, Sandate & Lander, 2020).

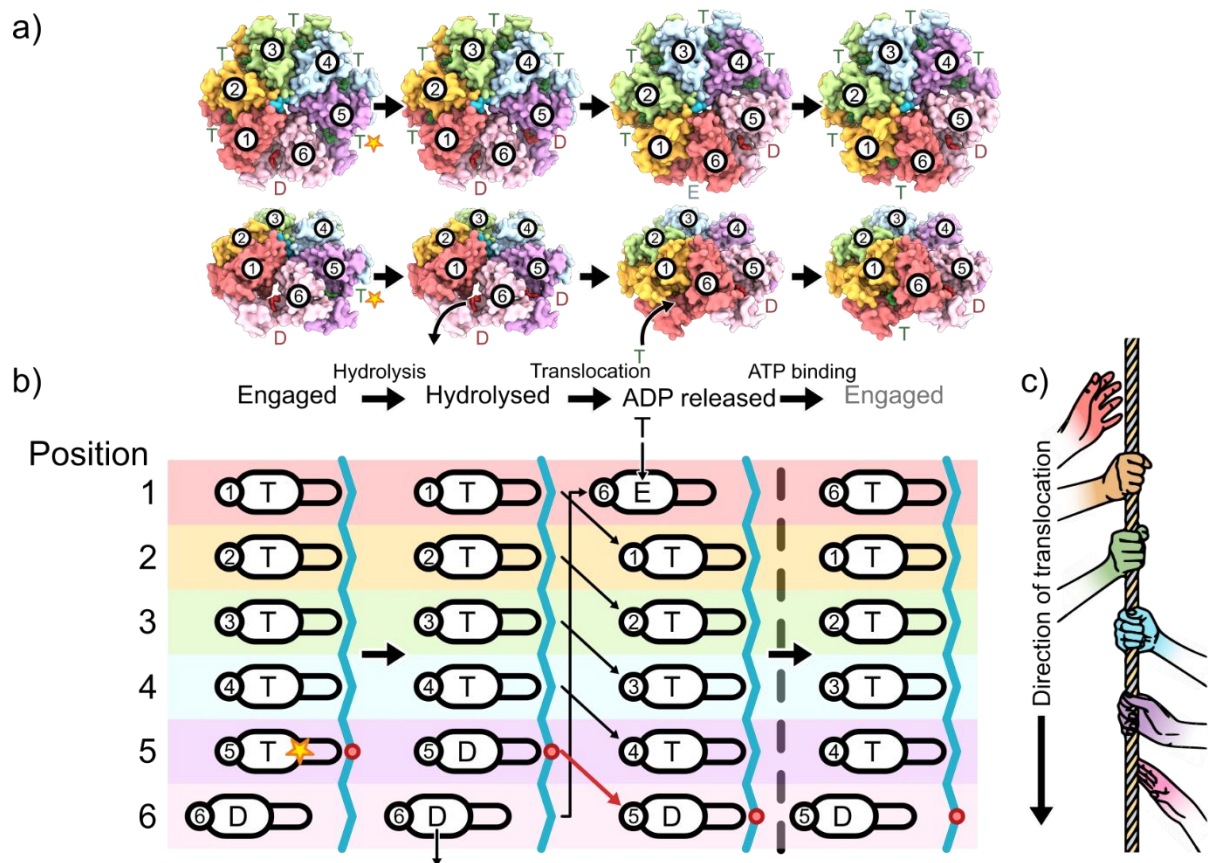


Figure 1.7.2: Classical mechanism of AAA+ driven substrate translocation using a sequential hand over hand mechanism. a) Hexamer model showing ATP hydrolysis driven movement of subunit 5 during 1 round of ATP hydrolysis. b) A schematic representation of the movements of the protomers during 1 round of ATP hydrolysis. c) hand-over-hand analogy of translocation.

Adapted from (Puchades, Sandate & Lander, 2020)

Clades 4-7 belong to the pre-sensor-1 β hairpin (PS1BH) superclade. This superclade is unique due to the presence of a different pore loop formed by an

insertion between $\alpha 3$ and $\beta 4$ that projects outwards from the AAA+ core, also known as the pre-sensor-1 β hairpin. Much like pore loop 1 in classical AAA+ proteins, the PS1BH lines the central pore of clade 4-7 members and makes direct interactions with substrate, similar to pore loop 1 found in clade 3 AAA+ proteins (Khan, White & Brunger, 2022).

1.7.1 Bacterial Enhancer binding proteins

bEBPs are a family of clade 6 hexameric AAA+ proteins. bEBPs generally contain 3 domains, with the central AAA+ domain (C) being absolutely conserved. The C domain is flanked by an N-terminal regulatory (R) domain and a C-terminal DNA binding domain (D). bEBPs are initially recruited via their D domain to upstream activating sequences (UAS) sites 80-150 bp upstream of the TSS; upon activation, by the R domain or *in trans* activation, integrative host factor (IHF) and DNA bending proteins facilitate looping of the bEBPs to the closed complex (Gao et al., 2020).

The R domain are the most diverse and can activate or inhibit the C domain in response to a plethora of signals, including but not limited to phosphorylation as part of two component systems, ligand binding to metal ions or benzene derivatives, or protein-protein interactions. The mechanism of regulation is also variable, with mechanisms ranging from UAS binding, hexamer formation, binding to σ^{54} and ATP hydrolysis (Table 1.7.2) (Bush & Dixon, 2012).

Table 1.7.2: An overview of the different groups of bEBPs. Abbreviations: RR = Response regulator, PRD = Phosphotransferase System Regulation Domain, PAS = Per-ARNT-Sim domain, V4R = 4-vinyl reductase, ACT = aspartokinase, chorismate mutase, and TyrA, GAF = cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA. (Reproduced with permission from (Gao et al., 2020))

Group	Examples	R domain	Mode of regulation	C domain	D domain
1	NtrC, NtrC1, NtrC4, ZraR, DctD	RR/PRD	Phosphorylation from 2 component systems	Clade 6 AAA+	HTH
2	DmpR, AcoR, PheR, PrdR, XylR	PAS/V4R/ACT	Ligand binding	Clade 6 AAA+	HTH
3	NorR, NifA, FhlA	GAF	Ligand binding	Clade 6 AAA+	HTH
4	PspF, HrpR/S	N/A	Protein binding <i>in trans</i>	Clade 6 AAA+	HTH
5	FlgR, CtcC	N/A	Phosphorylation from 2 component systems	Clade 6 AAA+	N/A

1.7.2 Conserved AAA+ motifs in the C domain

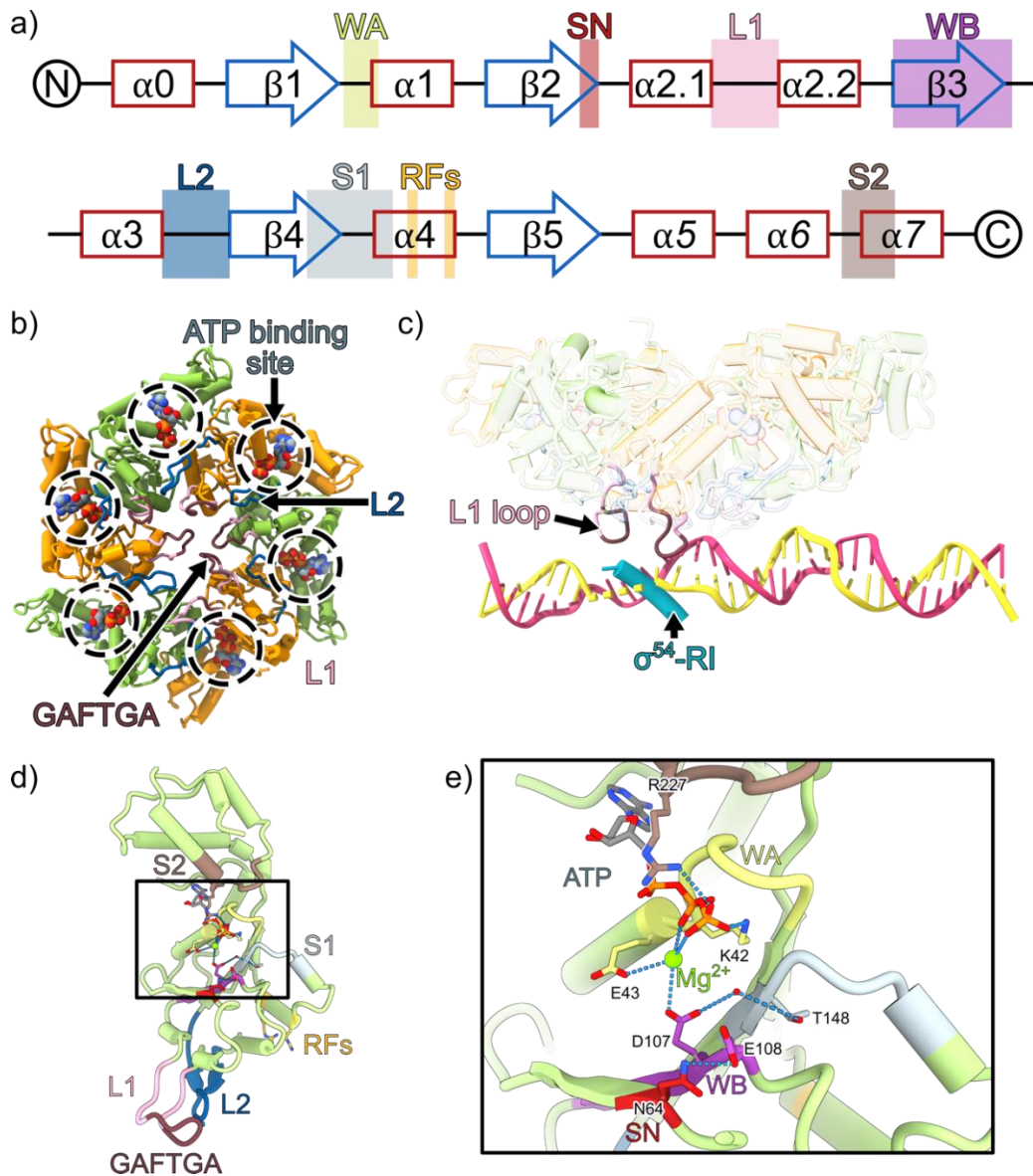


Figure 1.7.3: a) Schematic diagram of clade 6 AAA+ protein secondary structure. b) Model of PspF₁₋₂₇₅ hexamer showing ATP binding sites. Model was created by fitting in ATP molecules from the PspF₁₋₂₇₅ structure (2C96) onto the intermediate complex structure (5NSS). c) PspF₁₋₂₇₅ hexamer contacting DNA and σ^{54} -RI via its L1 loops. d) NtrC1 monomer crystal structure (PDB ID: 3M0E) showing the conserved domains found in AAA+ proteins, as well as the L1/L2 loops. Area highlighted is the active site, as enlarged in e). e) The active site of NtrC1. The E108 residue was modelled using PspF₁₋₂₇₅ crystal structure (2C96). The residues are numbered as in *E. coli* PspF₁₋₂₇₅. Figure adapted from Gao et al. (2020).

The nucleotide binding site consists of a variant walker A (GXXXXGK[D/E] where x = any residue) motif that binds ATP and a walker B motif (hhhhDE where h = any hydrophobic residue) that coordinates the Mg^{2+} cation to enable ATP hydrolysis (Figure 1.7.3a, b). Oligomerisation is facilitated by the sensor 2 motif, arginine finger motif and Mg^{2+} ion that forms salt bridges with the β and γ phosphate on ATP. The Mg^{2+} ion and walker A lysine stabilise the negative charge on the hydrolysis intermediates during the reaction mechanism (Figure 1.7.3d,e) (Gao et al., 2020).

bEBPs belong to clade 6 of the AAA+ superfamily alongside GTPase McrB; it is characterised by a β -hairpin in the $\alpha 2$ helix, also known as the L1 loop. It has been suggested that the L1 or L2 loops deliver the power stroke (Joly & Buck, 2011). The L1 loop contains the absolutely conserved GAFTGA motif, which has been shown to be essential for activation, sensing DNA conformation and a direct σ^{54} interaction (Figure 1.7.3c) (Dago et al., 2007; Zhang et al., 2009).

Other classical and HCLR AAA+ proteins were shown to channel substrate, either DNA or peptide, through a central pore via a sequential or partially sequential mechanism, with protomers being arranged in a spiral staircase arrangement (Figure 1.7.2). The pore loop found in other AAA+ proteins are analogous to the L2 loop in bEBPs, and delivers the power stroke (Lin, Shorter & Lucius, 2022).

However, the exact mechanism for ATP hydrolysis coordination for bEBPs were unknown, and it is not clear the exact roles of the clade 6-specific L1 or pore (L2) loops. It was previously shown that PspF interacts with both σ^{54} RI and DNA -11/-12 regions, and hexamerization was important for activity, suggesting a sequential or

partially mechanism (Joly & Buck, 2011; Joly, Schumacher & Buck, 2006; Schumacher et al., 2004). Furthermore, it is not clear whether bEBPs deliver the power stroke to promoter DNA, σ^{54} , or both, as previous structures were too low resolution to determine the nature of the interaction with the RNAP- σ^{54} -DNA and L1/L2 loops of PspF₁₋₂₇₅ (Glyde et al., 2017).

1.8 σ^{54} -mediated open complex formation

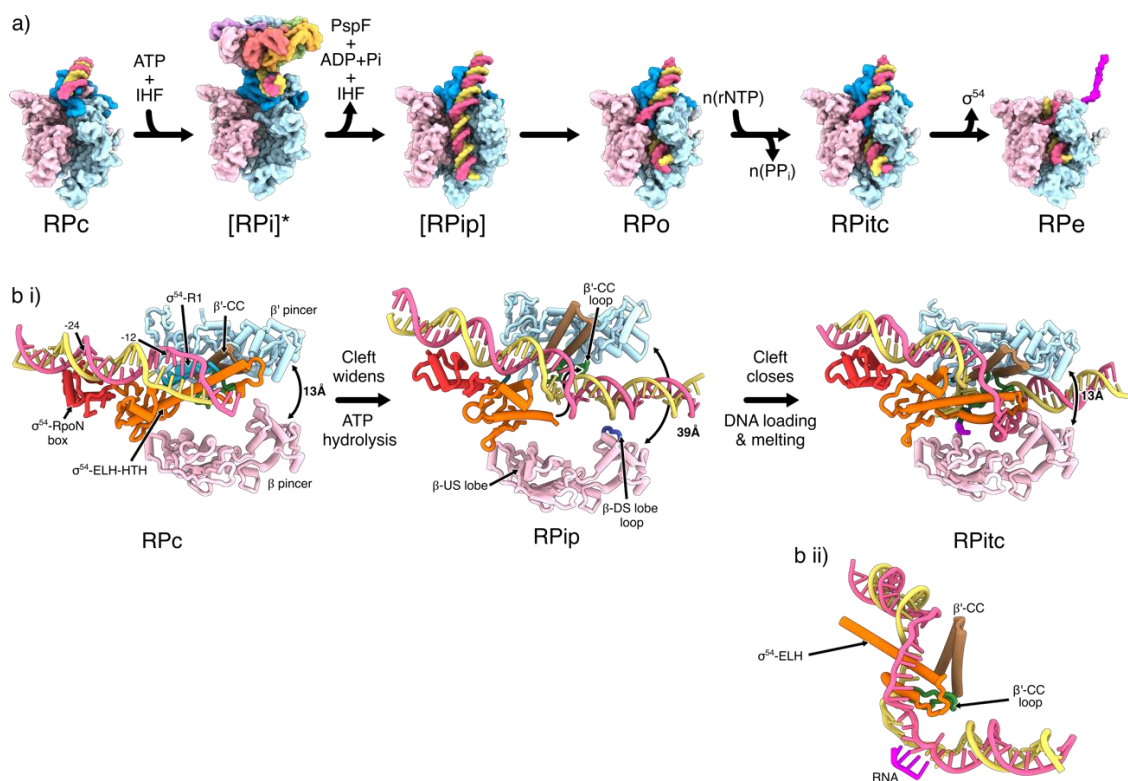


Figure 1.8.1: Structures in the open complex formation pathway. a) Overall structures of the different RPc (5NSR), RPi (5NSS), RPip (6GH6), RPo (6GH5), RPitc (6GFW) and RPe (taken from expressome structure 6XDR) complexes, viewed from the downstream face. [] = Transient intermediate. Asterisks indicates use of ADP.AIFx to capture hydrolysis intermediate state. b i) DNA melting and clamp opening during open complex formation. View from the top of the cleft, showing the RNAP pincers only for clarity. Region I is not resolved in RPip and RPitc. b ii) side view of the transcription bubble in RPitc structure. RNA shown in magenta.

The mechanism of σ^{54} mediated transcription initiation was first proposed by Glyde et al., from 4 separate trapped intermediate states - closed (RPc), intermediate (RPi), partially loaded (RPip) and open (RPo) - using a combination of R336A σ^{54} bypass mutants, transition state analogue ADP aluminium fluoride (ADP.AIFx, where x=3-4), preopened or mismatched *nifH* promoter DNA (Figure 1.8.1).

The structure of the RPc complex shows that the DNA is blocked from entering the RNAP catalytic site by the ELH and RI helices, which form a gate blocking DNA loading into the cleft, held together by conserved hydrophobic residues on $\alpha 1$ and $\alpha 2$, as well as ELH (Figure 1.8.1b, Figure 1.6.5e-f). RI and ELH are located at the -12 region and distorts the downstream minor groove, with the CBD blocking the RNA exit site. This mechanism is supported by mutagenesis studies of RI and ELH. Deletion of RI enabled open complex formation with preopened, but not wildtype DNA. Mutations of residues 33-37, located in RI $\alpha 2$, resulted in unstable open complexes and transient bubble melting (Gallegos & Buck, 2000) and mutation of leucine residues on 25-31 resulted in bypass phenotypes (Syed & Gralla, 1997). Substitution mutations of ELH, specifically R336A, also resulted in bypass phenotypes, albeit with reduced binding to core RNAP and DNA in all cases (Chaney et al., 2001).

The RPi structure was determined using a AAA+ truncation mutant of PspF (PspF₁₋₂₇₅) lacking the UAS binding domain, a bEBP involved in phage shock response. The structure was trapped using -12/-11 mismatched *nifH* 63bp DNA and ADP.AIFx. The structure, whilst low resolution, enabled docking of a PspF hexamer, showing that the L1/L2 loops descend from the hexamer plane and contacts the -

12 region of the DNA (Figure 1.7.3c, Figure 1.8.1a), which was further corroborated by biochemical and crystallographic studies (Bordes et al., 2003; Chaney et al., 2001; Sysoeva et al., 2013).

DNA loading intermediate and the open complexes were obtained using R336A bypass mutants of σ^{54} and preopened DNA (-10 to -1). In RPip, the clamp is in the completely open state, with the ELH retracted and lowered to facilitate DNA loading. The DNA is kinked at -10 to -5 by 30°, which aids the melting of the transcription bubble. In RPo, the cleft is closed around a 13 nt transcription bubble that is fully melted from -11 to +2. The ELH is inserted between the DNA strands and interacts with the β' -coiled-coil loop, aiding separation and preventing re-annealing (Figure 1.8.1b ii). In both structures, the RI and RII regions are unresolved, indicating this region is highly dynamic. Both RPip and RPo were obtained from the same dataset, suggesting that open complex formation occurs by clamp opening coupled with transcription bubble opening, with σ^{54} serving dual functions, by both facilitating DNA loading and stabilising the transcription bubble. The role of ELH was also supported by biochemical studies, showing that deletion of 310-328 on ELH, similar to 33-37 deletion mutants, produced unstable open complexes (Chaney et al., 2001). However, the role of ATP hydrolysis in DNA loading remains unknown due to the use of the R336A bypass mutant (Gao et al., 2020; Glyde et al., 2018).

1.9 Mechanisms of promoter escape

Promoter escape is the transition from the open complex to the processive elongation state. Promoter escape involves the accumulation of tension from

nucleotide addition that leads to three distinct possibilities: productive transcription, complex dissociation, or abortive/reiterative synthesis. In the productive escape pathway, the build-up of tension leads to the disruption of σ interactions with the promoter region as well as interactions between σ and the RNAP core, leading to σ displacement. Productive promoter escape has 5 key features: extension of RNA to a threshold length, usually 11-15 nt (Lee & Borukhov, 2016). Secondly, RNA-DNA hybrid extension and redirection of the RNA 5' end into the RNA exit channel after 9nt by the β -lid. Thirdly, displacement of the σ factor subdomains occupying the RNA exit channel (CBD in σ^{54} , R3.2 in σ^{70}), followed by displacement σ of the upstream promoter DNA-binding domains (RpoN and ELH-HTH in σ^{54} , R4 in σ^{70}). Finally, rewinding and relaxation of the upstream edge of the transcription bubble (Winkelman, Nickels & Ebright, 2021).

The exact mechanisms determining productive vs non-productive promoter escape remain poorly understood in bacteria; single molecule FRET experiments have shown that RNAP- σ^{70} undergoes a critical 6nt pause point, where transition into productive transcription is partially dependent on rNTP concentrations (Dulin et al., 2018). Other single molecule experiments have also shown that σ^{70} can remain bound throughout elongation, although the structural basis for σ^{70} retention remains unclear (Harden et al., 2016; Mooney & Landick, 2003). Furthermore, in certain promoters with repetitive G or C sequences, RNAP- σ^{70} can also undergo rounds of reiterative transcription, in which RNAP synthesis 2-3 nt of complementary RNA nucleotides followed by up to hundreds of non-complementary nucleotides as a result of RNA slippage (Liu et al., 2022).

There are two mechanisms of tension accumulation in RNAP that can lead to σ displacement: DNA scrunching, and RNA 5' displacement.

1.9.1 RNA 5' displacement

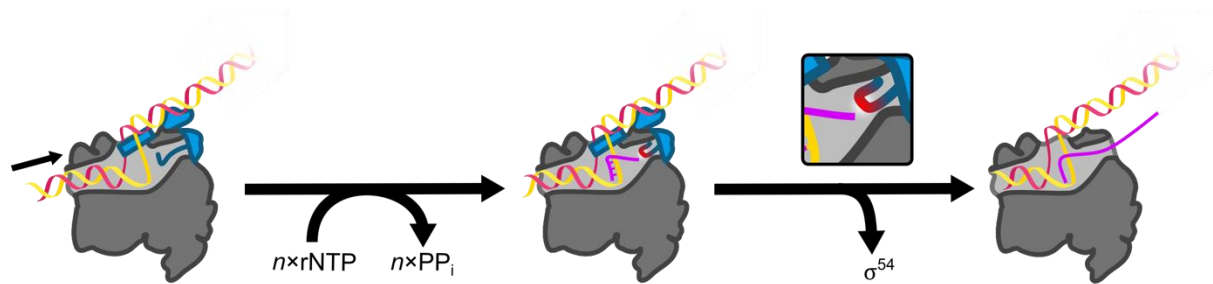


Figure 1.9.1: Schematic mechanism of RNA displacement.

The 5' RNA displacement mechanism posits that the 5' end of the RNA actively pushes R3.2 out of the RNA exit channel. As the RNA chain grows, tension is built up against the σ^{70} R3.2, leading to eventual displacement and dissociation. This theory is supported by mutagenesis and deletion mutants of R3.2 and recent crystallographic studies from the Ebright group (Li et al., 2020).

Crystal structures of σ^A , and ECF factors σ^H and σ^L demonstrates that starting from 5nt in σ^A , 8 to 9 nt in σ^H and 6nt in σ^L the R3.2 “ σ finger” folds back to accommodate the growing RNA chain. The length at which the σ finger made contact depends on the length of the R3.2 loops. The authors proposed that σ^{70} R3.2 folding backwards acted synergistically with the DNA scrunching to accumulate tension, resulting in a “protein spring” that leads to σ displacement during promoter escape (Li et al., 2020).

However, these structures were not determined using *de novo* synthesised RNA, and instead used a scaffold containing pre-synthesised RNA and without the upstream edge of the transcription bubble (Li et al., 2020). Consequently, the effects of DNA scrunching could not be determined.

1.9.2 DNA scrunching

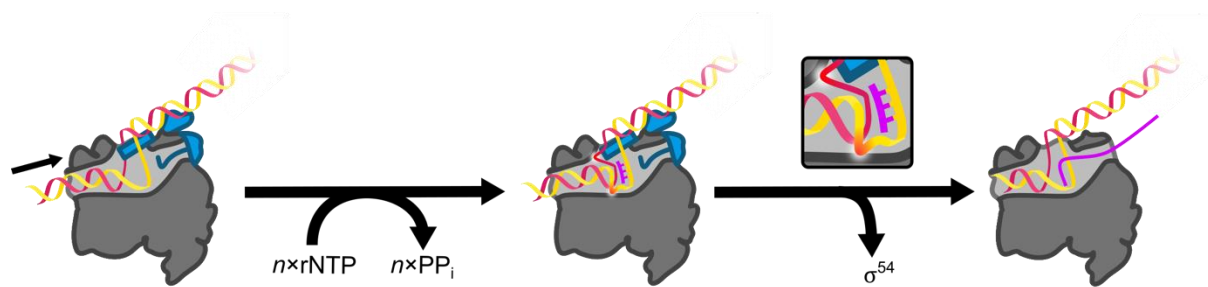


Figure 1.9.2: Schematic mechanism of DNA scrunching driven σ displacement.

DNA scrunching of the transcription bubble was proposed by a series of FRET and magnetic bead experiments performed by the Ebright and Strick groups. In both approaches they found that RNAP uses a scrunching mechanism during promoter escape, in contrast to inchworming (which would involve expansion of RNAP) or transient excursion (which would involve back and forth movement of RNAP along the promoter). In the scrunching model, the upstream promoter elements remain bound to the consensus regions, whilst downstream DNA is pulled into the active site, leading to enlargement and scrunching of the transcription bubble (Kapanidis et al., 2006; Revyakin et al., 2006).

This mechanism is supported by recent crystal structures from the Steitz group. The Steitz group determined 3-6nt RNA-bound σ^S complexes by using *in crystallo* RNA synthesis and dinucleotide primers that bind to different positions on the template

strand (Zuo et al., 2020). They observed expansion of the transcription bubble from 13nt to 16 nt upon addition of the appropriate NTPs. However, they found that soaking crystals in NTPs that was expected to produce a 7nt complex produced only 6nt with an additional NTP in the $i+1$ site. Similarly, addition of specific individual NTPs to a 4nt primer RNA led to crystal structures with 3nt, indicating backtracking and cleavage. At 6nt, they did not observe R3.2 blocking the path of the RNA, nor R3.2 folding between 4nt $\rightarrow$ 6nt, which lead the authors to suggest that R3.2 plays little to no role in the accumulation of stress (Zuo et al., 2020).

Mutation of residues located in σ^{70} R3.1 that interact with the β' lid increased the mobility of the $\sigma 3$ domain enabled a post-translocated 7nt and pre-translocated 8nt complex to be determined, suggesting an energetic barrier between 6nt $\rightarrow$ 7nt that requires displacement of σ^{70} R3.1 before longer extension can occur. It is also worth noting that although R3.2 was not resolved in these structures, the authors suggested this was due to mutagenesis, and the 7-8nt structures represented an artificial state. Instead, template strand scrunching against $\sigma 3$ was suggested to play a major role in displacement. However, due to the effects of crystallisation and the high concentrations of NTPs used, it was not ruled out that crystal contacts could have prevented RNA extension beyond 6nt (Zuo et al., 2020).

1.10 Project Aims

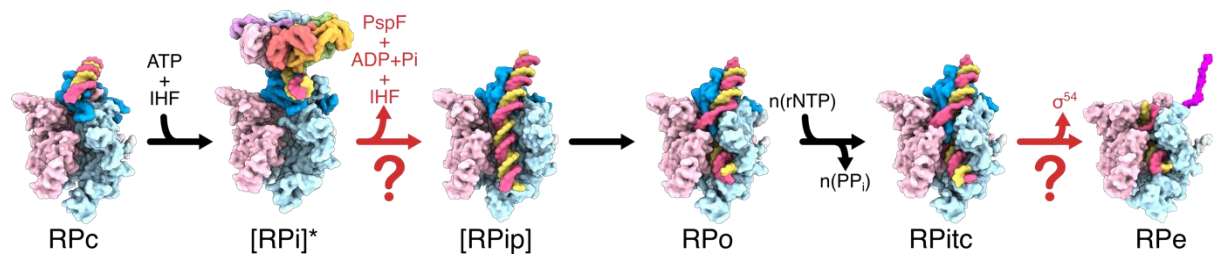


Figure 1.10.1: Summary of the structural transitions being studied in this project.

The aim of this project is to determine the mechanisms of PspF₁₋₂₇₅-mediated open complex formation. We determined 3 different PspF₁₋₂₇₅-engaged intermediate states using a range of different pre-opened DNA bubble sizes and ADP.AIFx nucleotide hydrolysis analogue elucidating the role of PspF₁₋₂₇₅ during DNA melting.

As shown in Figure 1.6.5d and Figure 1.6.2d, the interface between the σ^{54} and core RNAP are different from σ^{70} ; consequently, we sought out to determine the mechanism of σ^{54} -mediated promoter escape. We determined 6 structures of longer transcript states using R336A RNAP- σ^{54} R336A bypass mutants, determining new conformational changes in σ^{54} RII and ELH on the pathway to the elongation state.

Chapter 2

Cryo-electron Microscopy

2.1 Background

Single particle cryo-electron microscopy (cryo-EM) is the technique that enables the determination of 3D volumes of biomolecules by aligning noisy 2D projection images (Figure 2.1.1). Cryo-EM, as its name suggests, involves observing biological molecules in a 10-100 nm layer of vitreous ice at cryogenic temperatures (<123 K). Samples are flash-frozen by depositing samples onto 3 mm grids, blotting and plunging them into liquid ethane mixtures cooled to ~ 90 K. Vitrification preserves molecular structure to enable stain-free imaging without the need to crystallise the protein of interest (Dubochet et al., 1988).

The last decade has seen an exponential rise in the application of single particle cryo-EM to study biological molecules. Recently, multiple atomic resolution structures, on-par with X-ray crystallography has been achieved (Danev, Yanagisawa & Kikkawa, 2021; Nakane et al., 2020; Yip et al., 2020; Zhang et al., 2020). This is due to numerous advances in sample preparation, data collection and image processing algorithms. This chapter focuses on the fundamental theory underlying cryo-electron microscopy.

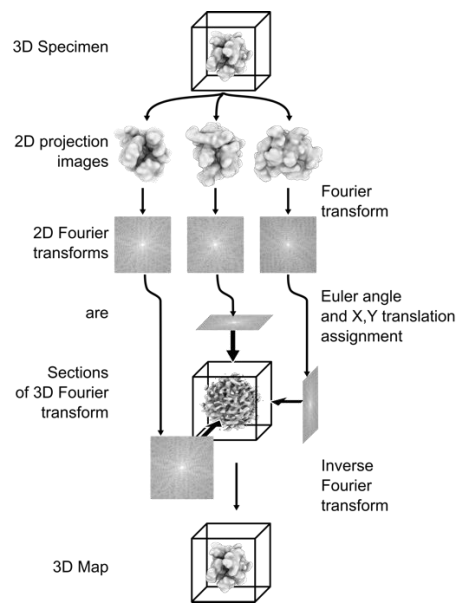


Figure 2.1.1: Schematic of 3D reconstruction method, demonstrating the principles of Fourier slice theorem - Fourier transforms of 2D projection images represent slices of a 3D Fourier transform. Discussed further in section 2.3.

2.2 Resolution limits

Electromagnetic radiation can exist as particles (photons) and waves (Figure 2.2.1).

Electromagnetic waves contain amplitude (A), wavelength (λ) and phase (ϕ), and

can be expressed as follows:

$$y = A \cos\left(\frac{2\pi x}{\lambda} + \phi\right)$$

Where A=amplitude, λ =wavelength and ϕ =phase

with frequency being defined as

$$f = c/\lambda$$

where c=speed of light

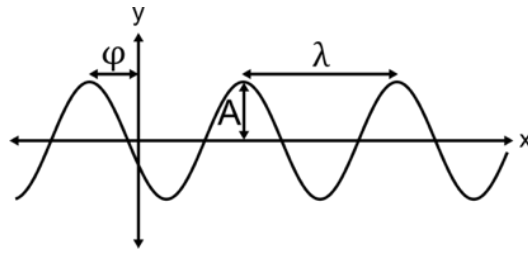


Figure 2.2.1: Overview of wave equation. All waves contain amplitude, phase and wavelength

All forms of microscopy utilise illumination and detection of incident electromagnetic radiation that interacts with a specimen to produce resultant waves. The resultant waves then pass through an aperture to reach the detector, that records the image (Egerton, 2016).

Specimens can be described as a collection of points arranged in 3D space, with different distances (D). Resolution is defined by the smallest distance that can be measured by the imaging system. When radiation hits a single point, this results in radiation of waves that emanates from the single point. In an evenly-spaced array of points, the radiation from each point will interfere with each other, either constructively or destructively dependent on the phase (Figure 2.2.2a). This results in diffraction, with the scattering angle (θ) being determined by the following formula:

$$\sin(\theta) = \lambda/D$$

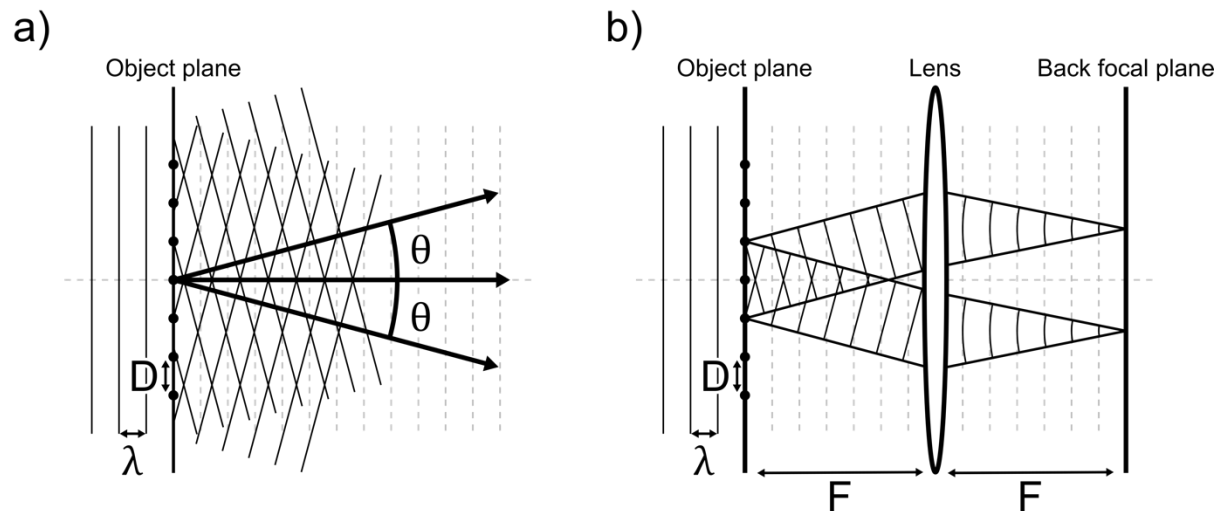


Figure 2.2.2: a) Incident wave hitting an array of evenly spaced points. The resultant radiation of waves from each point interfere constructively or destructively dependent of the wavelength and phase, resulting in diffraction at specific angles θ . b) Lenses focus the diffracted light onto a back focal plane.

Consequently, the scattering angle is inversely proportional the distance between two points (D), which means that the larger the scattering angle, the higher the resolution ($D_{\min}$) of the feature described by the two points. Therefore, apertures in the microscope result in occlusion of waves with large θ from hitting the detector, meaning that aperture diameter (d) is proportional to maximum achievable resolution. This diffraction limit of spatial resolution is defined by Releigh's criterion, which is defined by

$$D_{\min} = \frac{0.61\lambda}{NA}$$

Where $NA = \text{numerical aperture} = n\sin(\theta)$

Where $n = \text{refractive index of the medium}$

Hydrogen atoms have a diameter of $1 \text{ \AA}$. Electrons accelerated to 300 kV have a λ of $0.022 \text{ \AA}$ allowing them to surpass the atomic diffraction limit. In comparison,

visible light, with a wavelength of 400 nm, can only resolve up to 244 nm assuming a NA of 1 (Kohl & Reimer, 2008).

2.3 Amplitudes, phases and Fourier slice theorem

Electrons have advantages over X-rays in that they can be focused using electromagnetic lenses, enabling recovery of amplitude and phase.

Fourier transform decomposes waves into a series of sinusoidal functions of fixed frequency. Diffraction angles from lenses are dependent on the frequency (resolution) of the incident wave, resulting in separation of wave frequencies in space; an objective lens therefore projects the Fourier transform of the 2D projection of the object onto the back focal plane (Figure 2.2.2b). The projector lens performs a reverse Fourier transform to project the image of the object onto the image plane. Electron microscopes are therefore considered double-Fourier systems (Figure 2.3.1) (Heel, 2014).

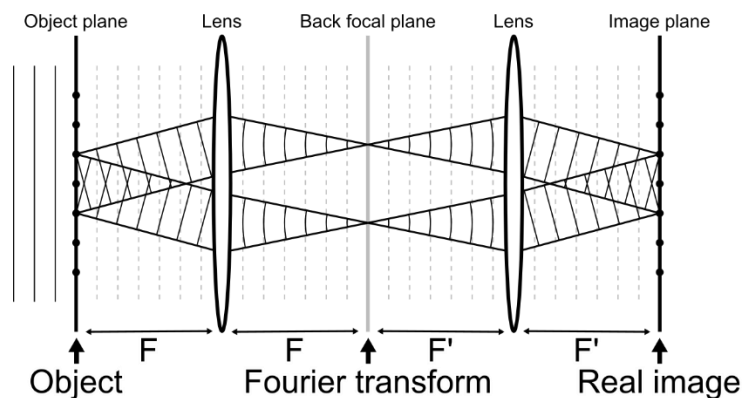


Figure 2.3.1: Schematic representation of a double Fourier system.

Fourier slice theorem states that Fourier transforms of 2D projection images from a 3D specimen is equivalent to a 2D slice its 3D Fourier transform. 3D reconstruction

in cryo-EM involves assigning Euler angles (α , β , γ corresponding to z, new y and the final z-axis) and in-plane XY translations (Scheres, 2012). Combining the signal from multiple views and calculating an inverse Fourier transform enables the reconstruction of a 3D map (Figure 2.1.1) (Glaeser, Nogales & Chiu, 2021).

In ideal circumstances, amplitudes are preserved at every spatial frequency, however due to detector limitations, lens aberrations and electron source imperfections, the output is a convolution between the actual image and contrast transfer function that is discussed further in section 0 (Erickson & Smith, 1988).

2.4 Electron microscope anatomy

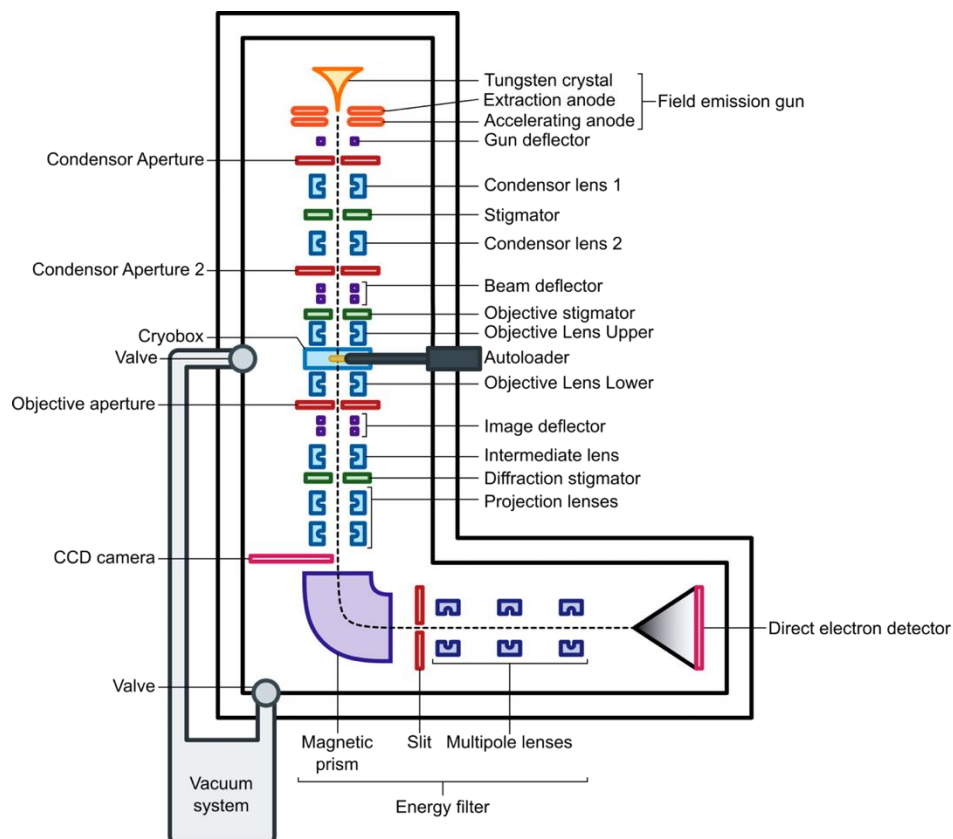


Figure 2.4.1: Anatomy of the cryo-electron electron microscope

A schematic of an electron microscope is shown in Figure 2.4.1. Electron microscopes all consist of a vacuum pump, electron source, condenser lens, sample stage, objective lens and projector lens, energy filter and a detector.

2.4.1 Vacuum source

Electrons have a high scattering cross-section, meaning they interact and scatter in air; consequently, microscopes must be kept under high vacuum (10^{-8} Pa) for imaging to occur and to ensure the longevity of the electron source. This is achieved by ion getter and turbo-molecular pumps that expel air and trap air from the column, stage and gun chambers. Turbomolecular pumps are physical fans, containing a series of blades spinning at 20,000-50,000 rpm to push air and water vapour out.

Ion getter pumps are non-physical and use a strong electrical current to ionize air molecules, which are drawn into the ion pump chamber and stick to the cathode surface. The chamber contains titanium electrodes, that sputter titanium atoms upon application of high voltage. The titanium atoms trap the air molecules and enable adsorption into the anode surface. The trapped air is then released upon shutdown and warming up of the microscope during cryo-cycle (Williams & Carter, 2009).

2.4.2 Electron sources and coherence

The electron source is the source of electrons that are focused by lenses to image the sample. The electron source should produce spatially and temporally coherent electron waves. Spatial coherence refers to the correlation between waves at

different points in space. Spatial coherence is inversely proportional to the area electrons are able to emanate from, therefore high spatial coherence relies on electrons emanating the smallest possible point. Temporal coherence, also known as monochromaticity, describes the difference in wavelength between electron waves.

High resolution microscopes use cold field emission guns. Field emission guns contain an extraction and acceleration anode that utilise high voltage electrical fields. The extraction anode is located close to the <100 nm tungsten crystal tip and applies a 300 kV voltage to draw electrons out. The electrons are then focused and accelerated by the accelerator anode and gun deflectors before entering the column (Crewe et al., 1968; Glaeser, Nogales & Chiu, 2021).

2.4.3 Electron lenses

The lens system in an electron microscope focuses the beam and is responsible for determining the magnification. Electron lenses are electromagnets that focus the electron trajectory by adjusting the current, therefore modulating the strength of the magnetic field. Electron trajectories are determined by the Lorentz force law, which states that the force (F) applied to an electron (q_e) travelling through an electric field (E) is perpendicular to the magnetic (B) field and velocity (v).

$$\mathbf{F} = -q_e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

This results in electrons following a spiral-like trajectory through the lens as the beam is focused (Glaeser, Nogales & Chiu, 2021).

The lenses are interspersed with apertures, stigmators and deflectors. Apertures remove electrons that have a scattering angle greater than a specified threshold to improve image contrast. Stigmators are weaker electromagnets that reduce astigmatism by applying a small magnetic field to correct for imperfections in the electron lens. Deflectors centre the electron beam and reduce spherical aberration, which is discussed in more detail in section 2.5.2 (Glaeser, Nogales & Chiu, 2021).

2.4.4 Cryobox and sample holders

Inserting samples into the grid introduces environmental moisture into the column. To prevent ice crystals forming on the sample, a LN₂ cooled cryobox encases the sample and prevents melting of the sample during data collection and acts as an ice deterrent, with ice sticking to the cryo-box rather than the sample (Egerton, 2016).

The sample holder can be manual side entry or autoloader. In both cases, the specimen is on top of a stage, with an optical tilt axis and motors to translate the sample in the z axis to find eucentric height, which is defined as the position where the electron beam is coincident with the optical axis, such that tilting the stage does not translate the observed area in x or y (Egerton, 2016; Elmlund, Le & Elmlund, 2017).

2.4.5 Energy filters

Following the sample chamber, the scattered electrons travel through an energy filter that removes inelastically scattered electrons.

Energy filters function as a prism that bends the electron trajectory through a bend in the column. After the bend, there is an energy slit that will only allow a narrow spectrum of electrons through to hit the detector. Inelastically scattered electrons are bent more due to their lower energy and therefore cannot pass through the energy slit, and do not hit the detector (Elmlund, Le & Elmlund, 2017).

2.4.6 Detectors and Direct Quantum Efficiency (DQE)

Detectors are a 2D array of pixels that measures the energy deposited onto pixels when hit by an incident electron. An integrated circuit then converts electron impact events into positional signals that can be interpreted by computers to produce an image (Faruqi & McMullan, 2018). Modern microscopes use direct electron detectors (DEDs), that consist of a back-thinned CMOS chip that directly converts electron impacts into a voltage signal for each pixel. This not only improves signal-to-noise ratio (SNR) relative to film and charged coupled devices (CCDs), but also vastly improves readout (Figure 2.4.2). The improved SNR from DEDs results in a vastly improved signal per electron exposure, which is also referred to as direct quantum efficiency (DQE). DQE is measured as a fraction of Nyquist frequency (section 2.7.1) (Li et al., 2013).

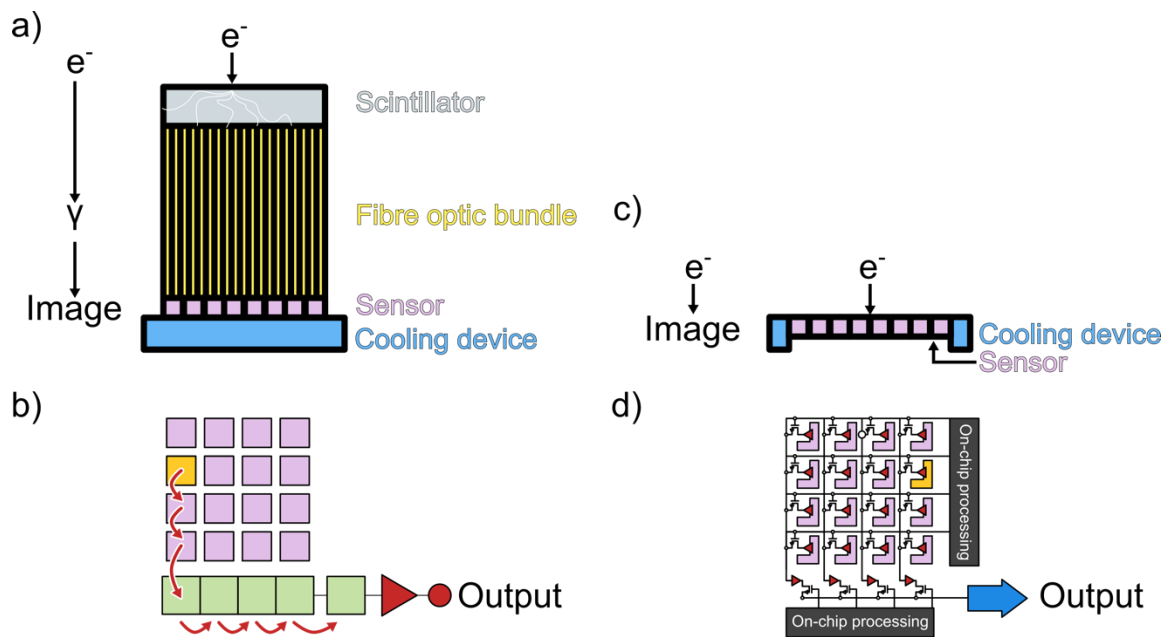


Figure 2.4.2: a) Overall structure of CCD chip. Incoming electrons must first enter a scintillator, which causes the electrons to scatter and emit photons. Often, spatial information is degraded due to backscattering and multiple photon emissions per electron scattering event. Photons then travel down the fibre optic bundle to hit the sensor. b) In the CCD sensor, each photon is counted and summed after a specified exposure time and amplified to produce the image. c) In DEDs, the incoming electrons hit the sensor directly. d) Each pixel in the sensor converts the electron to charge directly, and on-chip processing enables interpolation of electron position during the exposure.

The faster readout enables a "counting mode" which involves splitting exposures into dose-fractionated movie frames, which enables tracking of beam-induced motion. The readout also vastly improves acquisition time, allowing for more particles to be collected in shorter time frames, using lower dose, with higher weighting on high spatial frequency information captured during earlier frames with less damage, and later frames used to capture beam induced motion (BIM) and improve contrast. Furthermore, the integrated circuit in the CMOS chip can also further interpolate the position of electron impact events based on the strength of

signal from adjacent pixels, to effectively double the resolution by splitting each pixel into 2×2 subpixels. This is called "super-resolution" mode (Li et al., 2013).

2.5 Image formation

2.5.1 Scattering and contrast

When electrons pass through the specimen, three events can occur: transmission without scattering, inelastic scattering and elastic scattering.

The proportion of elastic to inelastic scattering depends on sample thickness and acceleration voltage. In general, inelastic scattering occurs 3 times more often than elastic scattering; for a 50 nm specimen imaging at 300 kV, ~80% of electrons are transmitted without scattering. Another 15% undergo inelastic scattering, and 5% undergo elastic scattering (Kohl & Reimer, 2008).

Elastic scattering occurs when the electron approaches the nucleus and are redirected with the same energy (frequency) but different direction.

Inelastic scattering occurs when the electrons deposit energy and displace electrons in the sample, causing secondary electrons. C-C bonds in proteins have a dissociation energy of 4 eV, and each inelastic scattering event deposits ~20 eV energy, which breaks ~5 covalent bonds. This results in radiation damage, which contributes to lowering the SNR.

Due to the 1:3 elastic:inelastic scattering ratio, every elastic scattering event coincides with ~15 covalent bonds being broken. Consequently, electron dose and exposure times are optimised to reduce damage (Glaeser, Nogales & Chiu, 2021).

There are 2 types of contrast: amplitude contrast and phase contrast. Phase contrast occurs due to elastic scattering, as the scattering angle is proportional to the resolution, resulting in a phase shift, and altered constructive interference. This causes enhanced contrast relative to the background.

However, due to the chemical properties of biological molecules, cryo-EM samples are weak phase objects and have negligible phase contrast when at focus. Elastic scattering results in a phase shift, however, the resultant wave does not have a significantly increased intensity compared to the unscattered waves, meaning that the foreground is often difficult to distinguish from the background. Consequently, images are taken at -0.5 to -3.5 μm defocus to improve contrast. Volta-phase plates can be used to apply an additional $-\pi/2$ phase shift, which significantly enhances phase contrast. However, for most >200 kDa samples, volta-phase plates are seldom used due to the inconsistencies in manufacturing quality, electron transparency and longevity, as well as difficulties maintaining optimal alignment (von Loeffelholz & Klaholz, 2021).

Amplitude contrast occurs due to inelastic scattering, as the inelastically scattered electrons are lower energy and subsequently blocked from hitting the detector by apertures and energy filters. Inelastic scattering occurs more often in higher density molecules such as heavy metals, causing fewer electrons from more electron dense biomolecules to reach the detector, resulting in a darker appearance relative to the background. Biological molecules, being made mostly of organic matter, are

considered phase objects with very minimal amplitude contrast (Elmlund, Le & Elmlund, 2017).

2.5.2 Spherical and Chromatic aberration

Spherical aberration (C_s) arises due to imperfections in the magnetic field and incoherence in the electron source (Figure 2.5.2a). This results in increased refraction of electrons near the edge of the lens, resulting in overfocussing of electron waves further away from the optical axis, resulting spreading of points. Spherical aberration is somewhat beneficial, as it adds phase contrast. In ideal circumstances, it is possible to tune defocus at a given C_s to achieve optimum contrast across spatial frequencies of 4-20 Å, which is also known as Scherzer focus.

Chromatic aberration arises from electrons of different wavelengths being focussed to different focal points. The longer the wavelength, the greater the electron wave is overfocussed (Figure 2.5.2b) (Egerton, 2016).

2.5.3 Contrast transfer function estimation and correction

$$CTF = -\sin(-\pi\Delta f\lambda k^2 + \frac{\pi}{2}C_s\lambda^3 k^4)$$

Where Δf = defocus

k = spatial frequency

C_s = spherical aberration

The equation for the contrast transfer function is shown above. The contrast transfer function is a description of how path length differences due to scattering, imperfections from lens aberrations, gun coherence, defocus and alignment result in contrast inversion at different spatial frequencies following a sinusoidal curve (Figure 2.5.1).

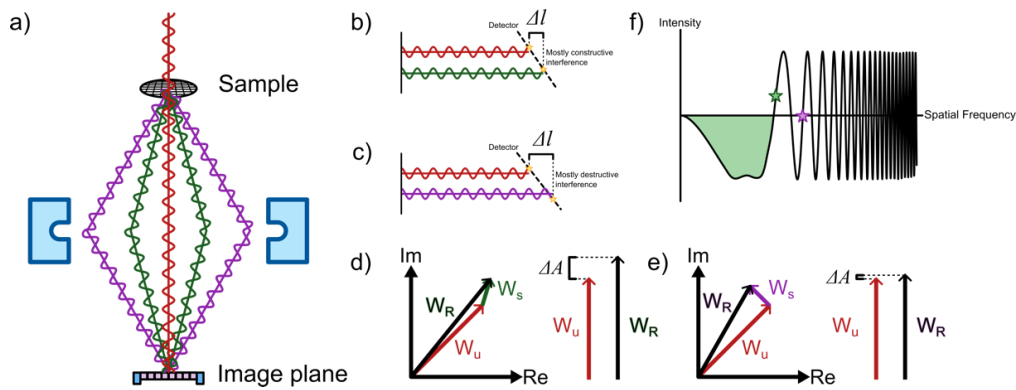


Figure 2.5.1: a) Schematic of scattering in a microscope. Unscattered beam shown in red. Higher spatial frequencies (purple) scatter at higher angles. b-c) Scattering results in a difference in path length Δl , that is dependent on spatial frequencies, which results in the resultant wave being imaged by the detector having different difference resultant intensities due to wave interference as it is focused into the detector. Star indicates position along the wave at point of impact. d-e) Argand diagrams showing that the difference in intensity ΔA between the unscattered background (W_u red) and resultant foreground (W_R black) creates the contrast in the image. f) Plotting the resultant intensities vs spatial frequency results in the contrast transfer function. Area shaded green represents the spatial frequency range that is unaffected by contrast inversion.

The contrast transfer function is best visualised by Thon rings observed in the power spectra ($[CTF]^2$) of micrographs (Figure 2.5.2d). The CTF is enveloped resulting in lower intensity at higher spatial frequencies, arising from chromatic and spherical aberrations, specimen charging, sample thickness and image drift. Consequently, microscopes must be carefully calibrated, and Thon rings of test micrographs must be inspected to ensure high resolution information is preserved (Egerton, 2016).

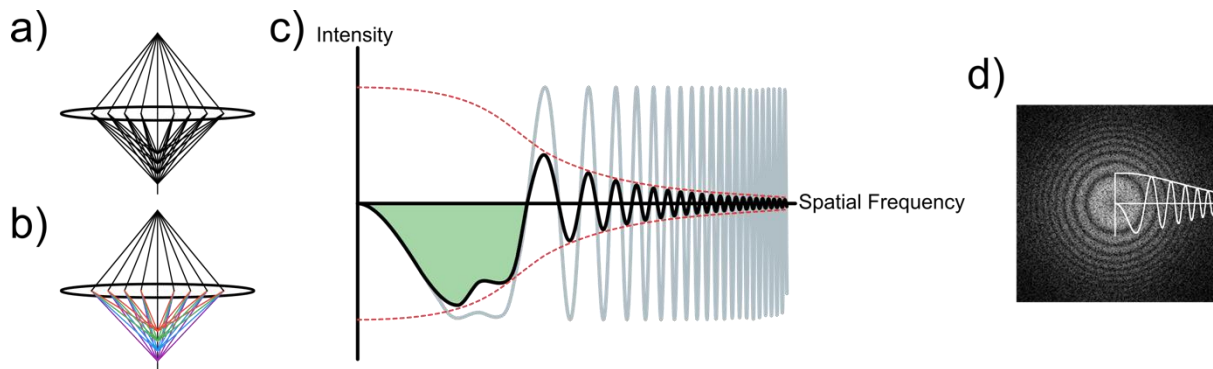


Figure 2.5.2: a) Spherical and b) chromatic aberrations create an envelope that dampens the CTF (c). d)

Thon rings observed in the power spectra of a micrograph, with the CTF inset. CTF image taken from Orlova & Saibil (2011).

To obtain the aberration-corrected image, a range of defocus values are used and the non-zero contrast intensities are corrected by dividing the Fourier transform of the image by the CTF, and performing an inverse Fourier transform to obtain the corrected image. Alternatively, the negative phase information can be flipped to produce a positive phases. The zero contrast spatial frequencies are then filled by the different defocus values (Orlova & Saibil, 2011).

2.6 Cryo-EM sample preparation

Samples in cryo-EM are deposited onto copper or gold grids and plunge frozen to preserve the frozen hydrated state and protect the sample from the vacuum and radiation damage. Liquid ethane or ethane-propane mixtures are used to vitrify the sample, ensuring that cooling can be achieved rapidly, without the Leidenfrost effect impeding thermal conductivity (Russo, Scotcher & Kyte, 2016).

The aim of cryo-EM sample preparation is to create an even distribution of homogeneous particles suspended in the middle of a 10-100 nm thick vitreous ice layer at a variety of particle orientations (Carragher et al., 2019).

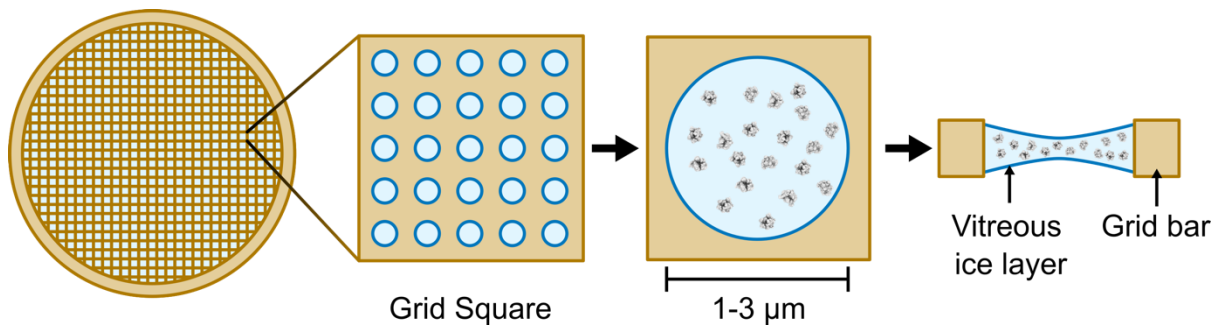


Figure 2.6.1: Schematic diagram of a holey grid

Grid preparation involves a fine optimisation between ice thickness and particle orientations, due to the air water interface and the sample affinity for the grid bars and support. There are 5 parameters that need to be optimised during grid preparation: grid type, grid pre-treatment, blotting parameters, buffer conditions and sample concentration (Weissenberger, Henderikx & Peters, 2021).

Grids are commonly made of copper or gold due to their conductive properties.

Grids are made up of squares, which are coated with perforated carbon or gold film where the sample is suspended. There are 2 main types of perforation: holey and lacey. Holey grids consist of evenly spaced holes of uniform size (Figure 2.6.1), whereas lacey grids are an interconnected network that have a variety of different hole shapes and sizes. Gold grids have the added advantage of greater conductivity, resulting in less BIM and radiation damage. The most common grid types are C-flat and Quantifoil.

Grids usually need to be rendered hydrophilic in order for biomolecules to enter the hole. This includes glow discharging, plasma cleaning, ethyl acetate washing or graphene oxide coating. 3-4 μl of sample is applied to the grid in a temperature and humidity controlled environment, after which filter paper is used to blot sample away to achieve the desired thickness. This can be done manually or done with an automated machine such as a Vitrobot. Blotting parameters are split into 4 key events. Wait time is the time between sample application and blotting where the sample is allowed to disperse on the grid. Blot time is the wicking duration where the filter paper contacts the grid surface. Blot force is the force applied to the grid during wicking.

Drain time is the delay between blotting and plunging, which is usually left at 0 to reduce exposure to the air-water interface. Additional parameters include humidity (which is usually kept at 95%) and temperature.

Buffer conditions affect blotting conditions, particle distribution, image contrast and preferential orientation. Commonly used additives for cryo-EM include detergents which are essential in solubilizing membrane proteins, as well as protecting biomolecules from the air-water interface.

Sample concentration dictates the number of particles per movie and is optimized to ensure maximum number of particles without aggregation or loss of contrast (Weissenberger, Henderikx & Peters, 2021).

2.7 Data collection

2.7.1 Pixel size and Nyquist frequency

Electron detectors convert analogue signal into discrete (digital) values. The Nyquist sampling theorem states that the sampling rate must be at least twice the highest frequency of the analogue wave to avoid aliasing. Aliasing is misidentification of analogue signal. The sampling rate is the inverse of the pixel size and is proportional to the magnification. For example, if the detector is set such that each pixel spans 2 Å (a pixel size of 2 Å/pixel) then the Nyquist frequency is $1/4 \text{ Å}^{-1}$, and the maximum sampling resolution is 4 Å. The maximum sampling resolution for a given pixel size is commonly referred to as Nyquist limit. Although it is possible to surpass the Nyquist limit using super-resolution mode, for most datasets, resolution is limited by sample homogeneity and ice quality rather than pixel size. Consequently, movies collected in super resolution are often binned to improve SNR and accelerate data acquisition and processing times.

Magnification affects both pixel size and the number of particles per movie. Given an unlimited timeframe and ideal sample, it would be optimum to use as high magnification as possible to get a better DQE, however given the limitations on microscope time, computing resources, and data storage, a compromise must be met between magnification and field of view (Glaeser, Nogales & Chiu, 2021).

2.7.2 Nominal Defocus

Movies must be collected at a range of defocus values to improve contrast at the lower spatial frequencies to separate the particles from the background. However, too high defocus results in loss of high spatial frequency information as the CTF crosses the X axis (0 amplitude), therefore a compromise must be met between defocus and desired resolution. Usually, most datasets use a defocus range between -0.5 to -3.5 μm . Alternatively, microscopes equipped with volta phase plate can collect data with minimal (usually -0.5 μm) defocus (Glaeser, Nogales & Chiu, 2021).

2.7.3 Data acquisition rate and exposure time

During data acquisition, the user must identify and select areas of the grid with good ice thickness (based on observing thin rings on the power spectra) and particle concentration. Exposure time for data acquisition must also be optimised to balance radiation damage with image contrast and available microscope time. With the advent of improved detectors with the ability to collect movies, longer exposure times are permitted as the user can remove the over-exposed final frames.

Furthermore, the surface of grids have slight deformations resulting in different z heights, meaning that the eucentric height and focus must be calibrated in each area. However, due to radiation damage, it is important to limit pre-exposure of the sample before acquiring high-magnification movie. Data collection is performed using a "low dose" approach, with search, focus and acquisition modes, each using different magnification and exposure times.

Initially, a series of $\sim 100\times$ magnification, low resolution images are collected and stitched together to produce an atlas, which shows the overall ice distribution and grid quality for each grid square (Figure 2.6.1). From there, users select squares showing good ice, followed by acquiring $\sim 10,000\times$ magnification images and finding eucentric height and focus whilst centred on the carbon support to avoid exposing the holes. Finally, holes are selected for final acquisition, which can range from $60,000$ - $150,000\times$ magnification depending on the expected size of the particle and desired pixel size. This magnification is subsequently used during autofocus. The exposure time is adjusted so the total electron exposure is kept within 20 - $100\text{ e}^-/\text{\AA}^2$, depending on ice thickness and type of grid support (Glaeser, Nogales & Chiu, 2021). The electron dose must be carefully optimised, too low dose results in poor contrast and insufficient signal for motion correction. Conversely, too high electron doses results in radiation damage, resulting in loss of high resolution information.

2.8 Image Processing

2.8.1 Motion Correction

There are two types of BIM during electron exposure, global and local. During global motion, the ice exposed to the beam undergoes a bowing movement. This motion can be exacerbated by stage drift which can be caused by vibrations in the environment, causing all particles to undergo a concerted movement. Conversely, local motion occurs at a particle level, with each particle being hit by different electrons and therefore following different trajectories.

MotionCorr is a motion correction program that takes each frame of the movie and calculates a pairwise cross correlation. It then compares the cross correlation between each frame to determine a vector describing the global shift during an exposure. This was later extended to local motion by dividing the field of view into $n \times n$ patches and calculating motion vectors for each patch (Zheng et al., 2017).

Further motion correction can be achieved using particle polishing, which is a post-refinement procedure that uses the 3D reference to track movement of individual particles. Particle polishing uses a running average of frames and gets rid of beam induced damage to improve the contrast of images and create trajectories for each individual particle, greatly improving per-particle CTF (Zivanov et al., 2018).

2.8.2 Particle picking

Particle picking is the stage after CTF correction, where particle centres are identified in summed movies and outputted into coordinate files. The corresponding coordinates are then used along with a specified box size during extraction to

obtain particle stacks used for later refinement and classification steps. For completely novel structures, particles are often initially manually selected for each micrograph by the user. However, since the advent of direct electron detectors, an increasing amount of cryo-EM density maps are available as initial references in the EMDB, as well as improvements in supervised learning algorithms, the vast majority of particles are picked using reference or reference free automated approaches.

Reference based approaches take 2D projection images of a reference (usually a previously determined region of the biomolecule, or a homolog) and performs a sliding window cross-correlation between the segments of a low-pass filtered micrograph with the reference. Segments with a cross-correlation above a specified threshold are selected for extraction (Wong et al., 2004).

Reference-free approaches include using gaussian blobs of specified radius or using manually labelled subset of micrographs to train a convolutional neural network to pick particles. The latter has been implemented recently in numerous different programs, including Topaz, crYOLO and cryoSPARC deep picker (Bepler et al., 2019a; Punjani et al., 2017; Wagner et al., 2019). In each of these methods, a per-pixel probability distribution is calculated, where each pixel is scored on whether it contains the particle centre.

For TOPAZ and cryoSPARC deep picker, micrographs are positively labelled with the coordinates specified given a high probability of being the particle centre. Only particles are picked, with the unpicked regions classified as unlabelled. In crYOLO, particles and non-particles (i.e., background ice and contaminant) must be

separately labelled and used to inform the weighting of the probability distribution. Training of the neural network involves performing per-pixel probabilities for the subset of micrographs that have been pre-labelled. TOPAZ uses a non-maximum suppression algorithm to select high-scoring pixels and remove surrounding pixels from contention as particle centres. The initial probabilities are then compared to the labelled probabilities, and the weights of the probability distribution are iteratively improved until the difference between the neural network-assigned coordinates and the pre-labelled coordinates have converged to a minimum (Bepler et al., 2019b; Wu et al., 2022).

2.8.3 Alignment

Following particle extraction, particle images are aligned and averaged in Fourier space to average out noise and enhance real signal from features in the biomolecules. Following the Fourier slice theorem, the Fourier transform of particle images can be "back-projected" to fill in Fourier space, and a reverse Fourier transform is used to reconstruct the final 3D volume (Figure 2.1.1). However, to average particles, Euler angles and in-plane translations must be determined on noisy images. In order to accurately assign angles, a maximum likelihood approach is used (Scheres, 2012).

2.8.4 Maximum likelihood approaches for refinement

Maximum likelihood approach is used in both refinement and classification to determine angular assignment. Each particle projection undergoes a discrete angular search against a reference projection based on an angular sampling parameter, and the cross-correlation for each angle is calculated to determine the

most accurate orientation. Instead of assigning a single Euler angle based on maximum cross-correlation, each Euler angle for each particle image is assigned a probability distribution for each angle based on the cross correlation.

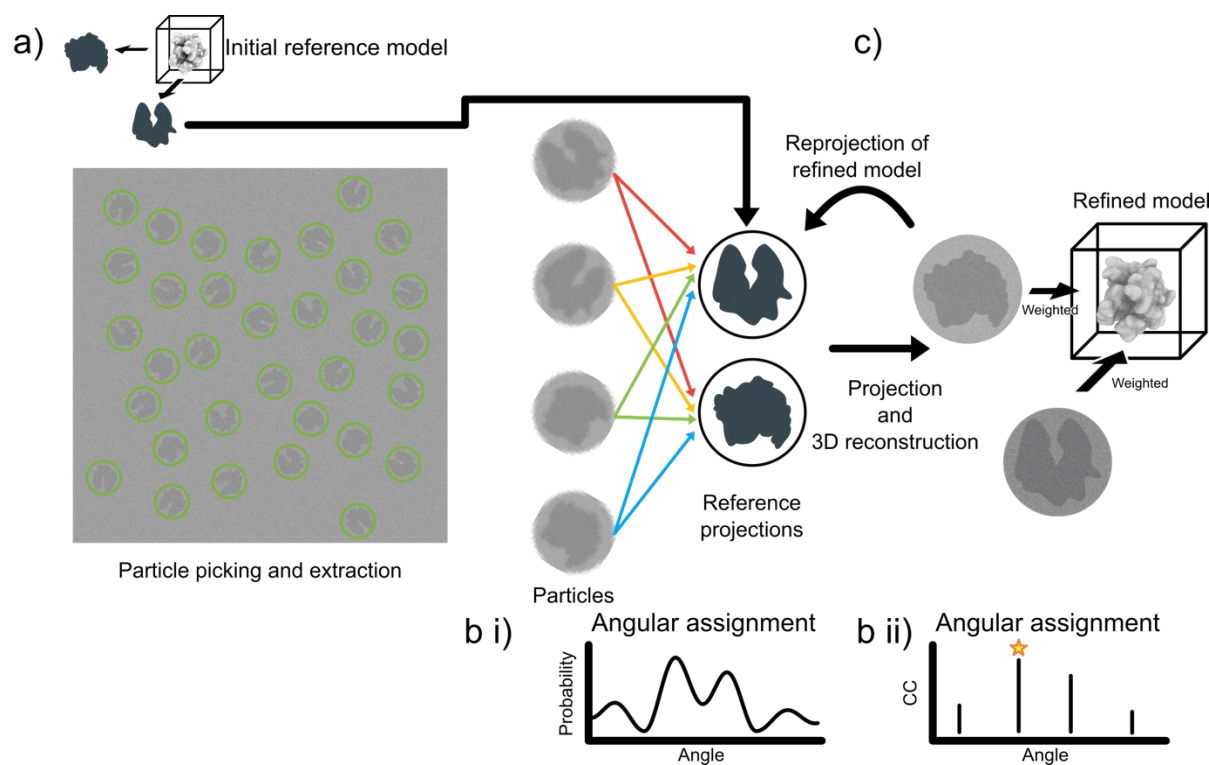


Figure 2.8.1: a) Single particle refinement involves aligning a set of noisy extracted particles with projection images from a reference. b i) Angular assignments are determined by calculating the probability of angles being correctly assigned. b ii) An alternative method of alignment is the maximum cross-correlation (CC) approach, which is suboptimal for noisy images. c) The aligned particles are then back-projected onto a low-pass filtered reference to produce the refined model, which can then be re-projected to produce improved reference projections, and the cycle is iterated until the accuracy in angular assignments converge.

The orientation and probabilities assigned to each particle projection are then used to create a new reconstruction and the process is repeated until accuracy of the angular assignment converges (Figure 2.8.1) (Scheres, 2012).

2.8.5 2D and 3D classification

2D and 3D classification enables identification of particles with different conformations or configurations and allows for discovery of new conformational states and removal of contaminants, which is one of the advantages of cryo-EM compared to X-ray crystallography. 2D classification is also used to determine whether there is any preferential orientation. 2D and 3D classification employs K-means clustering based approach to classify particle images into K number of groups as defined by the user. At first K particles are randomly selected to represent each class and are used as a reference. Each particle is then aligned to the reference particles and assigned to one of the classes. New class averages are then calculated using the maximum likelihood based refinement approach and these averages are then used as new representative references and the process is repeated for each round of classification. To avoid a phenomenon known as group collapse, which is often caused by preferential orientation or a major conformation dominating the dataset, a similar maximum likelihood approach is used, as well as running repeated rounds of classification to identify particles that are reliably assigned to particular classes at the same angular assignment values. 2D classification is also significantly sped up using stochastic gradient descent, where the orientation search angles are perturbed at random to ensure the entire orientational landscape is considered (Punjani et al., 2017; Yang et al., 2012).

2D and 3D classification can be performed without alignment and with masking. Masking enables classification to be focused on a particular region of interest within the refined class; by not performing alignment, the user can probe conformational changes within the sample that may have otherwise been averaged out. This is

further developed using particle subtraction and recentring, as well as multibody refinement, which subtracts the signal from the volume outside of the mask and recentres the particle coordinates onto the masked region, further improving accuracy of angular assignment for the masked region (Scheres, 2016).

More recent approaches using variational autoencoders-based neural networks are currently being developed to circumvent the need for specifying the number of classes. In most cases, the Euler angles from a previous consensus refinement are used to determine the initial "pose". Each particle is then fed into an encoder neural network that converts the information from each particle into a lower dimensional latent space using a dimensionality reduction algorithm, such as uniform manifold approximation or principal component analysis. The latent space can then be visualized on a 2D graph. A decoder neural network then uses the poses to convert the particles represented in latent space back into a 3D volume. The 3D volume produced is then compared to the initial consensus refinement and the weights for the encoder and decoder network are iteratively adjusted to minimise a loss function (Donnat et al., 2022).

2.8.6 Fourier shell correlation

To determine the resolution of the final 3D volume, the dataset is split into 2 random halves. Two independent 3D reconstructions are refined from two halves of the data and the Fourier transform of both "half-maps" are correlated at each resolution shell to determine the Fourier shell correlation (FSC). The FSC at each spatial frequency is then plotted, and the spatial frequency at a 0.143 cutoff is the given resolution estimate for the 3D reconstruction (Figure 2.8.2) (Rosenthal & Henderson, 2003).

This is referred to as the gold-standard FSC, as each half-map reconstruction is generated independently of the other.

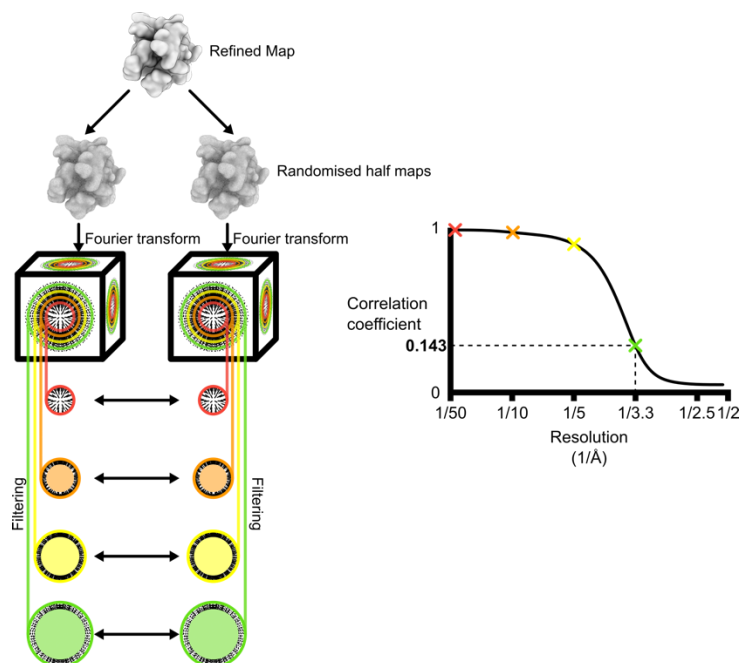


Figure 2.8.2: Fourier shell correlation measures the cross-correlation of the Fourier components of 2 independent halves of the refinement map. The correlation between each spatial frequency is then plotted and the nominal resolution is determined by finding the resolution at the correlation cutoff of 0.143.

The map can also be locally sharpened to further improve the visibility of high-resolution features by providing a solvent mask. Local sharpening utilizes a 3D sliding mask across the volume to calculate the FSC in the masked region, and then bandpass filtering the Fourier components of the masked volume within the resolution range calculated from the FSC curve to enhance the high-resolution signal present within the data (Kucukelbir, Sigworth & Tagare, 2014).

2.8.7 Model building

Depending on the resolution, PHENIX can be used to automatically build and refine a model from a protein sequence provided the resolution of the regions better than

3.5 Å (Liebschner et al., 2019). However, lower resolution regions must be manually edited in COOT (Croll, 2018; Emsley et al., 2010). Models are verified using a Ramachandran plot and MOLPROBITY software to verify the model abides by the stereochemical properties of proteins and nucleic acids. For density with resolution lower than 5 Å, rigid body fitting is often used to fit a previously determined portion of the structure into the density. More recently, AlphaFold models have also been used to verify the identity of lower resolution density and as starting models for model building in lower resolution reconstructions (Glaeser, Nogales & Chiu, 2021; Jumper et al., 2021).

Chapter 3

Materials & Methods

All protein statistics and buffer recipes can be found in Table 3.1.1 and Table 3.1.3.

All buffers were made to final volumes with MilliQ water unless stated otherwise.

3.1 Plasmid transformation

Recombinant plasmids containing genes for *Escherichia coli* RNAP α , β , β' and ω core subunits, PspF₁₋₂₇₅ AAA+ domain and *Klebsiella pneumoniae* σ^{54} and σ^{54} (R336A mutant) are listed in Table 3.1.2.

All plasmids were transformed into *E. coli* strain BL21 (DE3) (NEB) by heat shock following manufacturers' guidelines and plated on lysogeny broth agar plates supplemented with the appropriate antibiotic to select for transformants (Table 3.1.2). Plates were incubated at 37 °C for ~16 hours overnight.

Table 3.1.1: Proteins used in this investigation and statistics

Component	Molecular Weight (kDa)	pI	E (M ⁻¹ cm ⁻¹)*	No. Residues
σ^{54}	53.93	4.60	42400	477
RNAP ($\alpha_2\beta\beta'\omega$ subunits)	388.98	5.38	212130	3498
PspF ₁₋₂₇₅	31.10	5.06	21430	275
RNAP- σ^{54}	442.89	5.23	254530	3975
*Assuming all cysteines are reduced. Measured in water at 280 nm. pI = isoelectric point. E = extinction coefficient				

Table 3.1.2: List of plasmids used in this study and their corresponding genes.

Genes	Protein	Tag	Plasmid Vector	Antibiotic used in plates and LB media	Antibiotic Concentration (μ g/ml)
<i>rpoA</i> , <i>rpoB</i> , <i>rpoC</i> *	α , β , β'	None	pGEMABC (Murakami, 2013)	Ampicillin	100
<i>rpoZ</i> *	ω	None	pACYCDuet (Novagen)	Chloramphenicol	100
<i>rpoN</i>	σ^{54}	6 $\times$ His Tag	pET28b (Novagen)	Kanamycin	50
<i>pspF</i>	PspF ₁₋₂₇₅ (AAA+ domain)	6 $\times$ His Tag	pET28b (Novagen)	Kanamycin	50
* <i>rpoA</i> , <i>rpoB</i> , <i>rpoC</i> and <i>rpoZ</i> were coexpressed. <i>E. coli</i> BL21(DE3) were transformed with both pGEMABC (100 ng) and pACYCDuet (100 ng)					

Table 3.1.3: Buffer recipes

Buffers/Solutions	Composition
σ^{54} NiA	20 mM Tris-HCl pH 8, 500 mM NaCl, 5% v/v Glycerol, 10 mM Imidazole
PspF ₁₋₂₇₅ NiA	50 mM NaH ₂ PO ₄ , 300 mM NaCl, 5% v/v glycerol
σ^{54} NiB	70 mM NaCl, 500 mM Imidazole pH 8, 5% v/v Glycerol
PspF ₁₋₂₇₅ NiB	50mM NaH ₂ PO ₄ , 50 mM NaCl, 500 mM imidazole pH 8, 5% glycerol
HepA	20 mM Tris-HCl pH 8, 50 mM NaCl, 5% v/v Glycerol, 1 mM TCEP*
HepB	20 mM Tris-HCl pH 8, 1 M NaCl, 5% v/v Glycerol, 1 mM TCEP*
GF	20 mM Tris-HCl pH 8, 150 mM NaCl, 5% v/v Glycerol, 2 mM TCEP*
GF _{PspF}	10 mM Tris-HCl pH 8, 50 mM NaCl, 5% v/v Glycerol, 0.1 mM EDTA, 1 mM TCEP*
TGET	10 mM Tris-HCl (pH 8), 5% v/v Glycerol, 1 mM EDTA, 2 mM TCEP*, NaCl [♦]
Lysis Buffer	50 mM Tris-HCl pH 8, 1 mM EDTA, 2 mM TCEP*, 1 cOmplete™ protease inhibitor tablet per 50 ml*
STA Buffer	25 mM Tris-Acetate, 8 mM Mg-Acetate, 10 mM KCl, 1 mM TCEP*
2× Annealing buffer	20 mM Tris-HCl, 100 mM NaCl
Buffer EM1	20 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl ₂ , 1 mM TCEP
Buffer EM2	20 mM Tris-HCl, 150 mM KCl, 5 mM MgCl ₂ , 5 mM TCEP
Native-PAGE Binding buffer (5×)	25% v/v Glycerol, 750 mM KCl, 100 mM Tris-HCl pH 8, 10 mM MgCl ₂
* Added directly before use	
♦TGET was prepared with 100 mM, 250 mM, 500 mM, 800 mM and 1 M NaCl	

3.2 Cell culture and induction

Single colonies were used to inoculate 6×1 L of LB medium supplemented with the appropriate antibiotic (Sigma-Aldrich) in 1 L glass Erlenmeyer flasks (Corning) for all transformants. Gene expression was regulated by the *lac* promoter; expression was induced using 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich).

3.2.1 σ^{54} and PspF₁₋₂₇₅ induction and expression

Cells were incubated at 25 °C, 160 rpm overnight (~16 hours). Cells were induced at an OD₆₀₀ of 0.6-0.8 and the temperature was lowered to 22 °C; cells were left for at least 4 hours before harvesting.

3.2.2 RNAP core ($\alpha_2\beta\beta'\omega$) induction and expression

Cells were incubated at 25°C, 160 rpm for ~24 hours to an OD₆₀₀ of 0.8 before induction. Following induction, cells were left for 4 hours at 25 °C, 160 rpm before harvesting.

3.2.3 Harvesting

Cells were harvested by centrifugation at 3635 g for 20 minutes at 4 °C, and pellets were washed with σ^{54} NiA buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 5% v/v glycerol, 10 mM imidazole), PspF NiA buffer (50 mM NaH₂PO₄, 300 mM NaCl, 5% v/v glycerol) or lysis buffer (50 mM Tris-HCl, pH 8, 1 mM EDTA, 2 mM Tris(2-carboxyethyl)phosphine (TCEP)) each supplemented with 1 cOmplete™ protease inhibitor tablet per 50 ml (Roche) buffer before spinning at 8000 g for 10 minutes. The cell pellets were then stored at –80 °C.

3.3 Cell lysis

Each of the cell pellets were resuspended on ice with 10 ml of σ^{54} NiA, PspF₁₋₂₇₅ NiA or lysis buffers per litre of cell culture and supplemented with 1 cOmplete protease inhibitor cocktail tablet (Roche).

Cells were lysed on ice by sonication using a Q500 sonicator (Qsonica) at 50% amplitude, 3 seconds on, 4 seconds off for 15 minutes before centrifugation at 38,500 g for 45 minutes. The supernatant was then filtered with a 5 µm syringe filter, followed by a 0.45 µm syringe filter (PALL Life Sciences) to remove any remaining debris.

3.4 Protein purification

All chromatography was performed on a ÄKTA Pure (GE Life Sciences) at 4 °C. Flow rates and method details for each chromatography run can be found in Supplementary Table 1. After each chromatography run, fractions from selected peaks were then ran on Bolt 4-12% Bis-Tris Plus SDS-PAGE gels alongside PageRuler Plus Prestained 10 to 250 kDa molecular weight ladder (ThermoFisher Scientific™). All SDS-PAGE gels were run at 200 V for 35 minutes before being stained with InstantBlue (Expedeon) and imaged with Gel Doc™ XR+ Gel Documentation System (BioRad).

3.4.1 σ^{54} wildtype and R336A mutant purification

σ^{54} was initially extracted from lysate by nickel affinity chromatography, with a 5 ml HisTrap™ HP column (Cytiva) equilibrated with σ^{54} NiA buffer and eluted with a 5-100% linear gradient of σ^{54} NiB buffer (70 mM NaCl, 500 mM imidazole pH 8, 5% v/v glycerol). Following SDS-PAGE analysis, selected fractions were pooled and passed through a HiPrep™ 26/10 desalting column (Cytiva) to exchange into HepA buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 5% v/v glycerol, 1 mM TCEP).

σ^{54} was then further purified by heparin affinity chromatography with a 5 ml HiTrap™ Heparin HP column (GE Life Sciences) equilibrated with HepA buffer and eluted with a 5-100% gradient of HepB buffer (20 mM Tris-HCl pH 8, 1 M NaCl, 5% v/v glycerol, 1 mM TCEP). Fractions were pooled based on SDS-PAGE analysis.

Pooled fractions were then gel filtrated using a Superdex Hiload® 75 16/600 column (GE Life Sciences) equilibrated with GF buffer (20 mM Tris-HCl pH 8, 150

mM NaCl, 5% v/v glycerol, 2 mM TCEP). Fractions were pooled based on SDS-PAGE analysis and concentrated using a 30 kDa molecular weight cut-off (MWCO) centrifugal filter (Merck Millipore). The concentration was measured using a Nanodrop ND-1000 spectrophotometer (Thermofisher Scientific™). Protein was then aliquoted into 20 µl volumes, flash frozen in liquid nitrogen and stored at -80 °C.

3.4.2 PspF₁₋₂₇₅ purification

All studies presented here used the AAA+ truncation of PspF (residues 1-275). PspF₁₋₂₇₅ was initially extracted from lysate by nickel affinity chromatography; clarified lysate was passed through a 5 ml HisTrap HP column (Cytiva) equilibrated with PspF NiA buffer and eluted with 5-100% PspF NiB buffer (50mM NaH₂PO₄, 50 mM NaCl, 500 mM imidazole pH 8, 5% glycerol). The peak fractions were pooled, concentrated to less than 5 ml using a centrifugal filter and dialysed in PBS for 1 hour at room temperature with 10 kDa MWCO SnakeSkin™ dialysis tubing (Thermofisher Scientific™).

The N-terminal 6×His tag was then cleaved by adding 90 µl (18 NIH units) thrombin. The dialysis tubing was transferred to a fresh beaker of PBS for 2 hours at room temperature. Complete cleavage was verified by taking periodic aliquots for SDS-PAGE analysis and running for 50 minutes. Any uncleaved PspF₁₋₂₇₅ was then removed by reverse nickel affinity chromatography, in which the flowthrough was collected for gel filtration.

Flow through fractions were gel filtrated using a Superdex Hiload® 75 16/600 column (GE Life Sciences) equilibrated with GF_{PspF} buffer (10 mM Tris-HCl pH 8, 50

mM NaCl, 5% v/v glycerol, 0.1 mM EDTA, 1 mM TCEP). Fractions were pooled based on SDS-PAGE analysis and concentrated using a 10 kDa molecular weight cut-off (MWCO) centrifugal filter (Merck Millipore). The concentration was measured using a Nanodrop ND-1000 spectrophotometer (ThermoFisher Scientific™). Protein was then aliquoted into 20 µl volumes, flash frozen in liquid nitrogen and stored at -80 °C.

3.4.3 RNAP core ($\alpha_2\beta\beta'\omega$)

RNA polymerase was initially extracted from cell lysate using polymin P precipitation. Polymin P is a positively charged polymer that co-precipitates DNA-binding proteins. 10% (v/v) Polymin P was added dropwise to cell lysate at 4 °C on a magnetic stirrer to a concentration of 0.7% (v/v). The solution was then left at 4 °C for 15 minutes to homogenise, followed by centrifugation at 18,000 $\times g$ at 4 °C for 15 minutes. Cell pellet was then washed with 50 ml TGET buffer (10 mM Tris-HCl pH 8, 5% v/v glycerol, 1 mM EDTA, 2 mM TCEP). Centrifugation and washing was repeated twice to remove contaminants before a final wash with TGET containing 1M NaCl to elute RNAP.

RNAP was precipitated from the eluant by adding 60% ammonium sulfate. RNAP precipitate was pelleted by centrifugation as before and resuspended in 30 ml TGET buffer containing 100 mM NaCl. The resuspended solution was then desalted using a HiPrep™ 26/10 desalting column (GE Life Sciences) equilibrated with 100 mM NaCl TGET buffer. Selected fractions and aliquots were collected at each stage and run on SDS-PAGE to confirm the presence of RNAP.

To further improve the purity of the RNAP, pooled fractions were then run on a BioRex 70 exchange column (Bio-Rad), equilibrated with 100 mM NaCl TGET buffer. RNAP was eluted with an increasing gradient of 1M NaCl TGET buffer. Selected peak fractions were run on an SDS-PAGE gel before being pooled and dialysed overnight in 100 mM NaCl TGET buffer.

Pure RNAP was then finally obtained by HiTrap Q anion exchange chromatography, equilibrated with 100 mM NaCl TGET buffer. RNAP was eluted with a increasing gradient of 1M NaCl TGET buffer. Selected peak fractions were run on an SDS-PAGE gel before being pooled and concentrated using an 100 kDa MWCO centrifugal filter (Merck Millipore). The concentration was determined using a Nanodrop ND-1000 spectrophotometer (Thermofisher Scientific™) to measure the OD₂₈₀. Protein was then aliquoted into 20 µl volumes, flash frozen in liquid nitrogen and stored at -80 °C.

3.4.4 RNAP- σ^{54} formation

30 µM RNAP was incubated with 120 µM σ^{54} at 4 °C for 1 hour, at a final volume of 250 µl. The mixture was then loaded on a Superose 6 10/300 Increase column (GE Life Sciences) equilibrated with GF buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM TCEP). Selected peak fractions were run on an SDS-PAGE gel. Fractions containing both σ^{54} and RNAP core subunits were pooled and concentrated using an 100 kDa MWCO centrifugal filter (Merck Millipore), supplemented with 10% glycerol and aliquoted into 20 µl volumes, flash frozen in liquid nitrogen and stored at -80 °C.

3.5 Complex formation assays

3.5.1 Nucleic acid scaffold preparation

DNA in all experiments used 63 bp DNA from positions -35 to +28 of *nifH* promoter. Single stranded DNA (IDT) was resuspended in 20 mM Tris-HCl buffer and annealed in equimolar amounts in 1× annealing buffer to give a concentration of 200 μM. The mixture was then heated to 95 °C for 5 minutes, and then cooled to 4 °C in 2 °C steps, with 1 minute/step, before being stored in -20 °C. Final concentrations of dsDNA was 100 μM. For DNA-RNA scaffolds, all DNA and RNA was mixed simultaneous and cooled as before, and the final concentration was 66.7 μM.

3.5.2 Native-PAGE

All native-PAGE assays were run using 4.5% TBE gels in 1× TBE buffer (Invitrogen) running buffer. The recipe for native-PAGE gels and binding buffer used as the loading dye is shown in Table 3.5.1. Gels were run at 100V, 4 °C for 100 minutes. Gels were stained for DNA with SYBR Safe (Thermo Fisher Scientific™), and protein with InstantBlue (Expedeon). Final reaction volumes were 20 μl.

To form RNAP-σ⁵⁴-DNA complexes, 250 nM RNAP was mixed with 350 nM DNA for 1 hour at 4 °C in STA buffer (25 mM Tris-Acetate, 8 mM Mg-Acetate, 10 mM KCl, 1 mM TCEP) , native-PAGE binding buffer (25% v/v glycerol, 750 mM KCl, 100 mM Tris-HCl pH 8, 10 mM MgCl₂) for 4°C for 1 hour. To form RNAP-σ⁵⁴-DNA-PspF₁₋₂₇₅ complexes, ADP·AlFx was formed *in situ* from ADP, AlCl₃ and NaF to trap PspF₁₋₂₇₅ in a hexameric state. The R_{Pc} complex was mixed with 1.75 μM PspF₁₋₂₇₅, 8 mM ADP and 10 mM NaF and incubated for 5 minutes at 37 °C. Following the

incubation, the mixture was transferred to ice, and 2 mM AlCl_3 was added, and then the mixture was incubated for another 12 minutes at 37 °C.

Table 3.5.1: Recipe for native-PAGE gels and loading buffers

Gel type	Recipe (5 gels)
Native-PAGE (4.5% TBE gel)	7.5 ml 30% Acrylamide (Sigma-Aldrich) 5 ml 10× TBE 37 ml H_2O (to 50 ml) 0.5 ml 10% ammonium persulfate 50 μl Tetramethylethylenediamine
Native-PAGE Binding buffer (5×)	25% v/v Glycerol 750 mM KCl 100 mM Tris-HCl pH 8 10 mM MgCl_2

3.6 Complex formation

3.6.1 RP(-12/-11), RP(-11/-8) and RP(-11/-3) complexes

The molar ratios used in gel filtration was identical to the molar ratios used in native-PAGE. 20 μM RNAP- σ^{54} was mixed with 28 μM DNA in Buffer EM1 (20 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl_2 , 1 mM TCEP) for 1 hour at 4 °C, followed by addition of 140 μM of PspF₁₋₂₇₅, 8 mM ADP and 10 mM NaF. The mixture was incubated at 37 °C for 5 minutes, followed by addition of 2 mM AlCl_3 , and another incubation at 37 °C for 12 minutes. The final volume was 250 μl . The mixture was then centrifuged at 8000 $\times g$ for 5 minutes at 4 °C before loading onto a Superose 6 10/300 Increase column (GE Life Sciences) equilibrated with Buffer EM1. Selected fractions were analysed by SDS-PAGE to confirm the presence of all protein components, and the fraction at $\text{OD}_{280} = 200$ mAU (~ 0.8 μM) on the left shoulder of the peak was used immediately for grid preparation.

3.6.2 RP(-10/-1) complexes

RP(-10/-1) complexes were prepared with identical molar ratios to the above RNAP- σ^{54} -DNA-PspF₁₋₂₇₅ complexes, except using a final volume of 80 μ l and without the gel filtration step. 11.5 μ M RNAP- σ^{54} was used, with the same ADP, NaF and AlCl₃ concentrations as before. Following formation of ADP.AIFx, samples were buffer exchanged into Buffer EM2 (20 mM Tris-HCl, 150 mM KCl, 5 mM MgCl₂, 5 mM TCEP) using a 0.5 ml Zeba™ 7K MWCO desalting column as per the manufacturers' protocol. 8 mM CHAPSO was then added immediately before grid preparation.

3.6.3 RP5nt, 6nt, 7nt, 8nt and 9nt complexes

17 μ M RNAP- σ^{54}_{R336A} was added to 18.7 μ M of DNA-RNA scaffold in the presence of buffer EM1 and incubated for 1 hour at 4 °C. Following incubation, samples were buffer exchanged into Buffer EM2 (20 mM Tris-HCl, 150 mM KCl, 5 mM MgCl₂, 5 mM TCEP) using a 0.5 ml Zeba™ 7K MWCO desalting column as per the manufacturers' protocol. 8 mM CHAPSO was then added immediately before grid preparation.

3.7 Cryo-EM sample preparation

All grids were prepared using a Vitrobot™ Mark IV (FEI) at 4 °C and 100 % humidity, using Grade 595 Vitrobot™ filter paper (Electron Microscopy Sciences). All grids were plunge frozen using liquid ethane and stored in liquid nitrogen.

3.7.1 Sample preparation of RP(-11/-8) and RP(-11/-3) complexes

4 μ l of the $\sim 0.8 \mu\text{M}$ RNAP- σ^{54} -DNA-PspF₁₋₂₇₅ complexes directly from gel filtration were deposited onto 300 mesh holey copper Quantifoil R2/2 grids, that were argon cleaned for 45 seconds (Lambda photometrics). The blotting parameters were as follows: wait time 30 seconds, blot time 2 seconds and blot force -6.

3.7.2 Sample preparation of RP(-10/-1), RP5nt, 6nt, 7nt, 8nt and 9nt complexes

4 μ l of complex was deposited onto 300 mesh holey gold C-flat R1.2/1.3 grids, that were plasma cleaned in air for 30 seconds (Harrick Plasma). The blotting parameters were as follows: wait time 30 seconds, blot time 2 seconds and blot force -8.

3.7.3 Data collection

All datasets were collected on a Titan Krios (ThermoFisher Scientific), operated at 300 kV, equipped with a K3 direct electron detector and a BioQuantum energy filter (Gatan). Movies were collected at a nominal magnification of 81,000 $\times$ and a slit width of 20 eV. Data collection was carried out using EPU software (FEI).

Sample	Pixel Size (Å/pixel)	Total Dose (e ⁻ /Å ²)	Defocus Range (μm)	Total Movies
RP(-11/-8)	1.1	50	-3 to -1	11 634
RP(-11/-3)	1.06	40	-3 to -1	24 393
RP(-10/-1)	1.1	40	-3 to -1	15 534
RP5nt	1.1	30	-3 to -1	12 309
RP6nt	1.06	40	-2.7 to -1	9188
RP7nt	1.06	30	-3 to -1	4185
RP8nt	1.072	40	-2.7 to -1	11 198
RP9nt	1.06	30	-3 to -1	6121

3.7.4 Image processing

All image processing steps were carried out in RELION 4.0 (Kimanius et al., 2021). Contrast transfer function estimation was carried out with CTFFIND 4 (Rohou & Grigorieff, 2015). Particles in the RP(-11/-8) and RP(-11/-3) datasets were done using Gautmatch (<https://github.com/JackZhang-Lab/Gautmatch>) using RPi reference projections (Glyde et al., 2017), whereas RP(-10/-1) and RP5-9nt datasets were done with Topaz (Bepler et al., 2019b).

3.7.5 Model Building and refinement

RNAP- σ^{54} -DNA-PspF₁₋₂₇₅ complex structures were all built by first rigid-body-fitting RPi (PDB ID: 5NSS) into the cryo-EM density, followed by iterative cycles of manual refinement of selected regions in COOT (Emsley et al., 2010) followed by real space refinement in PHENIX (Liebschner et al., 2019) and validation in MOLBPRIORITY (Williams et al., 2018) to improve the fit and model geometry. A similar approach was used for longer transcript complexes, using RPitc as the initial starting model (Glyde et al., 2018). Figures were made in UCSF ChimeraX 1.4 (Pettersen et al., 2021).

3.8 *In vitro* transcription assays

In vitro transcription assays were used to detect short UpGGG RNA transcripts containing α -³²P labelled GTP. These assays were performed by Milija Jovanovic, a collaborator in Martin Buck's lab. 100 nM of RNAP- σ^{54} was mixed with 10 nM of either WT (negative control), -10/-1 (positive control), -11/-8, -11/-5 and -11/-3 *nifH* DNA for 5 minutes at 37 °C. In all lanes, 0.5 mM UpG, 0.2 mCi/ml [α -³²P] GTP and 0.2 mg/ml heparin. For PspF₁₋₂₇₅ activated complexes, 5 μ M of PspF₁₋₂₇₅ was added,

along with either 4 mM dATP or 8 mM ADP, 10 mM NaF and 2 mM AlCl_3 .

Complexes were incubated for 10 minutes at 37 °C for transcription to occur, and then the reaction was quenched in denaturing loading buffer and run on a 20% denaturing gel and imaged on a Fuji FLA-5000 Phosphorimager.

Chapter 4

Structural basis of σ^{54} -mediated open complex formation

4.1 Introduction

Open complex formation is the process in which the DNA is melted to form a 12-15 bp transcription bubble and the template strand is loaded into the RNA polymerase (RNAP) active site. In σ^{54} -mediated transcription initiation, σ^{54} recruits RNAP to conserved -24 (GG) and -12 (GCA) regions of promoters upstream of the transcription start site (TSS, +1). σ^{54} recruits RNAP to promoters associated with stress responses to form a stable, transcription incompetent closed complex (RPc). A specialised class of AAA+ protein called bacterial enhancer binding proteins (bEBPs) are required to remodel the RPc to the transcription competent open complex (RPO). bEBPs are members of clade 6 group within the pre-sensor 1 hairpin superfamily (PS1BH) and contains two defining insertions, the L1 loop (also referred to as the $\alpha 2$ insert), which is characteristic of clade 6, and the L2 loop (often referred to as “pore loop”), which is shared by PS1BH members.

In the absence of stress stimuli, most bEBPs are in an inhibited dimeric state and are recruited to regions of DNA 100-150 bp upstream known as UAS sites. Upon activation by specific signals and with the help of DNA looping factors such as IHF,

bEBPs hexamerize and binds to σ^{54} region I to form an intermediate complex (RPi). However, the mechanism of action of bEBPs, specifically how ATP hydrolysis facilitates DNA melting, loading and alleviation of σ^{54} inhibition remains unknown (Figure 4.1.1) (Gao et al., 2020).

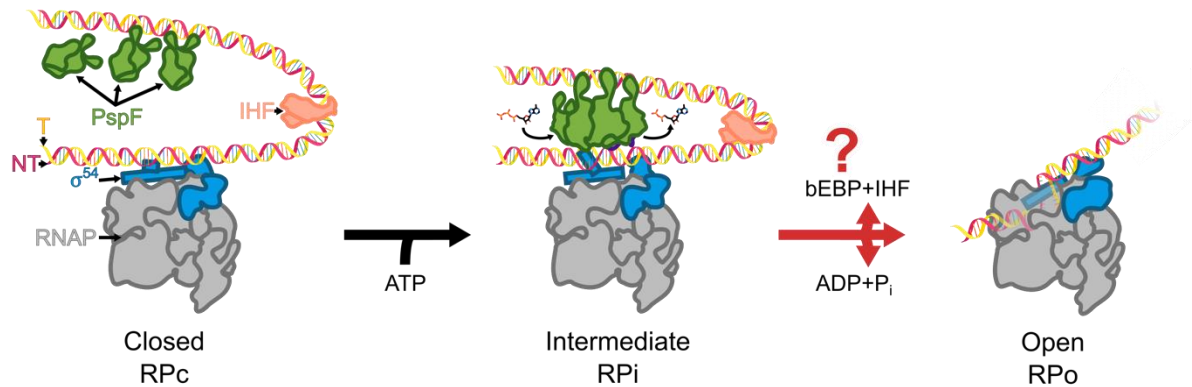


Figure 4.1.1: Schematic of σ^{54} mediated open complex formation.

4.2 High resolution structure of closed and PspF₁₋₂₇₅-engaged intermediate complexes

Recent work by Ye et al., has determined structures of the closed complex (RPC) and PspF₁₋₂₇₅-engaged intermediate complex (RPi(-12/-11)) states (Figure 4.2.1). Structure determination of the RPC and RPi(-12/-11) complexes was done by Fuzhou Ye, and I contributed to the model building, validation and performed the structural analysis.

For the closed complex, a fully duplexed *nifH* promoter DNA was used to determine the initial steps of σ^{54} binding and transcription bubble nucleation. For the RPi(-12/-11) structure, we used mismatched *nifH* promoter DNA (from -12 to -11, referred to as -12/-11) and ADP aluminium fluoride (ADP.AIFx) to trap PspF₁₋₂₇₅ in the

hexameric DNA-bound state. The -12/-11 mismatch was used due to its ~6-fold enhanced binding affinity to σ^{54} , and ADP.AIFx is a non-hydrolysable ATP-hydrolysis transition state analogue used to trap PspF₁₋₂₇₅ (Chaney et al., 2001).

These structures reveal the dual role of σ^{54} in preventing premature open complex formation as well as nucleating DNA melting, aided by PspF₁₋₂₇₅. These structures were determined to 3.4 and 3.5 Å nominal resolution for the RPc and RPi(-12/-11) complex respectively, enabling determination of side chain interactions (Ye et al., 2022).

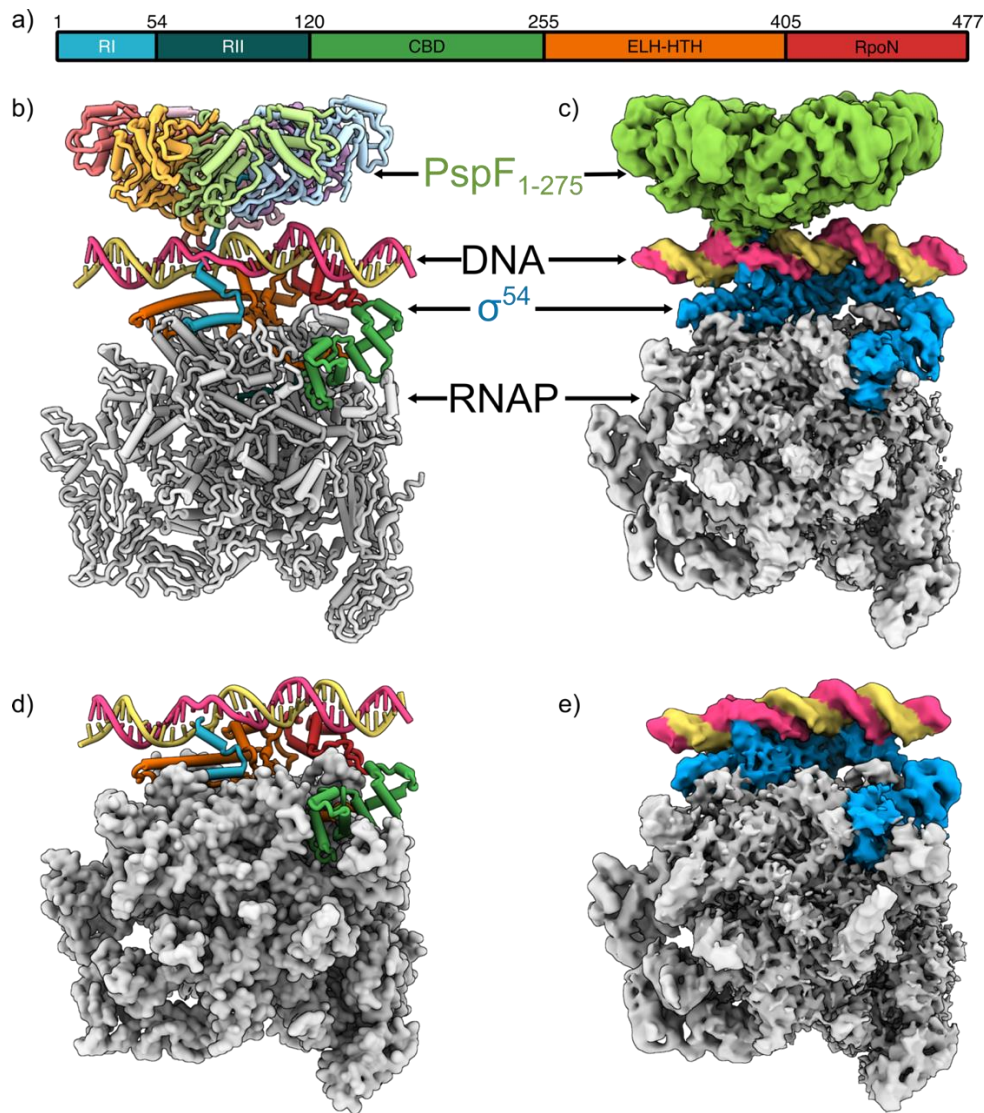


Figure 4.2.1: a) Domain map of σ^{54} . b) Model and c) composite map of RPI(-12/-11). d) Model and e) map of RPo(WT)

σ^{54} imposes tight spacing requirements between the -24 and -12 elements of the promoter and both structures show the mechanism DNA distortion by σ^{54} region III. The RpoN domain binds to the -24 region of the promoter DNA via conserved R455 and 456 residues that form base-specific interactions with the major groove from -27 to -23 (Figure 4.2.2a-b). The HTH domain binds the adjacent major groove by binding to conserved 5'-T₋₁₅G₋₁₄-3' bases on the non-template strand (Figure 4.2.2c).

The transcription bubble is nucleated by region I and region III-ELH by inserting the N-terminus of region I in between the DNA strands, separating the two DNA strands to create a fork junction structure (Figure 4.2.2d). The -12T base on the template strand is flipped out and stabilised by S335 on ELH, and the 10 Å-wide bubble is stabilised by M15 on region I which forms hydrophobic interactions with the DNA strand at the -12/-11 position (Figure 4.2.2e,f). The remainder of region I is flexible and unresolved in the RPi state.

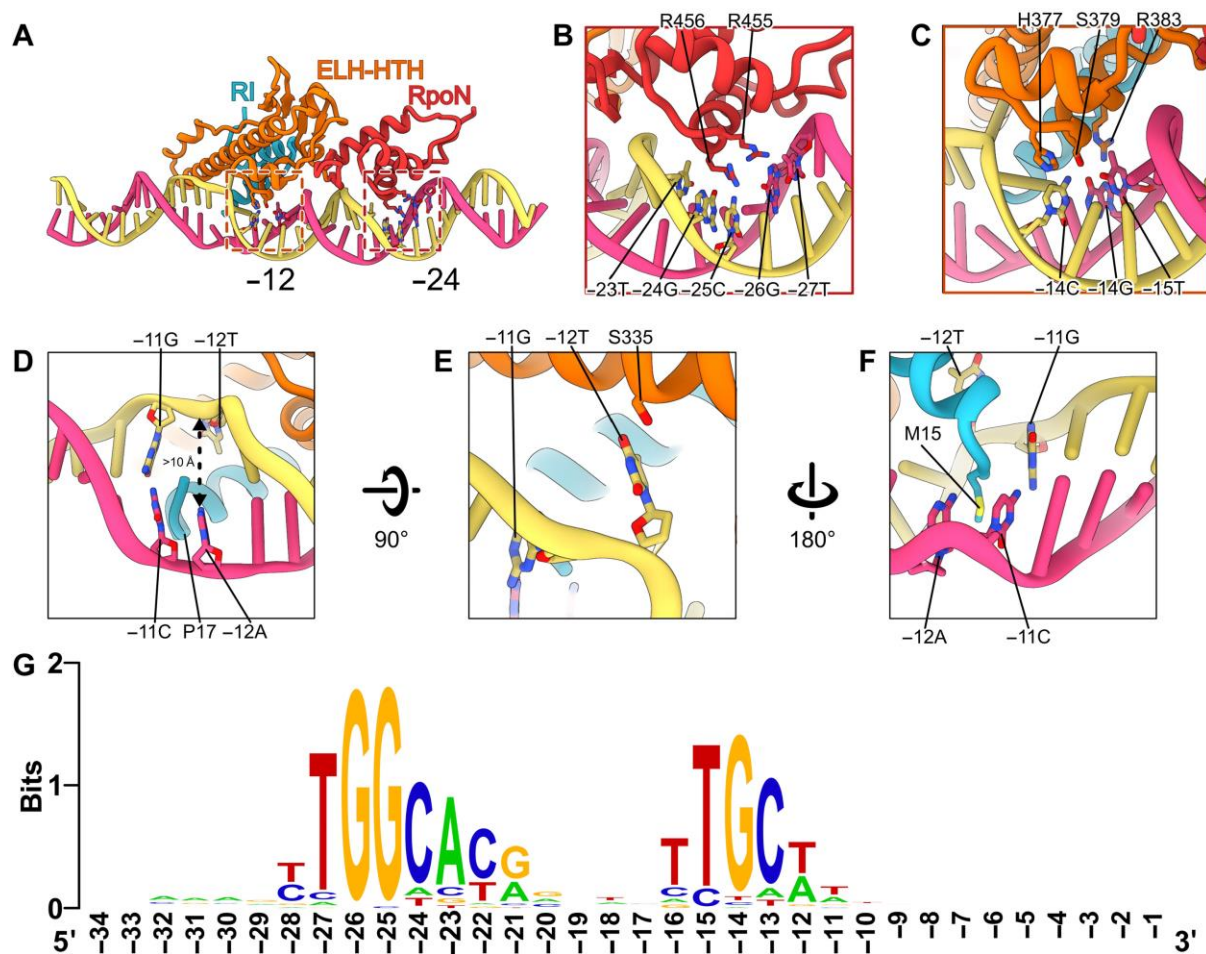


Figure 4.2.2: σ^{54} interaction with the *nifH* promoter DNA reveals nucleation of the transcription bubble by region I and ELH, stabilised by major groove interactions via region III-RpoN and HTH. a) overview of the DNA binding elements of σ^{54} in the RPi conformation. b) RpoN and c) HTH interactions. d, e) ELH stabilises the flipped out base in the -12 position. f) M15 inserting in between the 2 DNA strands.

In contrast, the RPi(-12/-11) complex enabled tracing of the entire, σ^{54} region I N-terminus. The focused refinement map of the PspF₁₋₂₇₅- σ^{54} -DNA also allowed for determination of nucleotide occupancy within the PspF₁₋₂₇₅ hexameric ring (Figure 4.2.3a-c).

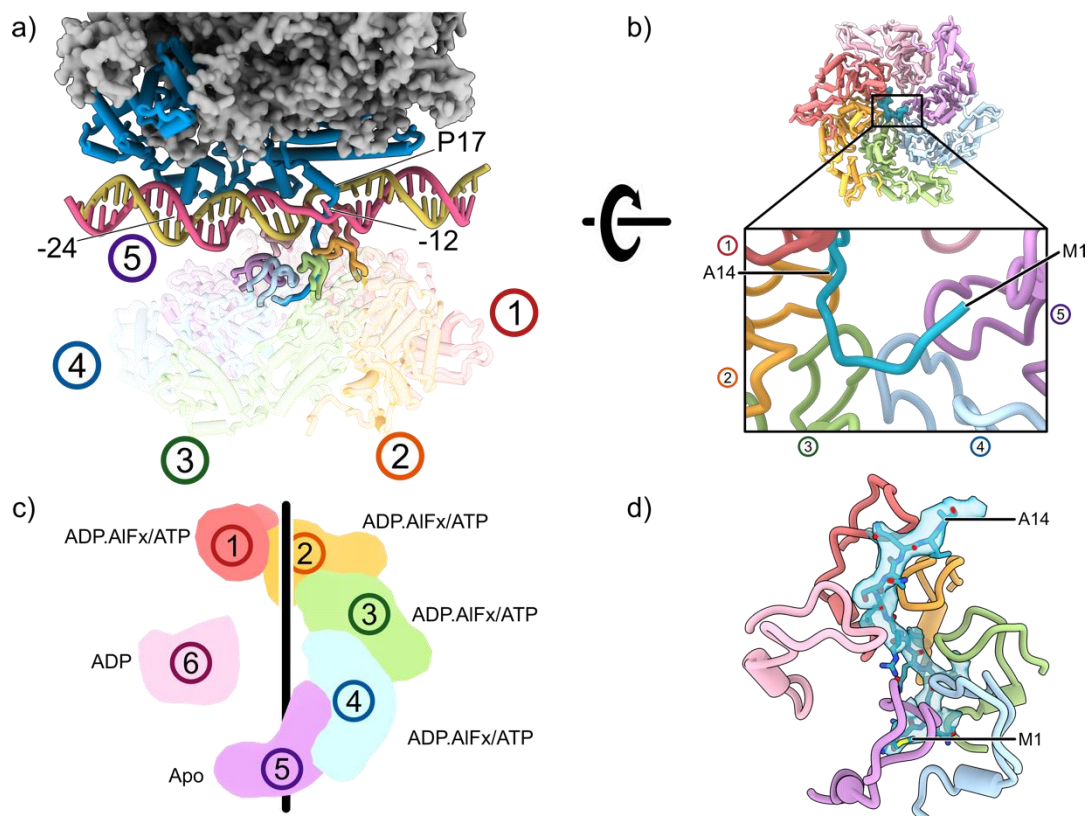


Figure 4.2.3: a) PspF₁₋₂₇₅ ring highlighting arrangement of L1 loops contacting RNAP- σ^{54} -DNA. b) Region I N-terminus as viewed from the top of the PspF₁₋₂₇₅ ring. c) Schematic depiction of promoter arrangement with the nucleotide occupancies. d) Cartoon of the L1 loops contacting the region I N-terminus.

The new model provided evidence that PspF₁₋₂₇₅-mediated remodelling of the closed complex involves the L1 loops interacting with the mostly hydrophobic residues on the N-terminus of region I (Figure 4.2.3d). Based on previous high resolution ATP- and ADP- bound NtrC and PspF₁₋₂₇₅ crystal structures, we observed that the ATP-

bound form is correlated with the raising of the L1 loop, whilst the ADP-bound form has a folded down L1 loop (Rappas et al., 2006). We proposed that the nucleotide states control L1 loop conformations with an extended L1 loop that allows protomers to be ATP hydrolysis ready (Rappas et al., 2006; Zhang & Wigley, 2008).

However, how ATP hydrolysis is linked to DNA bubble melting was unclear.

Consequently, we analysed the RPi(-12/-11) model to determine a potential mechanism for protomer coordination. This analysis reveals the precise nucleotide bound and functional states of PspF₁₋₂₇₅ subunits within the hexamer ring.

Correlating these functional states with DNA and σ^{54} interactions, we propose a mechanistic model that couples the ATP hydrolysis cycle within the bEBP hexamer with σ^{54} RI unfolding and DNA opening. Furthermore, we assembled RPi complexes with increasing lengths of DNA mismatches to simulate increased DNA unwinding and determined cryo-EM structures of intermediate complexes with -11 to -8 (-11/-8) and -10 to -1 (-10/-1) mismatches.

4.3 Analysis of RPi(-12/-11) PspF₁₋₂₇₅ protomer coordination and interactions

After determining the structure of the RPi(-12/-11) structure, we analysed the inter-protomer interfaces, nucleotide binding states and conformational differences to try to determine how ATP hydrolysis is coordinated between protomers within the PspF₁₋₂₇₅ hexameric ring (Figure 4.3.1a). The nucleotide is coordinated by conserved

walker A, walker B and sensor 2 motifs *in cis*, with *in trans* interactions from R-fingers in the adjacent protomer binding to the β and γ phosphate groups (in this instance, AlFx) (Figure 4.3.1b,c).

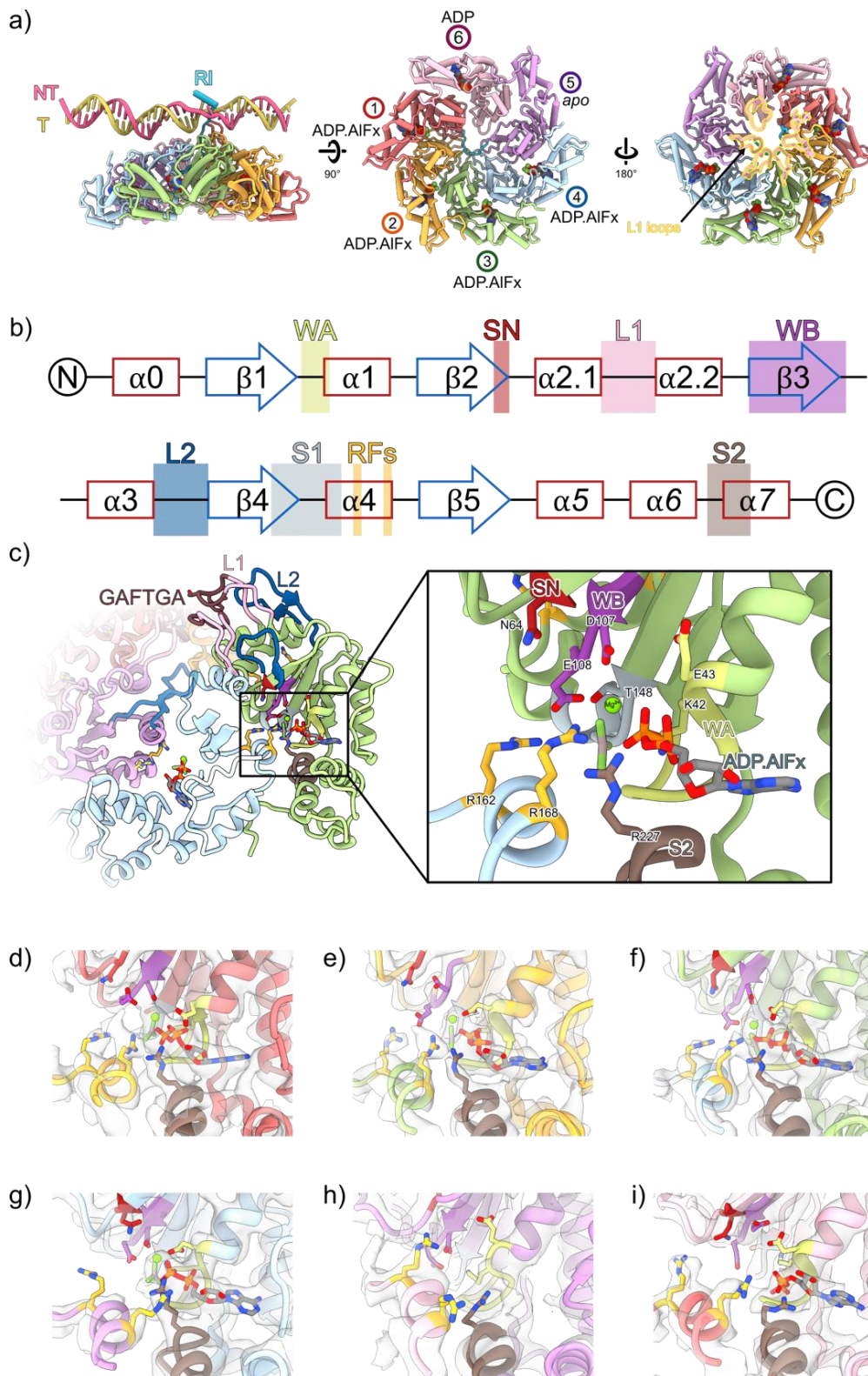


Figure 4.3.1: a) overview of the PspF₁₋₂₇₅ hexameric ring side and top views b) secondary structure schematic of PspF₁₋₂₇₅. c) active site interactions include *in trans* contributions from R finger in adjacent protomers. d-i) Zoomed in views of the nucleotide binding pocket from each protomer.

Based on the nucleotide density, we found that protomers 1-4 were bound by ADP.AIFx (Figure 4.3.1d-g), whereas protomer 5 was empty (*apo*) and protomer 6 was ADP-bound (Figure 4.3.1h,i).

Previous studies on AAA+ proteins have revealed that the relative closure between the smaller alpha-helical lid (α lid) and the larger $\alpha\beta$ sandwich subdomains is correlated with ATP-hydrolysis state (Chen et al., 2010; Han et al., 2017; Lo et al., 2019; Majumder et al., 2019; Wald et al., 2022). Protomers 1-4 were in a similar, more open subdomain conformation (Figure 4.3.2a, b) whilst in protomers 5 and 6, the subdomains are more closed, correlating with their nucleotide states (Figure 4.3.2c).

Furthermore, the interface between protomer 1 and protomer 6 ($928 \text{ \AA}^2$) as well as protomer 6 and protomer 5 ($860 \text{ \AA}^2$) is reduced compared to the other interfaces ($1484 \text{ \AA}^2$ between 4 and 5, $1707 \text{ \AA}^2$ between 3 and 4, $1515 \text{ \AA}^2$ between 2 and 3, $1509 \text{ \AA}^2$ between 1 and 2) suggesting protomers 5 and 6 are less constrained inside the hexamer. Viewed from the side, the PspF L1 loops from protomers 1-5 form a spiral with protomer 1 at the highest point (closest to σ^{54} RI) while protomer 5 at the lowest point. Protomer 6 is between protomer 1 and protomer 5 (Figure 4.2.3c-d).

The nucleotide states of the PspF hexamer suggest that protomer 1 is likely the subunit to hydrolyse ATP next, then becoming an ADP-bound state. Comparing the structures of protomer 1 and protomer 6 reveals the conformational changes in PspF protomer upon hydrolysis (Figure 4.3.2d). Consistent with previous crystal structures of PspF₁₋₂₇₅, ATP to ADP transitions causes the domain closure between

α/β subdomain and the α -helical domain (Supplementary Movie 1) (Rappas et al., 2006). Significantly we also observe the σ^{54} interacting L1 loop folds downward from an extended upward conformation in protomer 1 as it transitions to protomer 6 (Figure 4.3.2f).

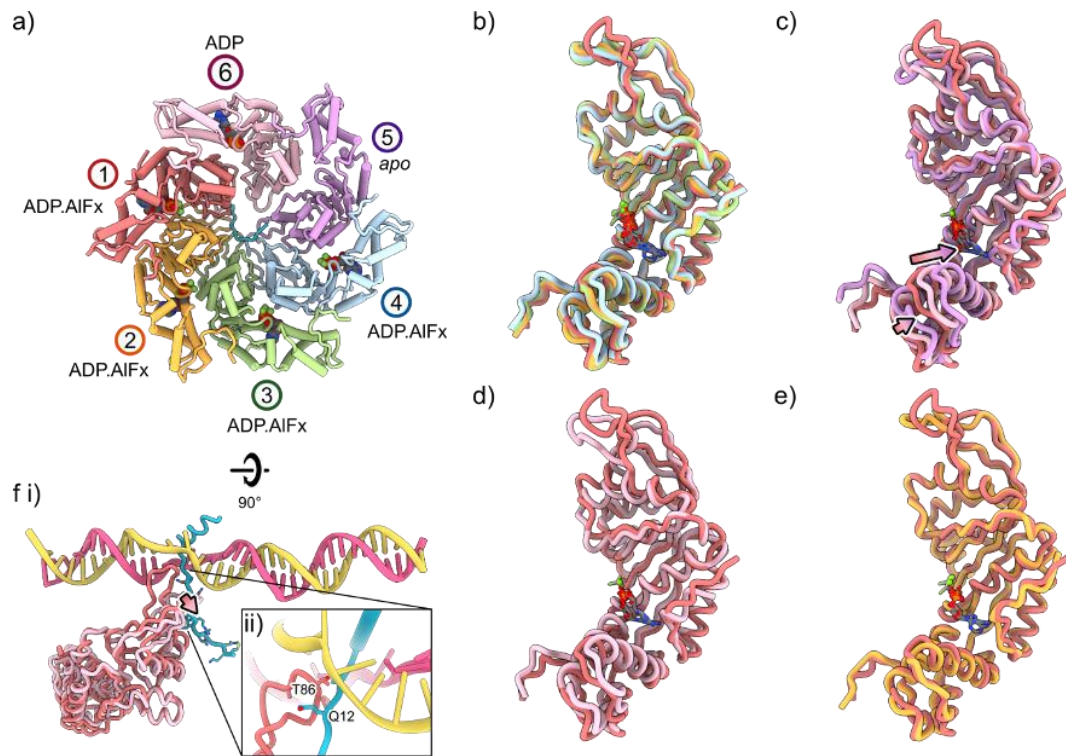


Figure 4.3.2: a) overview nucleotide binding state in the hexamer. b-e) alignment of the protomers by the $\alpha\beta$ sandwich subdomain. b) ATP bound protomers show a similar level of scrunching of the $\alpha\beta$ sandwich and α lid subdomains. c) ATP (protomer 1), ADP and *apo* bound comparison shows scrunching, with ATP $\rightarrow$ ADP $\rightarrow$ *apo* being the order from least to most closed. d) the proposed power stroke, comparing ATP and ADP-bound states. e) proposed priming transition, comparing protomer 2 and protomer 1, showing that the L1 loop in protomer 1 is in a more raised position. f i) The proposed power stroke on region I would pull region I through the PspF₁₋₂₇₅ central pore. f ii) T86 from L1 packs against Q12 on region I, and likely exerts force on Q12 upon hydrolysis.

Upon hydrolysis of protomer 1, the conformational changes in protomer 1 would cause steric clashes with protomer 2 at multiple positions on the α/β subdomain, as well as the GAFTGA motif within the L1 loop (Figure 4.3.3a, panel 1). Given that protomer 2 is tightly packed against protomer 3 while the interface between protomer 1 and protomer 6 is looser, upon hydrolysis, protomer 1 will detach from protomer 2. The closing of the protomer 1 subdomains upon hydrolysis would also create steric clashes with the protomer 6 α -lid (Figure 4.3.3a, panel 2), suggesting that protomer 6 must also shift slightly towards protomer 5 to enable ATP hydrolysis from protomer 1.

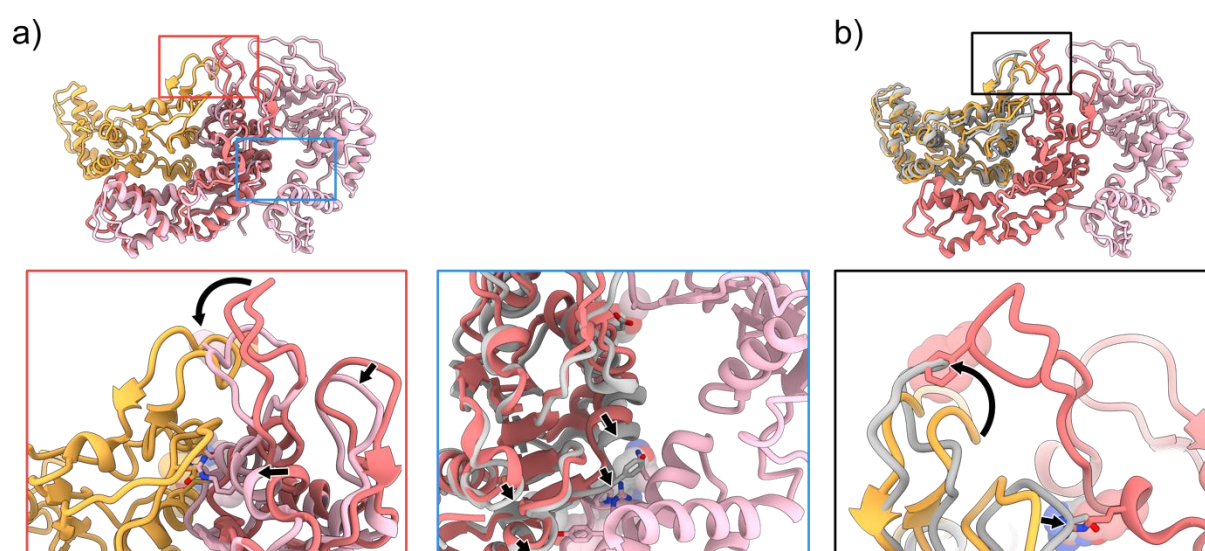


Figure 4.3.3: The proposed conformational changes within the hexameric ring suggests that a) ATP hydrolysis from protomer 1 leads to detachment of protomer 1 from protomer 2 (red panel) and requires protomer 6 to move towards protomer 5 to prevent steric clashing (blue panel). Upon shifting of protomer 6, protomer 1 is free to hydrolyse and move to position 6 (shown in grey) b) Protomer 2 is prevented from adopting a raised position by protomer 1; upon detachment, protomer 2 is free to move up into the raised position (shown in grey).

There are extensive interactions between protomer 1 through to protomer 4 (Figure 4.2.3d, Figure 4.3.1a). The conserved L1 and L2 loops, which point towards the hexamer pore (Figure 4.3.1a), contribute to the interfaces. The conserved phenylalanine on the L1 GAFTGA loop, S135 of L2 and R98 on α 2.2 of protomer 1 sterically blocks the L1 loop of protomer 2 from raising upwards. L1 loops in protomers 2-4 are therefore constrained and adopt a folded conformation. The detachment of protomer 1 from protomer 2 would relieve the constraint on L1 and L2 loops in protomer 2, allowing the L1 loop to be extended, reaching to interact with σ^{54} RI, adapting new position. The extended L1 loop would also reposition the nucleotide binding pocket, making it hydrolysis ready (Figure 4.3.3b).

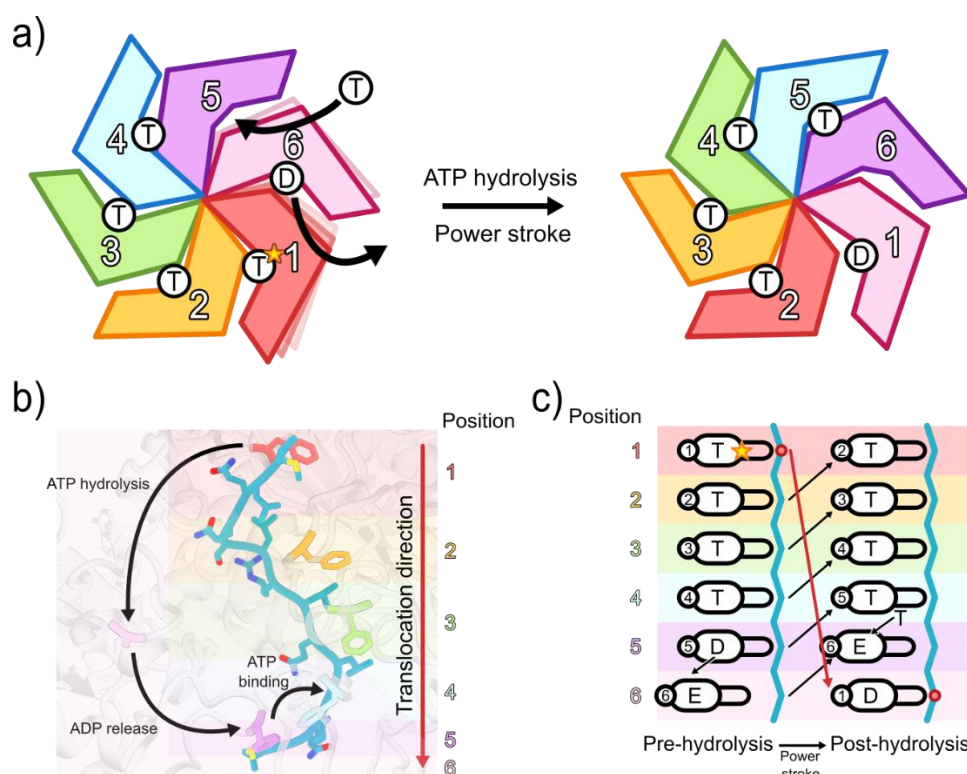


Figure 4.3.4: Proposed mechanism of ATP hydrolysis and substrate translocation. a) Cartoon schematic of movement of protomer 1 during the power stroke leads to dissociation from protomer 2 and shifting of protomer 6 towards protomer 5, facilitating protomer 2 L1 loop opening, and nucleotide release by protomer 6. b) Structural schematic showing direction of translocation facilitated by ATP hydrolysis. c) Cartoon schematic showing the movement of protomers down the central axis of PspF₁₋₂₇₅ during ATP hydrolysis.

The detachment of protomer 1 from 2 not only enables protomer 2 to adapt hydrolysis ready, σ^{54} RI engaged state, it would also cause clashes with protomer 6, causing protomer 6 (ADP bound) to move towards protomer 5 and release ADP, thus becoming nucleotide free (*apo*) state (Figure 4.3.4a). The engagement of protomer 6 with protomer 5 would promote ATP binding to protomer 5, completing the shift of nucleotide states within the hexamer (Figure 4.3.4b-c, Supplementary Movie 2-3).

The proposed mechanism suggests the power stroke occurs when protomer 1 hydrolyses ATP to adopt the folded-down ADP conformation. The L1 loop in protomer 1 has the most extensive interactions with σ^{54} Region I and DNA. Indeed, Thr86 of L1 loop packs against Gln12 of σ^{54} RI (Figure 4.3.2f ii). Upon hydrolysis, the L1 loop folds downwards, and this movement would thus push Gln12 and RI with it. The remaining PspF protomers form primarily hydrophobic “slippery” interactions with the conserved hydrophobic residues on σ^{54} region I (residues M1,L5,L7,L13). enabling sliding of the rest of region I (Figure 4.3.2f i). The translocation of RI would eventually cause unfolding of RI H1, which has been shown to be crucial in maintaining the stable closed complex.

Significantly, in addition to their interactions with σ^{54} , L1 loops makes specific interactions with promoter DNA, thus the changes in L1 loop conformation within the hexamer would also change interactions with DNA. Further RI transverses through DNA strands and the translocating of RI could also change interactions with DNA. These altered interactions could contribute to the DNA melting induced by PspF ATP hydrolysis.

To investigate these mechanisms further, we determined structures of RNAP- σ^{54} -DNA-PspF₁₋₂₇₅ complexes with increasing stretches of DNA mismatch to mimic downstream DNA opening.

4.4 Purification of intermediate complex protein components

4.4.1 RNA polymerase

Expression and purification of RNAP, σ^{54} , and PspF₁₋₂₇₅ is detailed in section 3.4.

Briefly, RNA polymerase was isolated from cell extract by polymin P precipitation followed by ammonium sulphate precipitation, cation exchange and anion exchange chromatography using Biorex70 and HiTrap Q columns, respectively.

RNAP was obtained to high purity as indicated by the SDS-PAGE gel from the HiTrap Q column, showing clear bands for all RNAP subunits with minimal contamination, as well as a narrow elution peak in the elution profile (Figure 4.4.1).

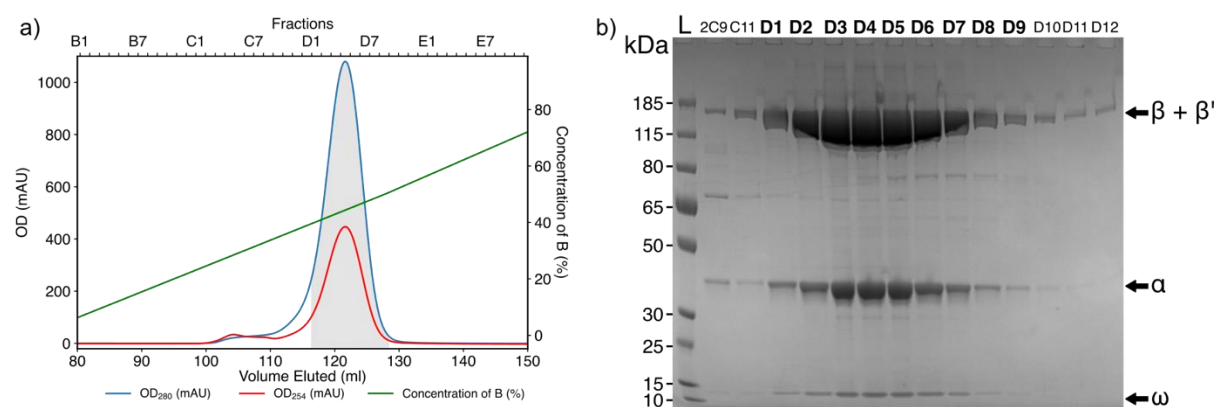


Figure 4.4.1. a) Elution profile of RNA polymerase from final anion exchange chromatography step.

Concentration of buffer B is 1M TGET buffer, and is shown in green, OD₂₈₀ is shown in blue and OD₂₅₄ is shown in red. Shaded region in the elution profile represent pooled fractions, as shown in bold in the SDS-PAGE profile in b). b) SDS-PAGE gel of fractions from HiTrap Q anion exchange column. L= molecular weight ladder.

4.4.2 σ^{54}

σ^{54} was expressed purified as described in section 3.4.1. Gel filtration with a Superdex Hiload 75 16/600 column and subsequent SDS-PAGE analysis demonstrated that σ^{54} fractions were pure and contained minimal contamination (Figure 4.4.2). Identical protocol was followed for σ^{54} R336A mutants. As shown previously, σ^{54} migrates at a higher than expected molecular weight (Chen et al., 2007; Ye et al., 2020). This could be due to the highly acidic region II, resulting in inefficient SDS-binding and lowered electrophoretic mobility (Tiwari, Kaila & Guptasarma, 2019). The identity of σ^{54} was later verified and supported by native-PAGE binding RNAP, DNA, PspF₁₋₂₇₅ and cryo-EM reconstructions.

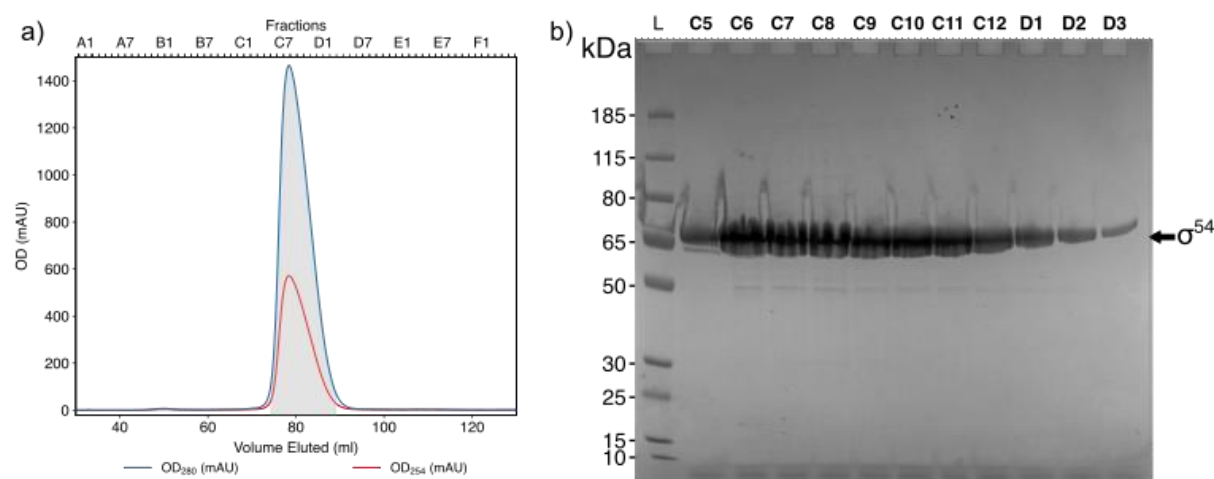


Figure 4.4.2. a) Elution profile of σ^{54} from final gel filtration step. Shaded region in the elution profile represent pooled fractions, as shown in bold in the SDS-PAGE profile in b). b) SDS-PAGE gel of fractions from gel filtration column. L= molecular weight ladder.

4.4.3 RNAP- σ^{54} holoenzyme

RNAP- σ^{54} holoenzyme was formed by mixing RNAP and σ^{54} in a 1:4 ratio, followed by gel filtration. As seen in figure 4.4.3, gel filtration produced 2 clear peaks in the

gel filtration profile, with minimal void volume. All components were present with minimal contamination, making it suitable for cryo-EM analysis.

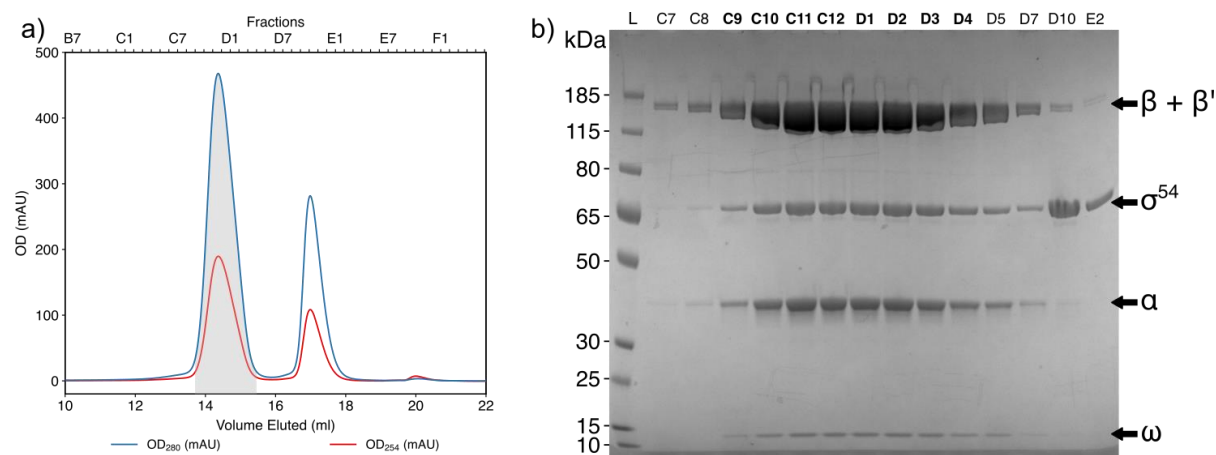


Figure 4.4.3. a) Elution profile of σ^{54} from final gel filtration step. Shaded region in the elution profile represent pooled fractions, as shown in bold in the SDS-PAGE profile in b). b) SDS-PAGE gel of fractions from gel filtration column. L= molecular weight ladder.

4.4.4 PspF₁₋₂₇₅

PspF₁₋₂₇₅ was expressed and purified as described in section 3.4.2. Briefly, N-terminal poly-His-tagged PspF was extracted from cell lysate by nickel affinity chromatography, followed by tag cleavage and gel filtration with a Superdex Hiload 75 16/600 column. The final gel filtration step demonstrates that PspF₁₋₂₇₅ was pure and with minimal contamination, as shown in Figure 4.4.4.

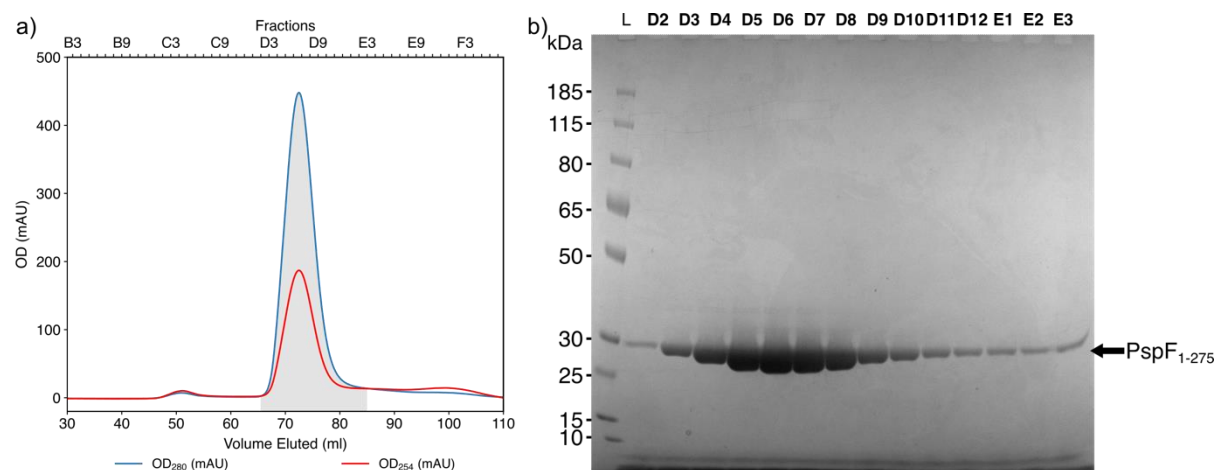


Figure 4.4.4. a) Elution profile of PspF₁₋₂₇₅ from final gel filtration step. Shaded region in the elution profile represent pooled fractions, as shown in bold in the SDS-PAGE profile in b). b) SDS-PAGE gel of fractions from gel filtration column. L= molecular weight ladder.

4.5 Optimisation of complex assembly

4.5.1 Choice of DNA scaffolds

Our previous RPip, RPo and RPitc structures used mismatches from -10 to -1 respectively (Glyde et al., 2018). Given that the RPitc complex includes a melted region from -10 to +4, we rationalised that the DNA mismatch from -12/-11 would have to be melted from the downstream edge to enable loading of the +1 base pair into the catalytic site. Consequently, we tested a range of mismatch lengths (Figure 4.5.1) to mimic the further opening of the nascent transcription bubble during open complex formation.

4.5.2 Native-PAGE and longer transcript gels of RNAP- σ^{54} -DNA-PspF complexes

We optimised the assembly RNAP- σ^{54} -DNA-PspF complexes using increasing lengths of DNA mismatches (from fully duplexed to opened from -10 to -1 and -11

to -3) using RNAP. We found that whereas ATP, ADP.BeF₃, ATP γ S and AMPPNP formed no or unstable complexes that did not yield promising 2D classes, ADP.AIFx formed stable intermediate complexes with a wide range of DNA mismatch lengths (Figure 4.5.1).

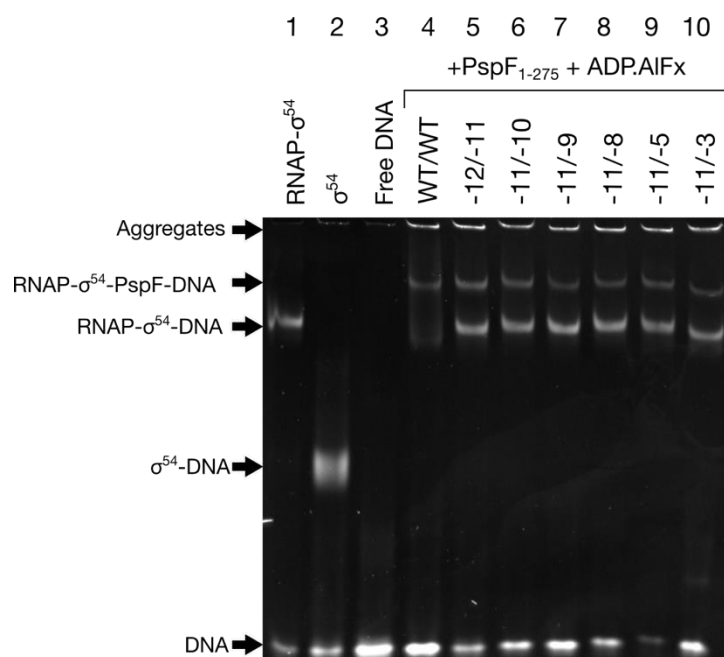


Figure 4.5.1. Native-PAGE gel of RNAP- σ^{54} -PspF₁₋₂₇₅ complexes trapped with ADP.AIFx and varying DNA mismatch lengths (lanes 4-10)

Following verification of complex formation, we then tested competency of transcription both with ATP and with ADP.AIFx. Interestingly, -11/-8 mismatched DNA was transcription incompetent in the presence of both ATP and ADP.AIFx. This suggests that RP(-11/-8) complexes represent a trapped state. On the other hand, -11/-3 mismatched and -11/-5 mismatched DNA was transcription competent in the presence of ATP and ADP.AIFx. As expected, fully duplexed DNA did not produce transcript except with ATP, and -10/-1 DNA was transcription competent with both ADP.AIFx, ATP and without PspF₁₋₂₇₅ (Figure 4.5.2).

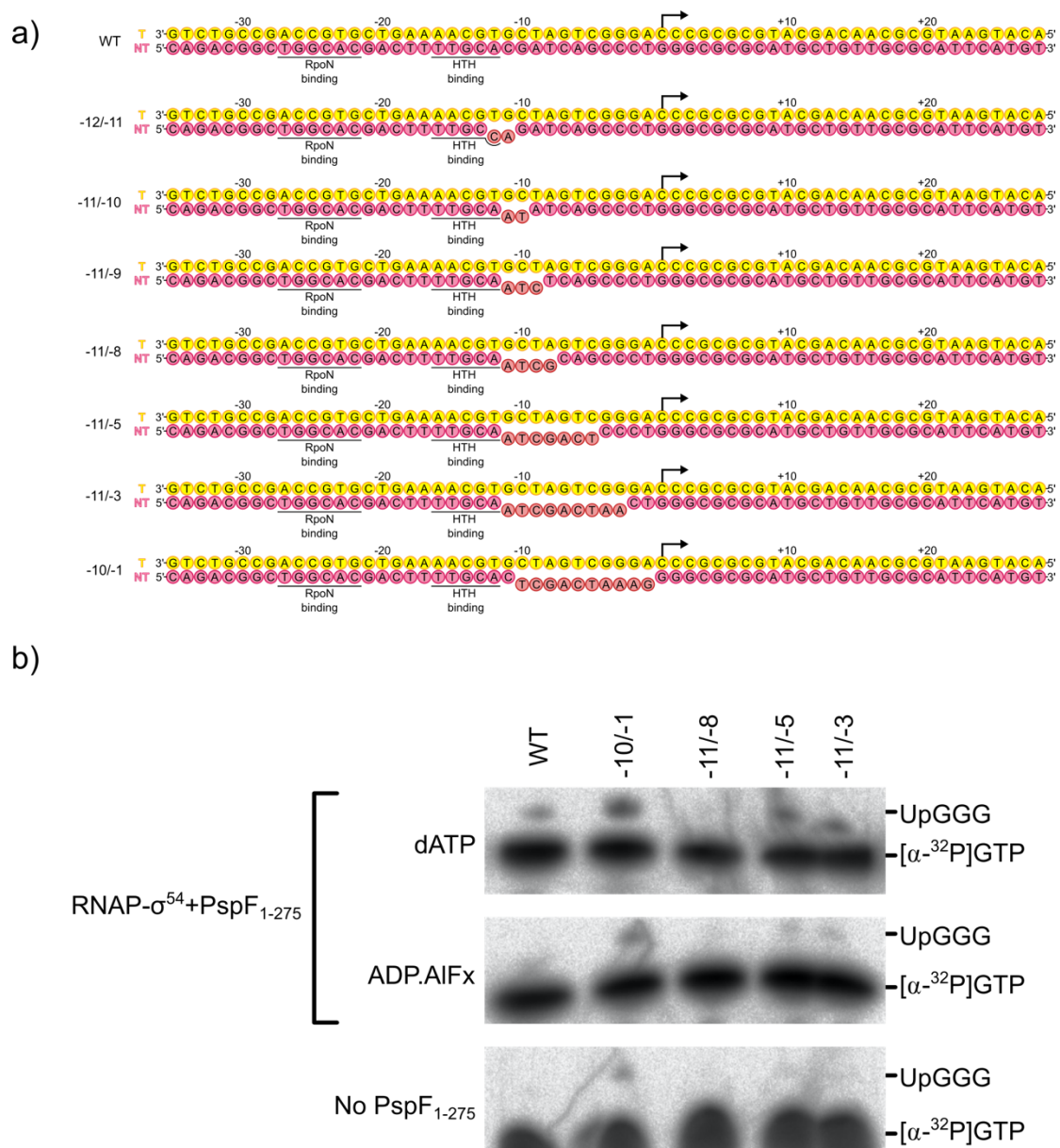


Figure 4.5.2. a) Design of *nifH* DNA sequences. All mutations were done on the non-template strand, indicated in red. Consensus σ^{54} binding sequences highlighted. b) Short primed RNA assay detects α -³²P labelled GTP in short RNA transcripts show that all except RP(-11/-8) are transcription competent with dATP. Pre-opened (-10/-1) and fully duplexed (WT) DNA was used as positive and negative controls respectively.

4.6 RP(-11/-8) complexes

4.6.1 Sample Preparation of the RP(-11/-8) complex

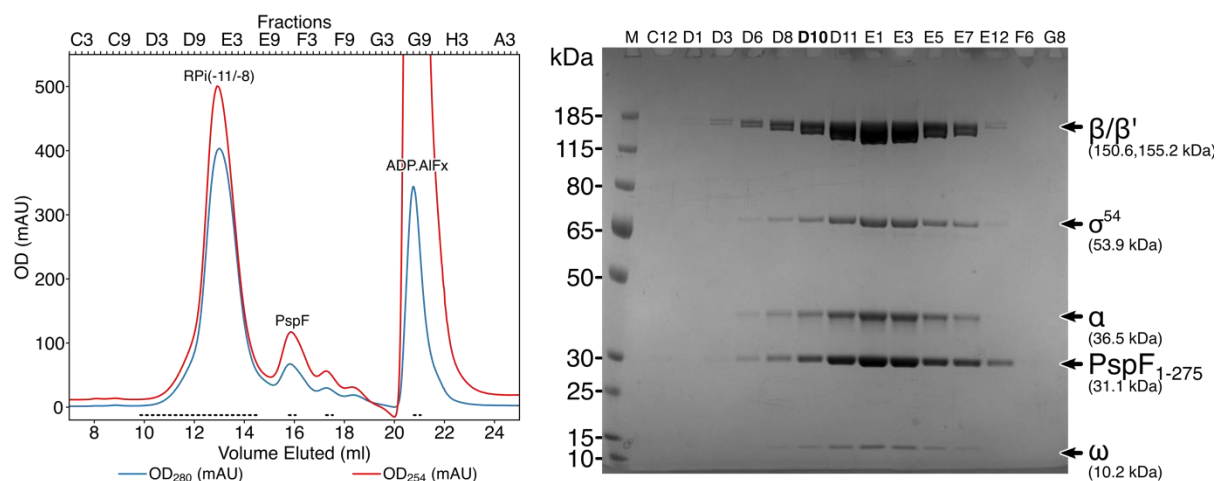


Figure 4.6.1: Gel filtration profile of RP(-11/-8) and SDS-PAGE analysis of fractions. Fraction used in grid preparation is highlighted in bold.

We initially used gel filtration to assess the purity of RP(-11/-8) complexes. The gel filtration profile demonstrated that RP(-11/-8) could be formed in a high purity, with peak fractions of suitable homogeneity and concentration for cryo-EM analysis. We used fraction D10, located on the left shoulder of the RP(-11/-8) peak, which had a concentration of ~0.5 mg/ml for grid preparation.

4.6.2 Image processing of RP(-11/-8) complexes

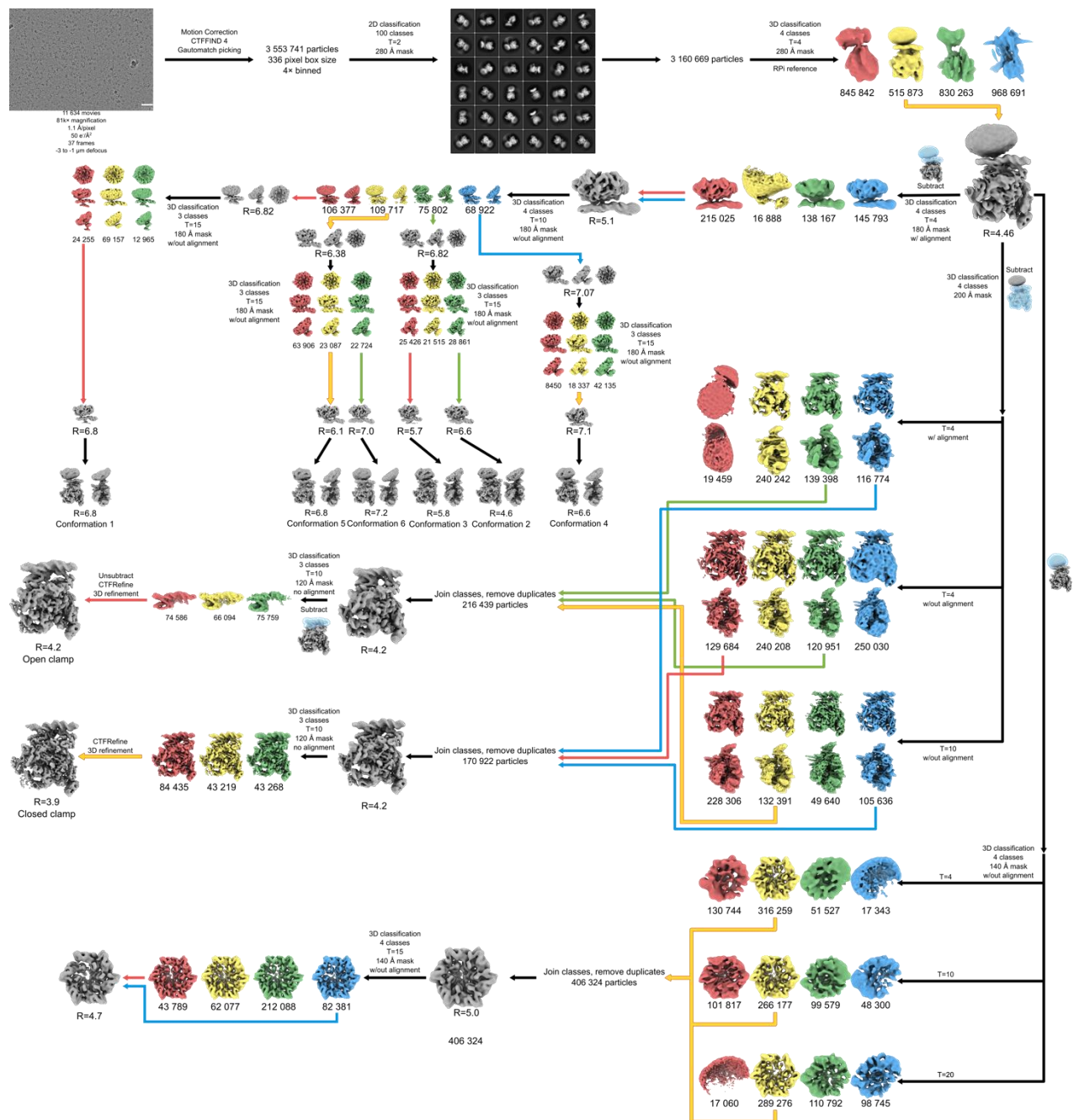


Figure 4.6.2: Overall processing flowchart for RP(-11/-8) complexes. Arrow colour indicated class

selected for further processing. Numbers below each class represents number of particles. For detailed breakdown, see figures below.

The overall workflow of image processing is shown in Figure 4.6.2. Following pre-processing (as described in section 3.7.4). We selected the initial 2D and 3D classes based on presence of clear density for both PspF₁₋₂₇₅ and DNA.

Following the first round of 3D classification and refinement (Figure 4.6.3a), we found that the PspF-DNA interactions suffered from a large amount of flexibility, which limited the resolution of the PspF ring, as well as preferential orientation (Figure 4.6.3b,c). Consequently, we used particle subtraction and recentring on 3 separate regions to obtain higher-resolution reconstructions of specific features within the complex.

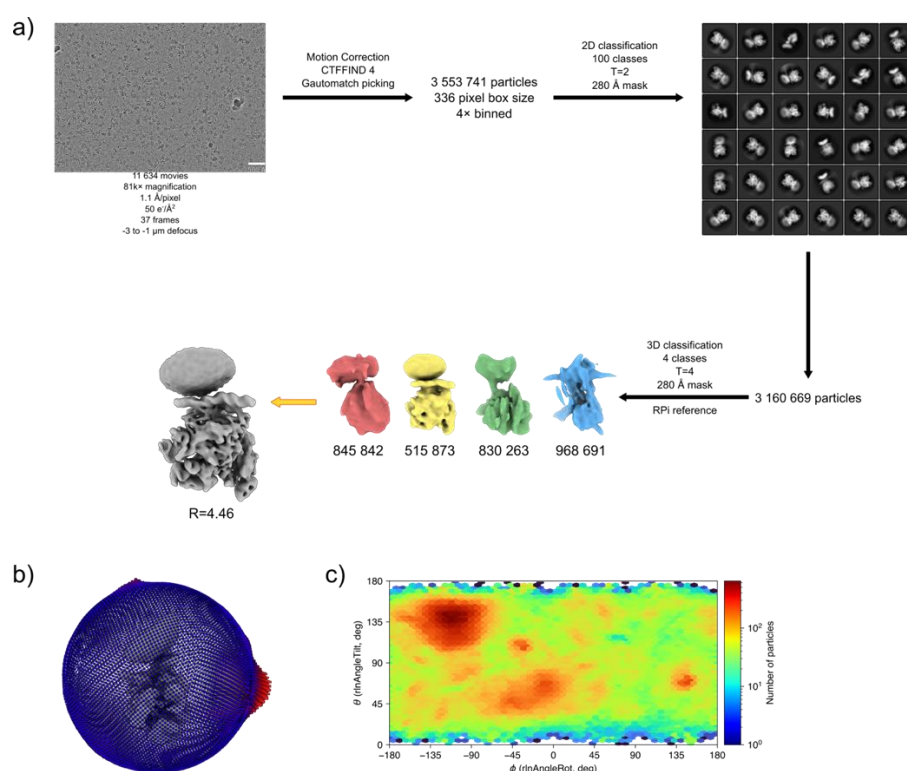


Figure 4.6.3: a) Pre-processing and 1st round of classification and refinement showed features of RNAP, σ^{54} , DNA and PspF₁₋₂₇₅. Numbers below each class represents number of particles. The first refinement showed preferential orientation, as indicated by the spherical (b) and flattened (c) angular distribution map of the first refinement.

The first subtraction and recentring focused on the PspF₁₋₂₇₅-DNA- σ^{54} region I to identify PspF₁₋₂₇₅-region I and PspF₁₋₂₇₅-DNA interactions (Figure 4.6.4a). We found that subtracting and focusing on the PspF₁₋₂₇₅-DNA- σ^{54} region I first and then reverting back to original particles to perform global refinements produced the best overall maps. The second subtraction and recentring focused on the RNAP- σ^{54} -DNA complex to identify clamp opening during PspF₁₋₂₇₅ remodelling. The third subtraction and recentring focused of PspF₁₋₂₇₅ alone, to identify interprotomer interfaces, as well as improve the overall resolution of the PspF₁₋₂₇₅ ring and region I.

From the PspF₁₋₂₇₅-DNA- σ^{54} subtraction processing strategy, we identified 6 separate conformations, varying from 5.7-8 Å (Figure 4.6.4b, Figure 4.6.6). By reverting back to original particles and performing global refinements, we could then correlate the PspF₁₋₂₇₅-DNA conformations with clamp opening.

The second subtraction and recentring focused on the RNAP- σ^{54} -DNA, to find conformations with altered states of DNA loading or clamp opening (Figure 4.6.7a). Whilst we could not find clamp loading intermediates, we did identify 2 conformations, open and closed, which was refined to 4.2 and 3.9 Å overall resolution respectively (Figure 4.6.7b,c). The final subtraction was on PspF₁₋₂₇₅ alone (Figure 4.6.8a); the PspF₁₋₂₇₅ subtracted processing resulted in a 4.7 Å final resolution reconstruction (Figure 4.6.8b-d).

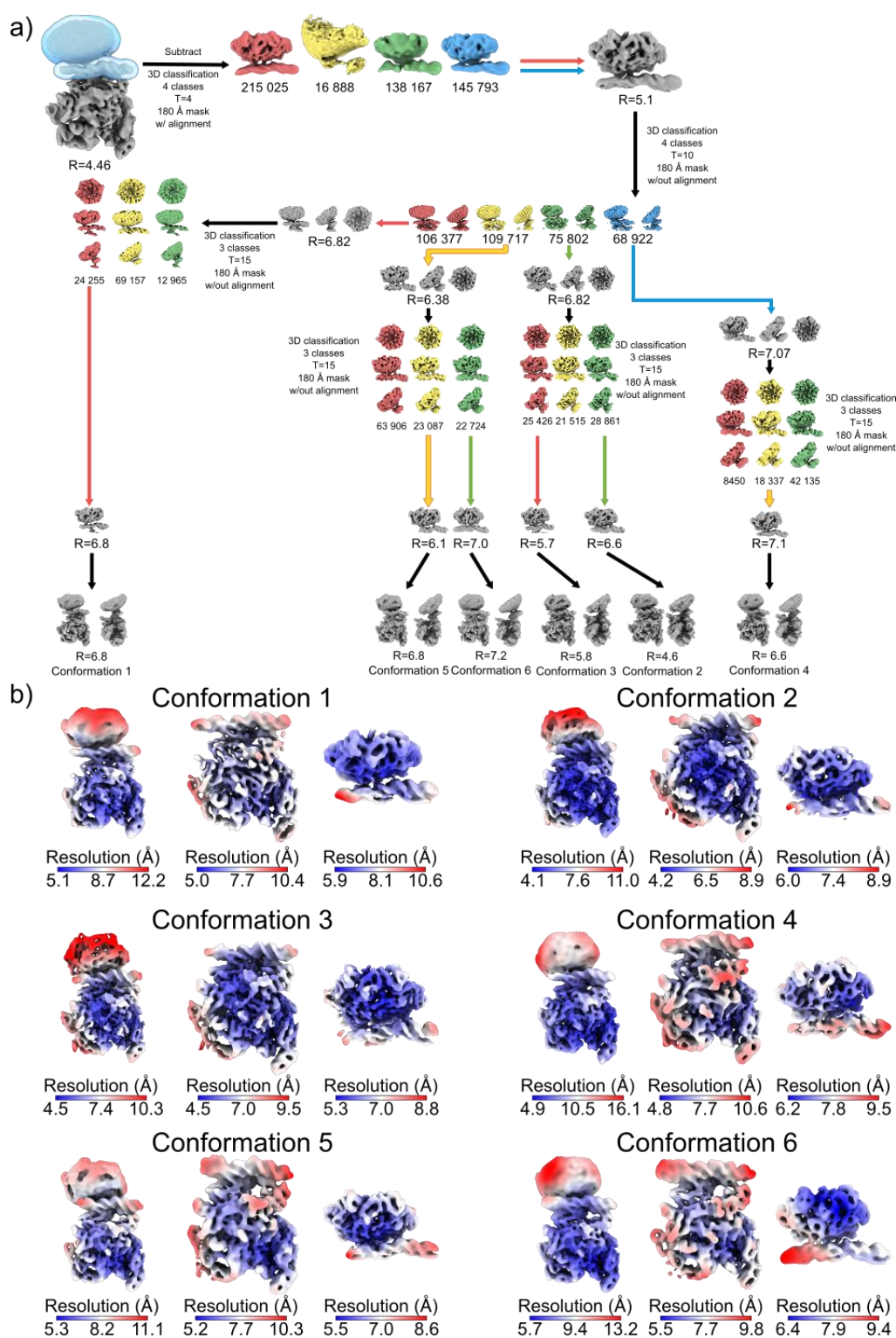


Figure 4.6.4: a) Flowchart of PspF₁₋₂₇₅-DNA- σ^{54} region I subtraction processing pipeline. Numbers below each class represents number of particles b) Local resolution maps of the 6 conformations found, including local resolution maps for global, PspF₁₋₂₇₅-DNA- σ^{54} region I-focused and RNAP- σ^{54} -DNA focused maps for each conformation.

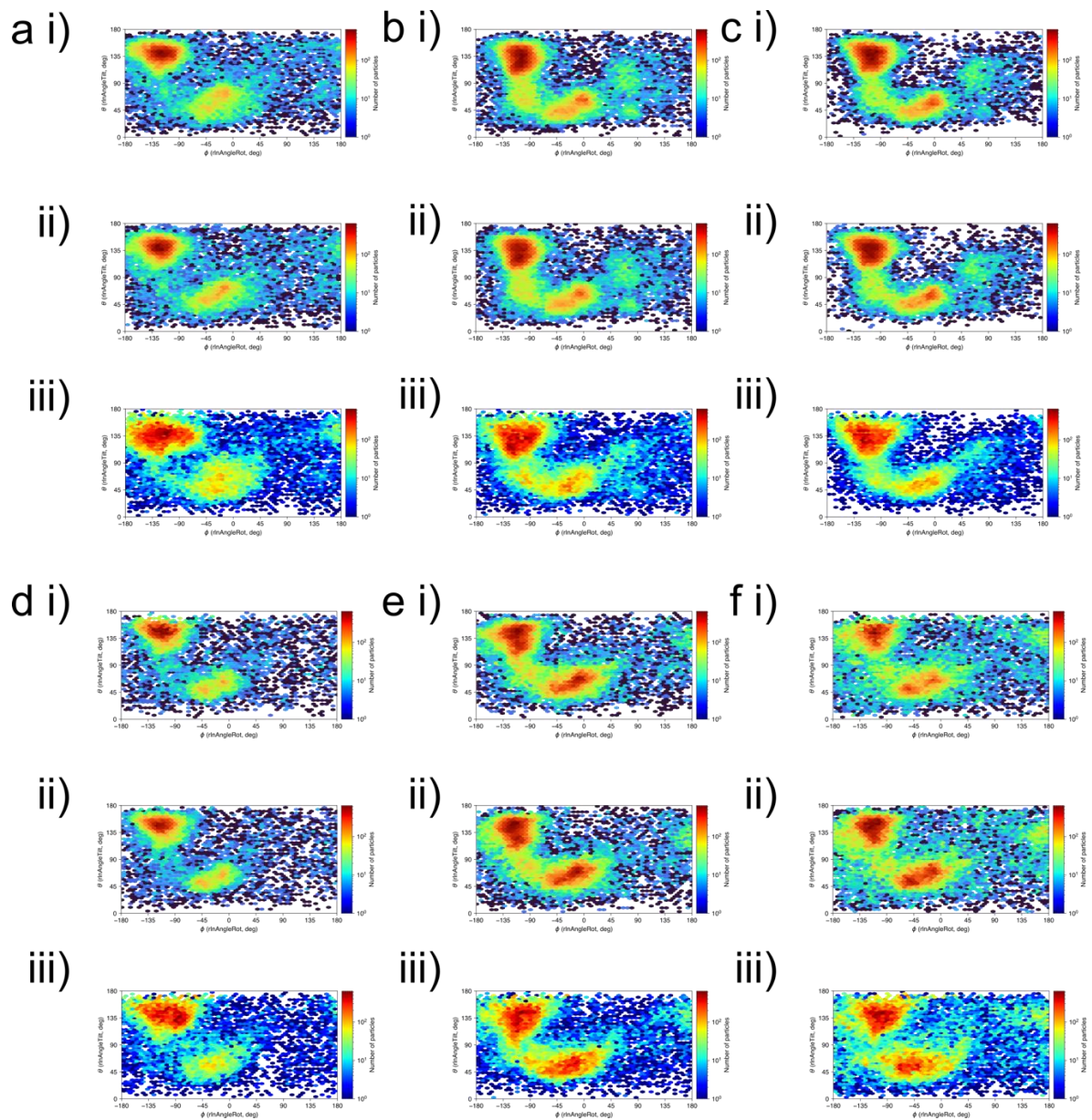


Figure 4.6.5: Angular distribution map of the global (i) RNAP- σ^{54} -DNA (ii) and PspF₁₋₂₇₅-DNA-region I (iii)

focused refinement maps of conformations 1-6 (a-f).

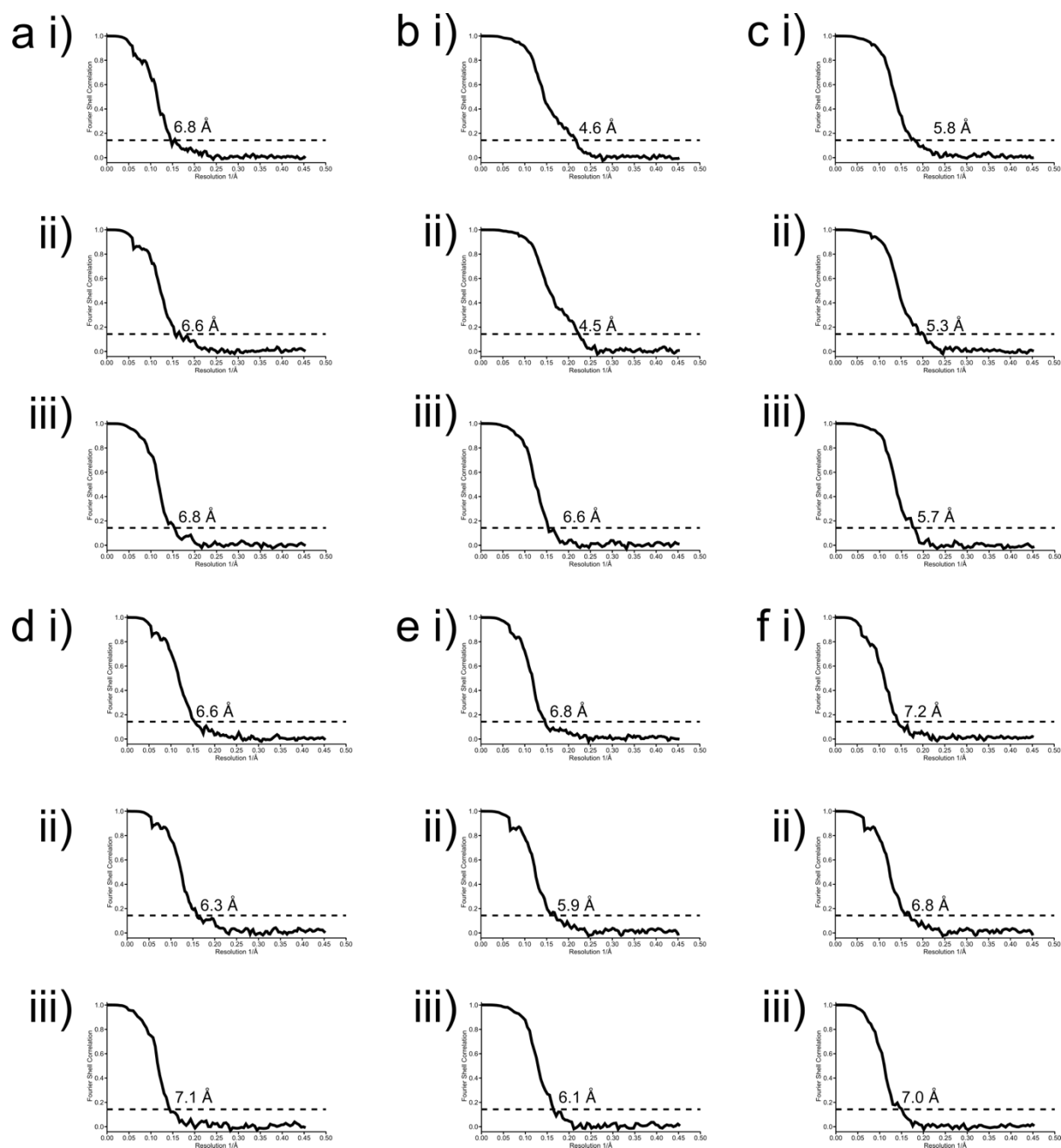


Figure 4.6: FSC curves of the global (i), RNAP- σ^{54} -DNA (ii), and PspF₁₋₂₇₅-DNA-region I (iii) focused refinement maps of conformations 1-6 (a-f).

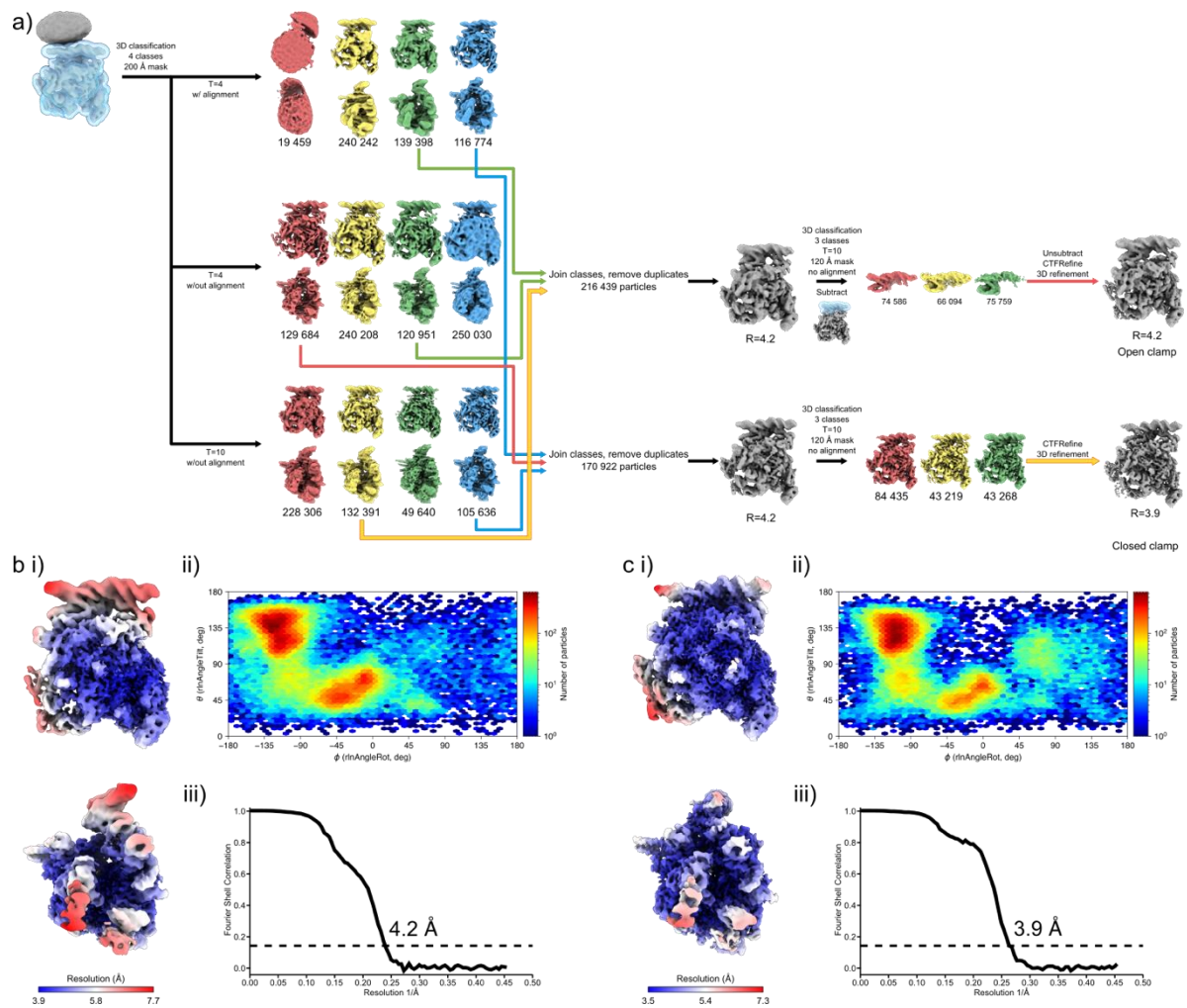


Figure 4.6.7: a) Flowchart of RNAP- σ^{54} -DNA subtraction processing pipeline. Numbers below each class

represents number of particles. b) open clamp and c) closed clamp maps showing i) local resolution map ii) angular distribution and iii) FSC curves.

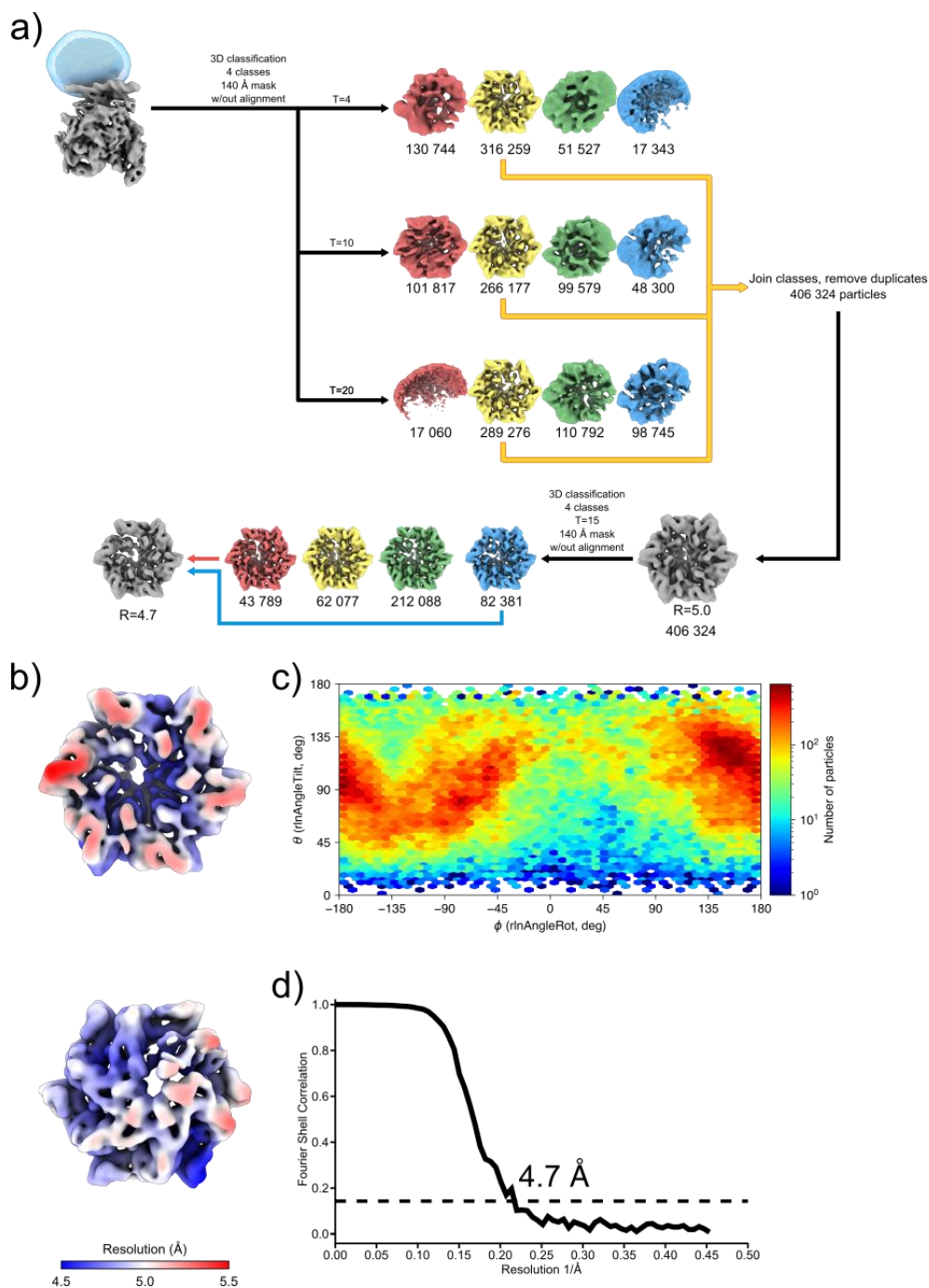


Figure 4.6.8: a) Flowchart of PspF₁₋₂₇₅ subtraction processing pipeline. Numbers below each class represents number of particles. b) local resolution map c) angular distribution and d) FSC curve of PspF₁₋₂₇₅ subtracted map.

The models were built by first docking the RPi(-12/-11) complex into the cryo-EM density maps. At first, focused maps were combined in PHENIX to produce a

composite map. Then the RNAP- σ^{54} was docked into the map using Chimera, followed by docking of PspF₁₋₂₇₅ and DNA upstream of the -10 position. The map and model fit was then adjusted manually to ensure the major and minor grooves fit into the density of DNA, and the PspF₁₋₂₇₅ seam formed by the lowest and highest protomers were aligned with the map. Docking was followed by rounds of real space refinement and manual refinement in COOT. Sidechains were not built in these models due to limitations in resolution. All model validation statistics can be found in Supplementary Table 2.

4.6.3 Structural analysis of RP(-11/-8)

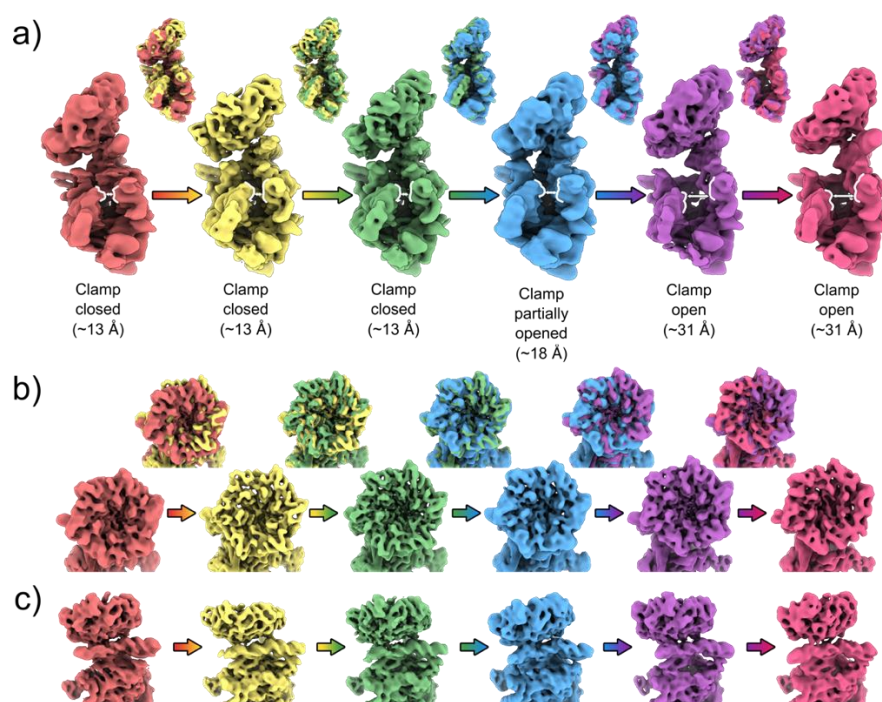


Figure 4.6.9: a) Clamp opening across different composite maps of the 6 conformations. Maps and models were aligned by the β' subunit. The outline of the inner edge of the pincers are highlighted in white, with the clamp width in brackets. The distances were measured from C α atoms between residue 372 on the β subunit and 137 on the β' subunit. b) PspF₁₋₂₇₅ ring conformations, as viewed from the top of the ring. c) PspF₁₋₂₇₅ ring conformations, as viewed from the top of the side. Each conformation has a different angle relative to the DNA.

We identified 6 different conformations, which we assign conformations 1-6 (Figure 4.6.9). The PspF₁₋₂₇₅ ring adopted different conformations and angles relative to the DNA in each conformation, suggesting that PspF₁₋₂₇₅ adopts a continuous range of motion in our datasets, limiting the resolution of the global reconstruction maps (Figure 4.6.9b, c).

Conformations 1-3 have closed RNAP clamps, similar to the RPc complex, at a width of ~13 Å. Conformation 4 is partially open, with the distance in between the pincers measuring ~18 Å. Conformations 5 and 6 are fully open, with the RNAP clamp in an intermediate state between RPc and RPip, with the clamp opening width of ~31 Å (Figure 4.6.9a).

As the sample was formed using ADP.AIFx, we cannot determine the timepoint along the reaction pathway each conformation represents and cannot determine whether clamp opening was a result of PspF₁₋₂₇₅, as ATP hydrolysis does not occur. Similarly, the lower resolution of certain conformations limits the insight into changes in interaction between the different conformations. Consequently, the bulk of the structural analysis focuses on comparisons between the PspF₁₋₂₇₅-DNA- σ^{54} region I conformations in the RP(-12/-11) complex and the highest resolution conformation in the RP(-11/-8) dataset, conformation 3.

4.6.3.1 PspF₁₋₂₇₅- σ^{54} -DNA interactions

Comparison of the DNA between the RP(-11/-8) highest resolution map (conformation 3) and the RPi (-12/-11) structures demonstrate that the PspF₁₋₂₇₅ ring is tilted at a ~52° angle towards the β subunit when viewed from the RNAP

downstream face (Figure 4.6.10a). In addition, we found that each conformation had a 29-46° bend in the DNA at the -11 position (Figure 4.6.10b).

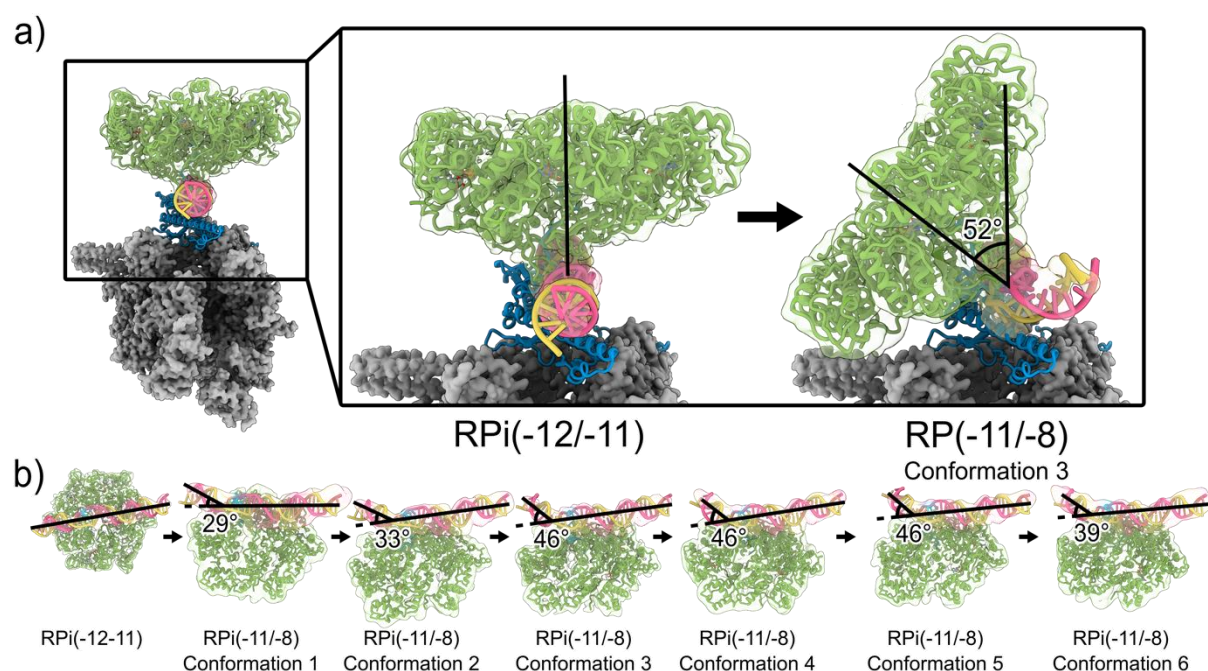


Figure 4.6.10: A comparison of the a) PspF₁₋₂₇₅ tilt angles and b) DNA bending downstream of the -10 position for RPi(-12/-11) and RP(-11/-8).

Interestingly, the tilted PspF ring now interacts with DNA at – 8/-9 instead of -12/-11, thus stabilising the enlarged transcription bubble. The protomer that engages with DNA is still at the highest position within the hexamer ring (position of protomer 1). However, by aligning the model by the PspF₁₋₂₇₅ ring, we found that the positions of the protomers, most notably the space between protomers 1 and 6, which forms the seam, is enlarged from ~19 Å in RPi(-12/-11) to 27 Å in RP(-11/-8) as we predicted (section 4.3, Figure 4.6.11a, Figure 4.6.12a-b). This coincides with the upwards movement of the L1 loop in protomer 6 (Figure 4.6.11b). Protomer 6 in RP(-11/-8) has increased scrunching of the $\alpha\beta$ sandwich and α lid subdomains compared to RP(-12/-11) and is in the *apo* state (Figure 4.6.12c,e). Furthermore, protomer 5 is more opened in the RP(-11/-8) complex and has density for

nucleotide, which is not observed in the RP(-12/-11) map, suggesting that protomer 5 is in an ADP or ADP·AlFx bound state (Figure 4.6.12c,d).

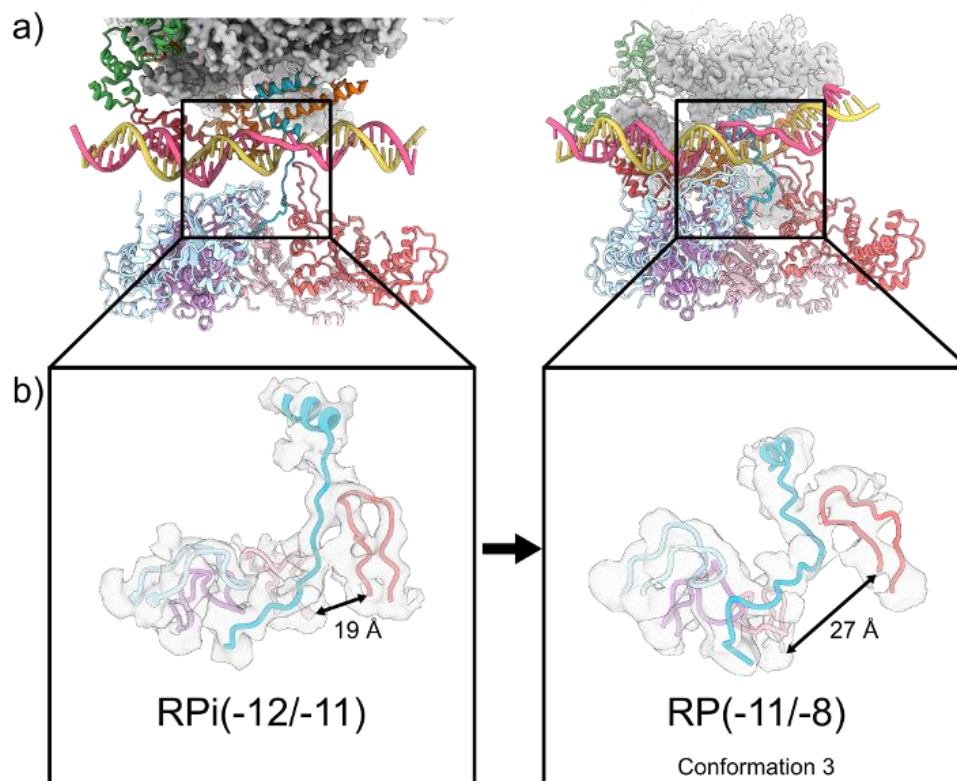


Figure 4.6.11: RPi(-12/-11) and RP(-11/-8) aligned by PspF₁₋₂₇₅ ring reveals differences in protomer arrangement. a) RPi(-11/-8) PspF₁₋₂₇₅ is tilted in comparison to RPi(-12/-11), with a larger seam between protomers 6 and 1. Protomers 2 and 3 are hidden for clarity. b) the L1 loops of protomers 4, 5, 6 and 1 are shifted in comparison to RPi(-12/-11), with the L1 loop of protomer 6 moving away from protomer 1.

The DNA is also significantly bent at a $\sim 46^\circ$ angle at the -11 position, with the density having lower resolution than the rest of the DNA, suggesting that this region is more flexible (Figure 4.6.10b). This is expected given the mismatches introduced from -11 to -8. In conformation 3, the -12T base is flipped out on the template strand similar to the RPi(-12/-11) structure; we also observe density suggesting that the base stacking in the -11G base on the template strand is disrupted, with

the -11G base flipped out to face the ELH (Figure 4.6.13). This supports the hypothesis that ELH and region I play dual roles in inhibiting DNA loading in the closed complex, as well as facilitating DNA melting and bubble propagation.

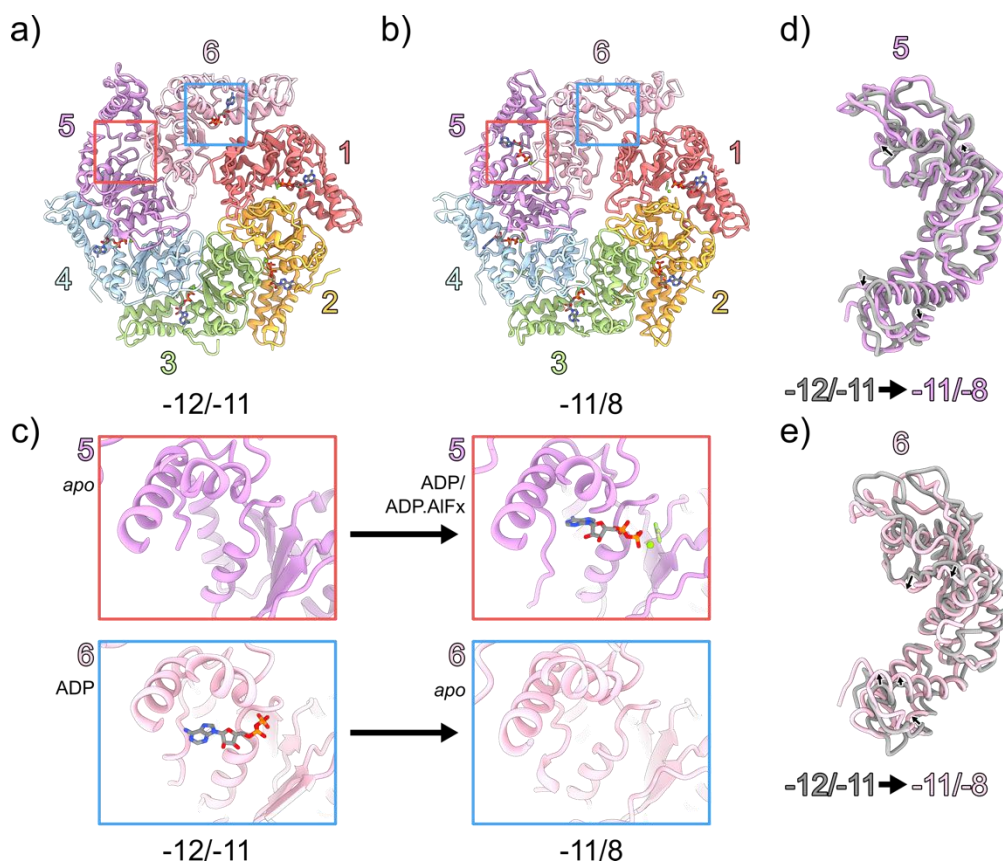


Figure 4.6.12: Nucleotide occupancy changes between protomers 5 and 6. a-b) Overview of protomer arrangement and occupancy. Zoomed in views of protomers 5 and 6 are highlighted red and blue respectively. c) Zoomed in view of the active site of protomers 5 (top panels) and 6 (bottom panels), showing protomer 5 becoming nucleotide bound, whereas protomer 6 becomes *apo*. d-e) Alignment of protomers by the walker A motif shows that from -12/-11 → -11/-8 is associated with slight opening in protomer 5 (d) and scrunching and closing of the α lid and $\alpha\beta$ sandwich subdomains in protomer 6 (e).

Interestingly, we also observed the PspF₁₋₂₇₅ ring coming into contact with the RpoN domain. Contact between the RpoN domain and protomer 4 of PspF₁₋₂₇₅ was observed in both closed and open clamp conformations, in conformations 2, 3 and

5. In all 3 instances, contacts occurred in the same region, between $\alpha 1$ helix and $\beta 2$ sheet on protomer 4 of PspF₁₋₂₇₅ and loop on the N-terminus of the RpoN domain. Although the resolution is insufficient, superimposing the subtracted PspF₁₋₂₇₅ map and the closed and open clamp RNAP- σ^{54} holoenzyme suggests that the interaction is between residues W56 on PspF₁₋₂₇₅ and G411 on the RpoN domain. This interaction is only present in the tilted ring conformation, suggesting this interaction contributes to the stabilisation of this conformation.

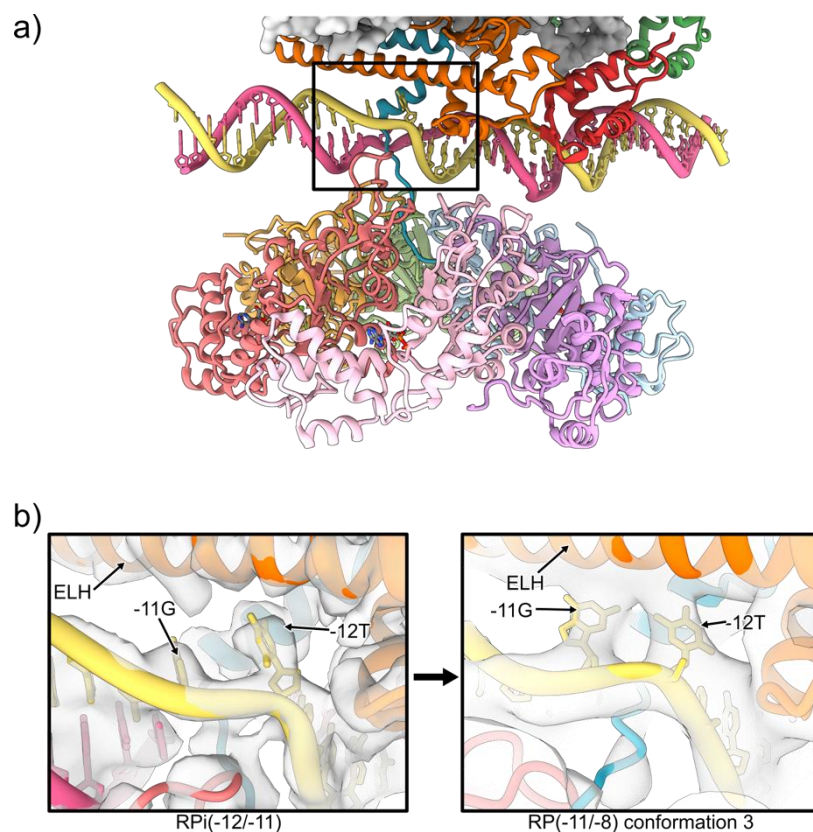


Figure 4.6.13: a) View of the RPi(-12/-11) complex from the β face. Zoomed in area in b) is highlighted. b)

In the RP(-11/-8) conformation 3 structure, the -11G base on the template strand flips out and faces the ELH.

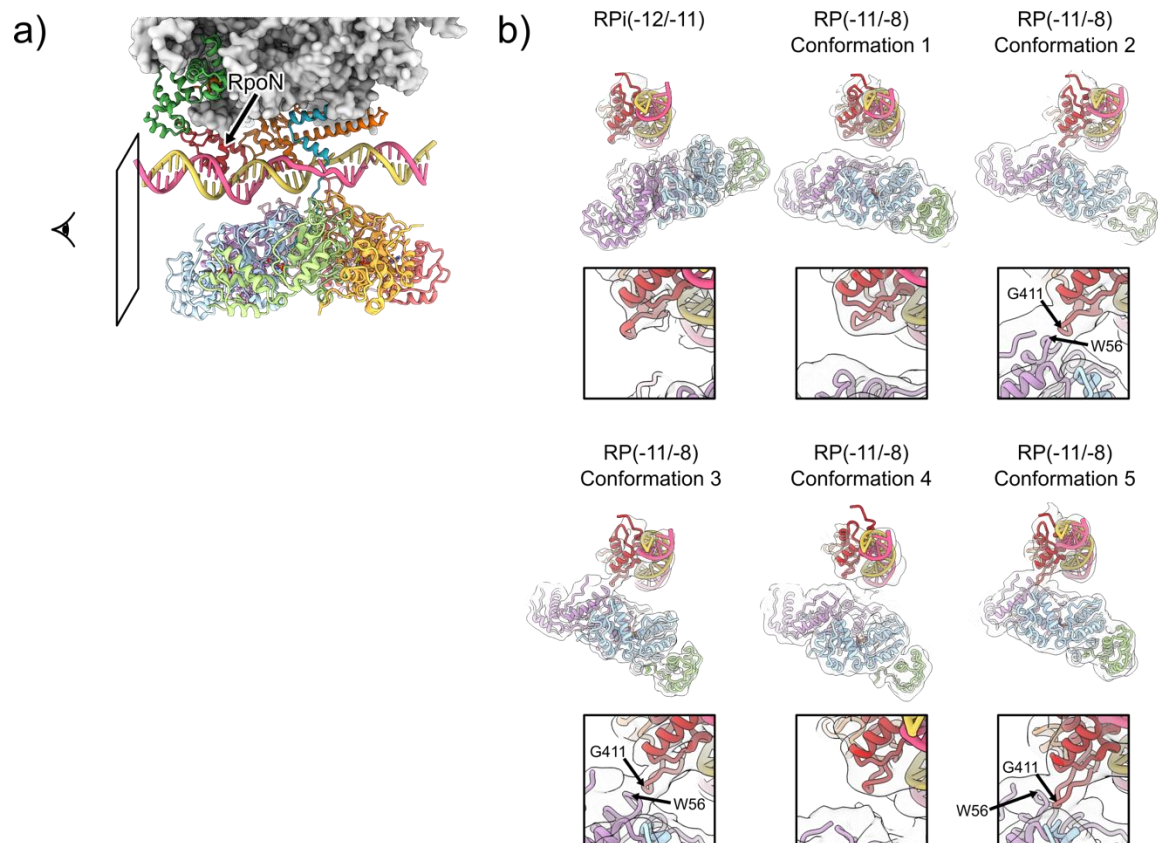


Figure 4.6.14: Overview of potential RpoN and protomer 4 interactions in the RP(-11/-8) dataset. a)

Zoomed out view of RP(-12/-11).b) Density and models of the RpoN domain interactions

with DNA and PspF₁₋₂₇₅. Bottom panels are zoomed in views of protomer 4 and RpoN.

The core binding domain and ELH are hidden for clarity.

We could not find RPip- and RPo-like states within the dataset. As a result, we could not verify the order of the conformations within a reaction pathway nor confirm whether these intermediate states represented on-pathway conformational states. We also collected data for RP(-11/-3) and RP(-11/-5), however, these dataset exhibited large amounts of complex dissociation and our 3D reconstructions demonstrated severe preferential orientation, that hampered the resolution and further processing and identification of new intermediates.

Consequently, we performed a screen of different detergents, and subsequently determined the structure of a RP(-10/-1) intermediate, which has been proven to be transcriptionally active, to verify if these conformations were on pathway.

4.7 RP(-10/-1) complex

4.7.1 Sample preparation

We found that CHAPSO, a zwitterionic detergent, along with a 10× higher protein concentration enabled us to obtain sufficient particle concentration and distribution with a significant improvement of different views in our reconstructions, thus overcoming preferential orientation issues. As a result of the higher concentration required, and gel filtration resulting in a complex dissociation, we opted to skip the gel filtration step and apply mixed sample directly onto grids.

4.7.2 Image processing of RP(-10/-1)

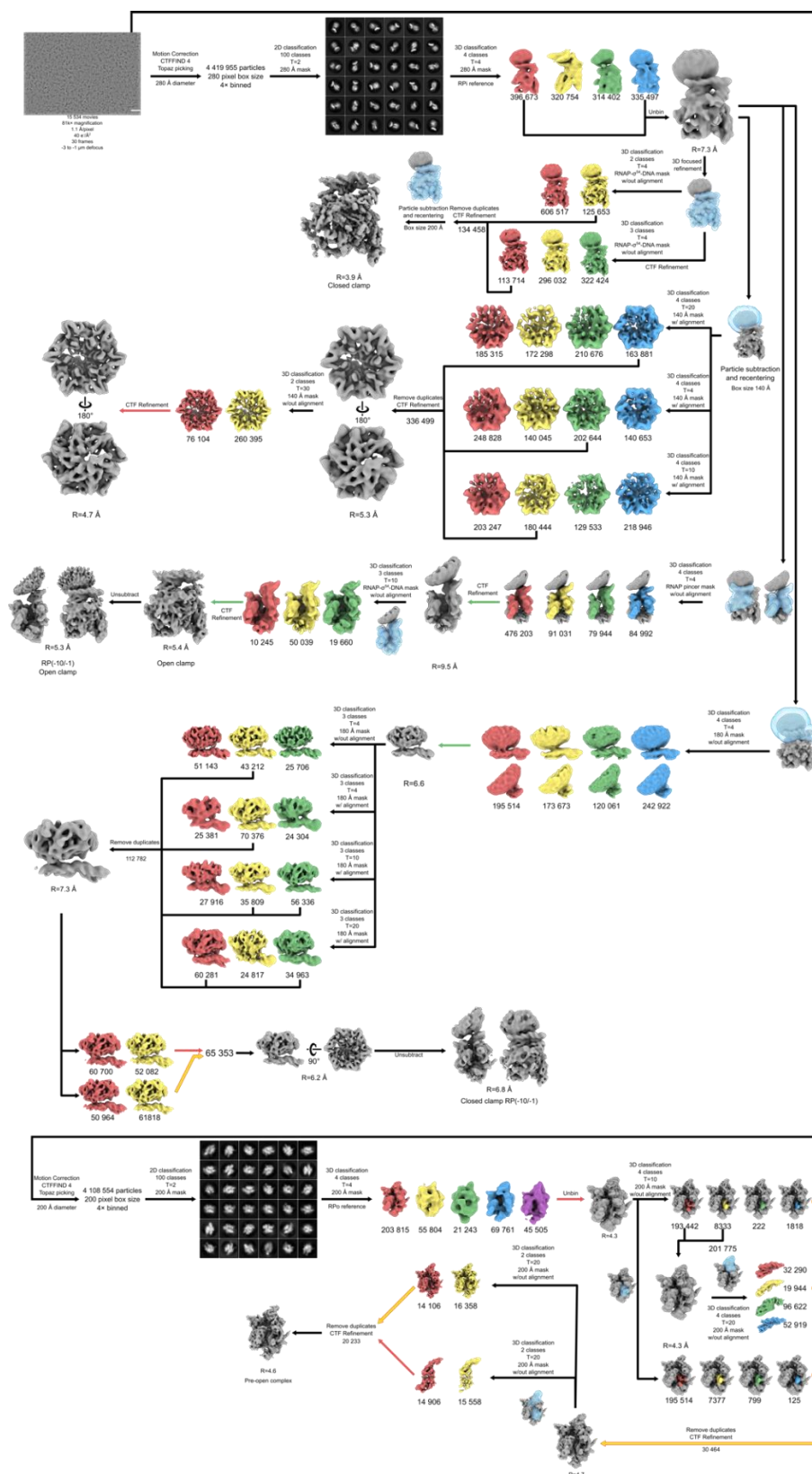


Figure 4.7.1: Overall processing workflow for RP(-10/-1) dataset. Arrow colour indicates class selected.

The imaging processing workflow is shown in Figure 4.7.1. We found a good distribution of views from the initial angular distribution map (Figure 4.7.2, Figure 4.7.3ii). We performed an identical processing strategy to the RP(-11/-8) complex dataset. Unlike the previous datasets, we found only 1 conformation of PspF₁₋₂₇₅ - DNA which corresponded to a closed clamp conformation. We did find an open clamp conformation by from un-subtracting the open clamp subtracted refinement; however, the PspF₁₋₂₇₅ density was much weaker and lower resolution (8.7-13.7 Å) in comparison to the RP(-11/-8) maps.

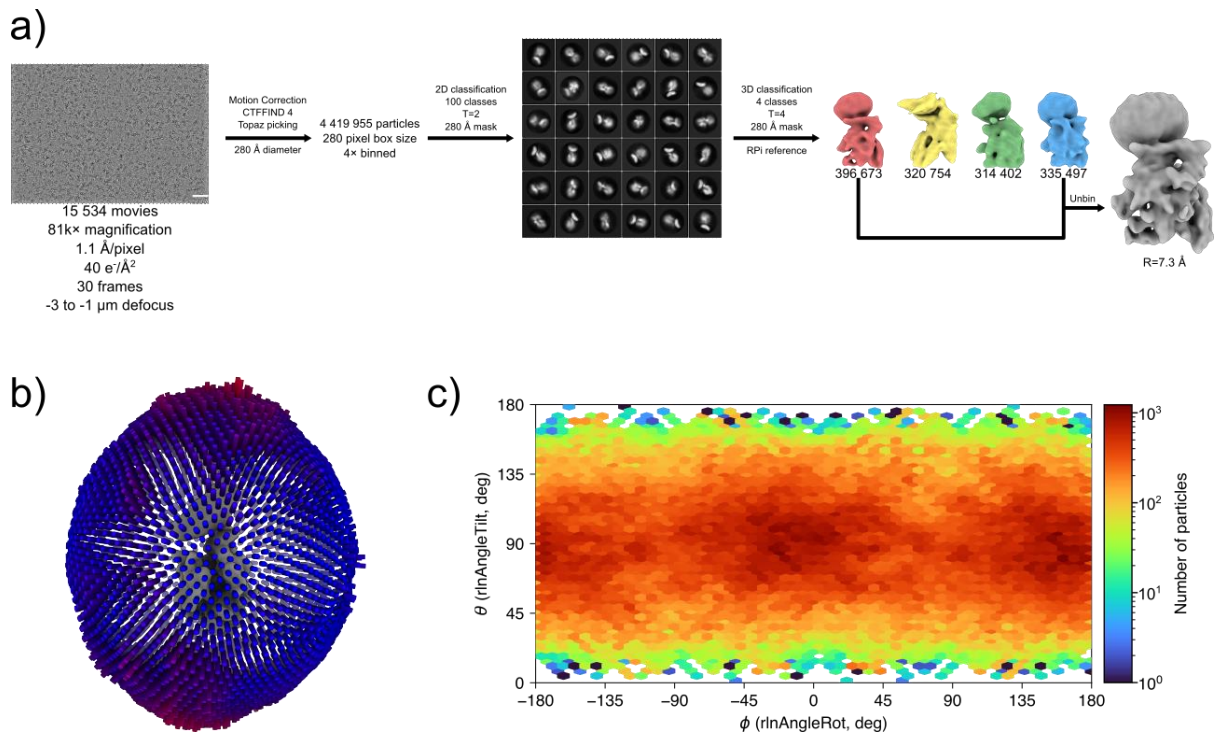


Figure 4.7.2: a) Pre-processing and 1st round of classification and refinement of the RP(-10/-1) dataset showed features of RNAP, σ^{54} , DNA and PspF₁₋₂₇₅. The first refinement showed preferential orientation, as indicated by the spherical (b) and flattened (c) angular distribution map of the first refinement.

As expected from our *in vitro* transcription assay (Figure 4.5.2b), we found RPO- like states inside the dataset using the general model in the TOPAZ particle-picker without training. Subtraction of the RNAP- σ^{54} -DNA and PspF₁₋₂₇₅ ring resulted in a 3.9 and 4.7 Å reconstructions, allowing unambiguous assignment of secondary structure and tracing of the C α backbone densities for both RNAP- σ^{54} and PspF₁₋₂₇₅ protomers (Figure 4.7.3b-c).

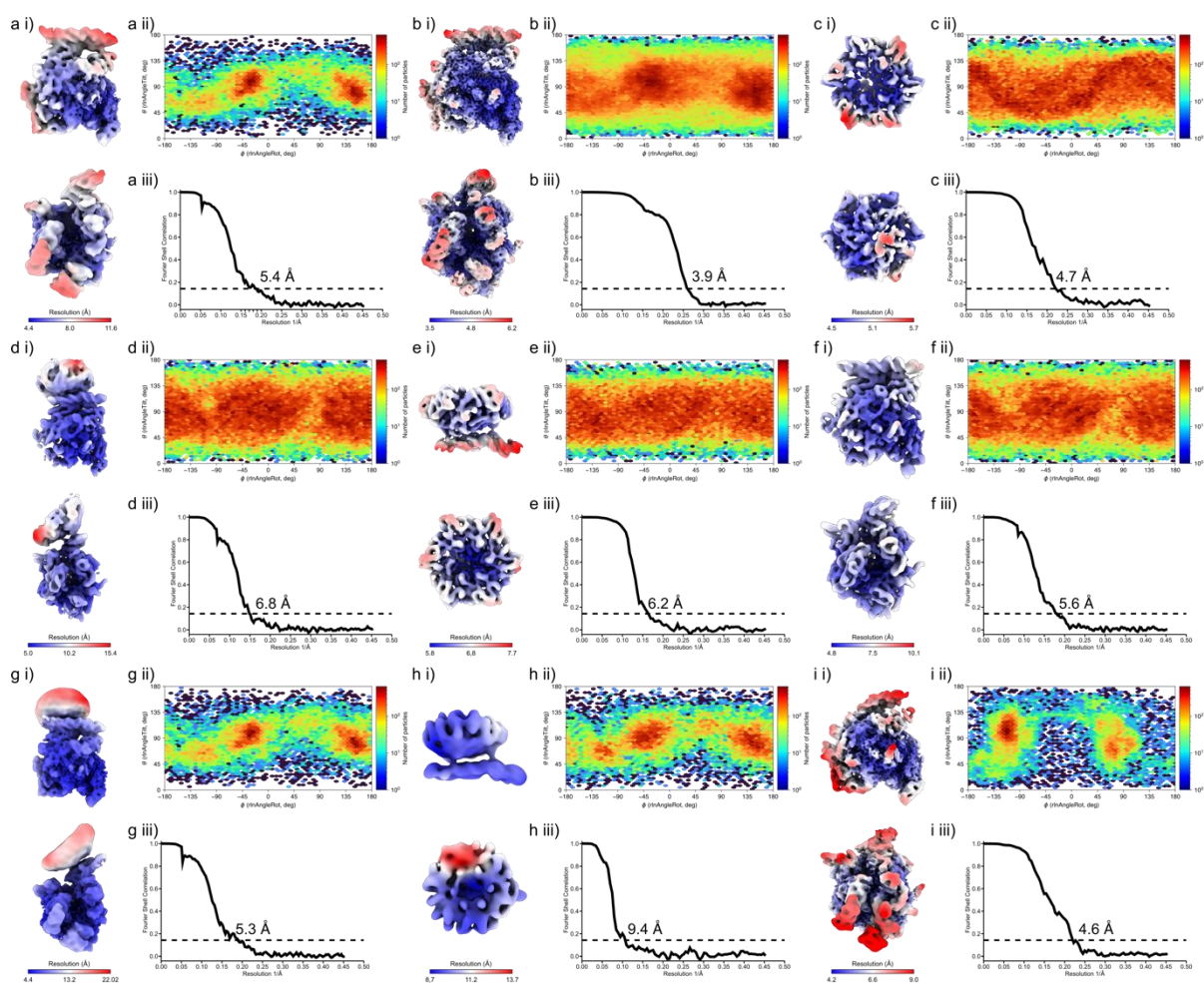


Figure 4.7.3: i) Local resolution maps ii) angular distribution maps and iii) Fourier-shell correlation curves of a) open clamp conformation. b) closed clamp conformation c) PspF₁₋₂₇₅ ring d-f) Conformation 1 and g-i) conformation 2 of the RP(-10/-1) complex.

4.7.3 Structural analysis of RP(-10/-1)

Much like the RP(-11/-8) structures, the PspF₁₋₂₇₅ ring exhibits a tilted conformation, with the ring tilting 52° towards to β subunit (Figure 4.7.4a) and the trajectory of the DNA being bent downstream of the -10 position, at a more acute angle of 73° (Figure 4.7.4b). We found only 1 conformation of PspF₁₋₂₇₅-DNA in the dataset, suggesting that the intermediate states in RP(-11/-8) were transient melted intermediates that could not be trapped using the preopened DNA bubble. Similar to RP(-11/-8) conformation 3, we found that the RpoN domain was in a raised conformation and formed contacts with the same regions of protomer 5 of the PspF₁₋₂₇₅ ring.

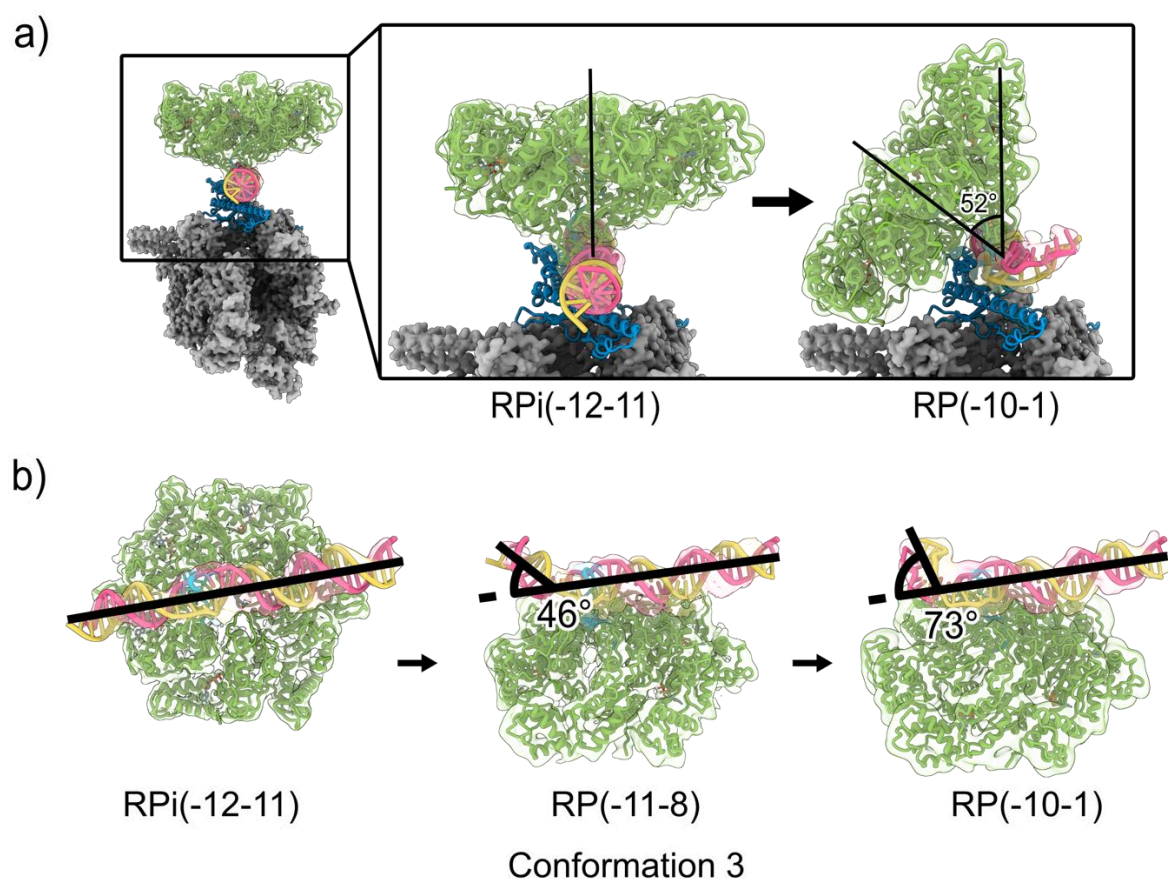


Figure 4.7.4: a) The PspF₁₋₂₇₅ tilt for RP(-10/-1) conformation is the same angle as RP(-11/-8) conformation 3. b) the DNA bending angle of the RP(-10/-1) is 73°, a sharper bend corresponding to a larger mismatch length and bubble size.

However, the arrangement of protomers in the PspF₁₋₂₇₅ ring in the -10/-1 is different from RPi(-12/-11) and RP(-11/-8). Most notably, the position of the protomers within the ring is shifted 1 unit anticlockwise (Figure 4.7.5a). Region I N-terminus is shifted as protomer 1 now occupies the same vertical position along central pore axis to protomer 6 and protomer 2 moves to the same position as protomer 1 and closest to the DNA bubble and region I (Supplementary Movie 4-6). The resolution makes it difficult to determine the exact number of residues shifted, however, the backbone trace suggests that the N-terminus has been translocated by at least 1 residue (Figure 4.7.5a, b). This is consistent with the power stroke hypothesised Figure 4.3.3f, and strongly supports the mechanism where ATP-hydrolysis by PspF₁₋₂₇₅ translocates the L1 loops within the hexameric ring, pulls the N-terminus of region I through the central pore, although the resolution is insufficient to determine if region I α 1 is unfolded (Figure 4.7.5b and c).

The resolution for the DNA downstream of -10 (Figure 4.7.3e, i) is lower in comparison to maps in -11/-8 conformations 2 and 3 (Figure 4.6.4b), making it difficult to determine base flipping in RP(-10/-1) the loss of resolution suggests that this region is more flexible due to the loss of base pairing in the pre-opened bubble.

The resolution of RP(-11/-8) and RP(-10/-1) conformations are too low to accurately determine the nucleotide occupancy of each protomer, there is no nucleotide density for protomer 6 RP(-11/-8) conformation 3 and protomer 1 in RP(-10/-1), and density for nucleotide in protomer 5 in RP(-10/-1).

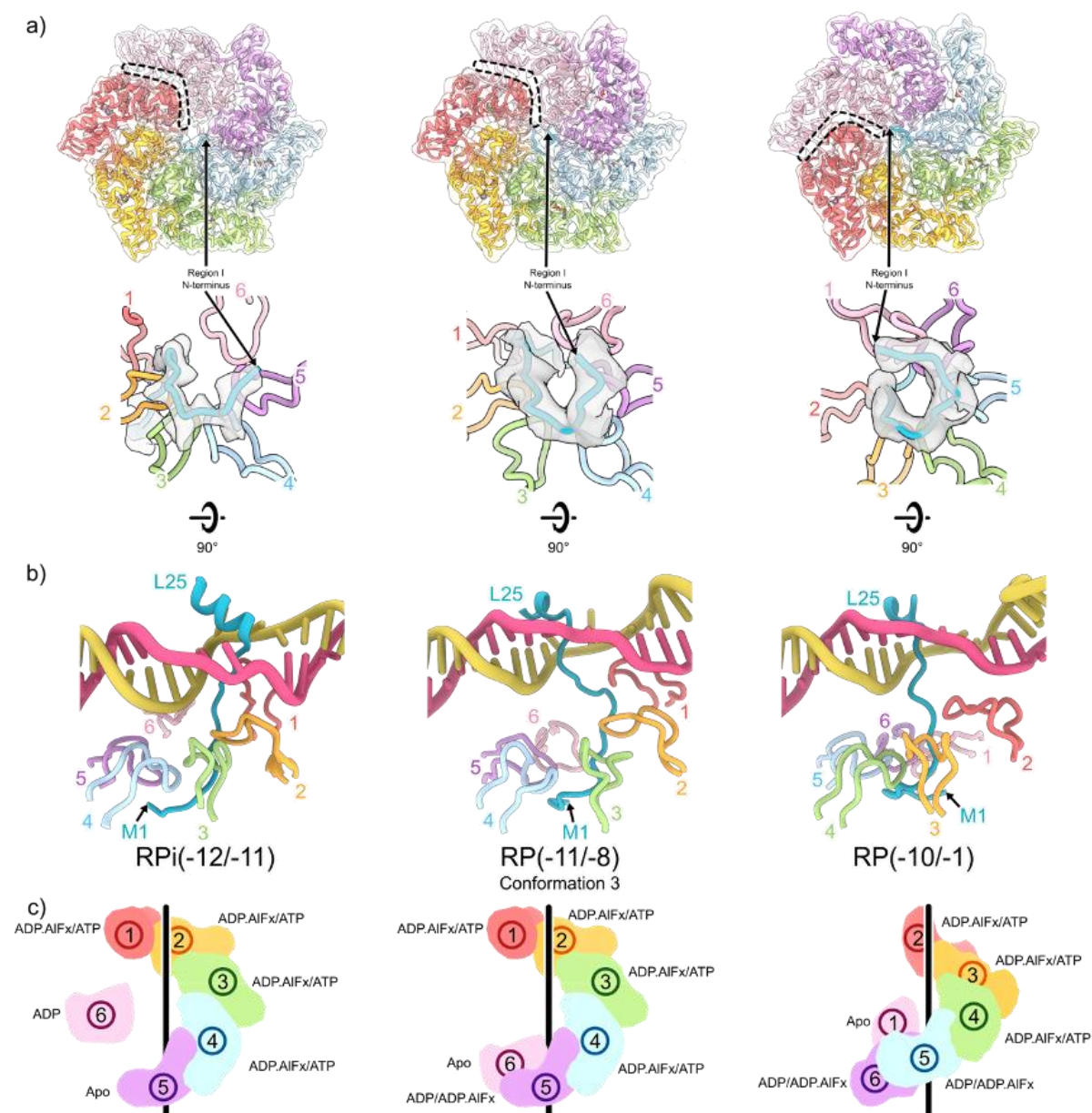


Figure 4.7.5: a) DNA melting corresponds with changes in the region I N-terminus position, which suggests a pulling and potential unfolding of the $\alpha 1$ helix (b) and protomer arrangement of the PspF₁₋₂₇₅ ring (c). a) Apo interface highlighted by dotted lines. Note that the colours of the protomer indicate its position along the central axis, shown by a vertical line in (c).

The RPo-like state was similar to the previously published RPo model (Figure 4.7.6a); however, the trajectory of the upstream DNA was slightly altered, and the density for ELH was disrupted, similar to the RPip map and model, suggesting that the ELH is partially unfolded, or in a “retracted” state. This conformation may

represent a pre-RPo conformation, that occurs before the ELH inserts in between the DNA strands, as seen in the RPo structure (Figure 4.7.6b, c). Surprisingly, we could not find any RPip like complexes within our dataset, suggesting that the process of DNA loading and clamp opening occurs too transiently to be captured using the WT σ^{54} without time-resolved approaches. Alternatively, RPip might be an off-pathway or artificial state that is caused by the R336A mutation or lack of initiating rNTPs to stabilise the RPo conformation.

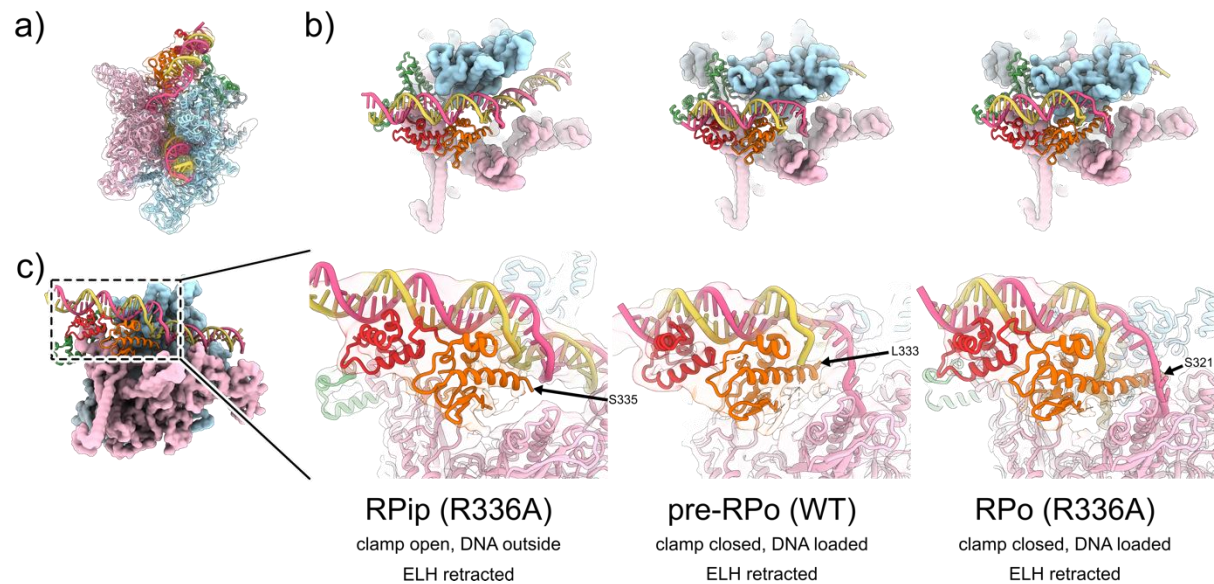


Figure 4.7.6: Comparison of the RPo(WT) structure with the previously published RPip(R336A) and RPo(R336A) mutant structures. a) The map and model of the RPo(WT) structure. b) Comparison of the upstream DNA path and clamp opening in the different structures. View looking at the upstream DNA from the top of the RNAP clamp. c) The ELH position and upstream DNA path in the different conformations.

4.8 Summary

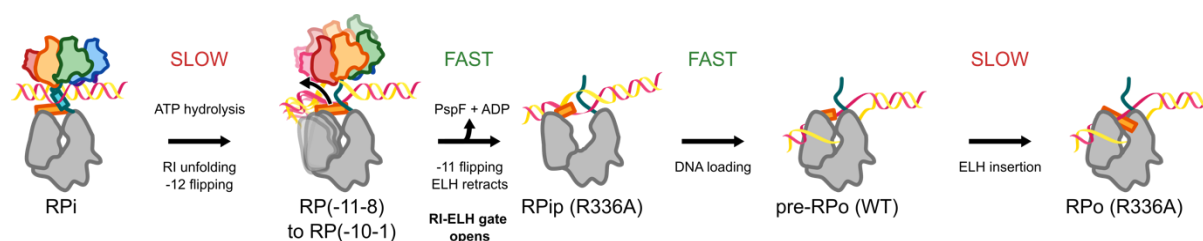


Figure 4.8.1: Schematic of open complex formation

In summary, these series of structures suggests that PspF₁₋₂₇₅ binds and pulls region I through its central pore to facilitate DNA melting, as demonstrated by the altered protomer arrangement in RP(-10/-1) and the movement of protomer 6 in RP(-11/-8) in comparison to RPi(-12/-11) and protomer 1. The translocation of the N-terminus of region I could lead to unfolding the α 1 helix, resulting in opening of the gate formed between region I and ELH, causing the ELH to retract. The melting of the DNA coincides with PspF₁₋₂₇₅ ring tilting and bending of the DNA downstream of the -10 region, which brings the RpoN into close proximity to PspF₁₋₂₇₅ leading potential interaction between protomers in position 4 and a loop on the RpoN domain. Both RP(-11/-8) and RP(-10/-1) datasets contain closed and open clamp conformations, with the RP(-10/-1) dataset containing weak PspF₁₋₂₇₅-DNA density in the open clamp conformation. The RP(-10/-1) dataset also contains a pre-RP_o conformation, but no RPip conformation, which could suggest that DNA loading and PspF₁₋₂₇₅ dissociation is a rapid process. Upon loading and formation of pre-RP_o, the ELH then inserts in between the two DNA strands to stabilise the transcription bubble as seen in the RP_o and RPitc structures.

Chapter 5

Structural basis of σ^{54} displacement in Bacterial Promoter Escape

5.1 Introduction

Transcription initiation involves promoter search, open complex formation, initial transcription and promoter escape (Herbert, Greenleaf & Block, 2008). Promoter escape is the final step in bacterial transcription initiation and describes the process in which RNA polymerase (RNAP) is released from the promoter region and translocates downstream to transcribe the gene. σ displacement is essential in enabling RNAP to be released from the promoter and for processive RNA synthesis during elongation (Winkelman, Nickels & Ebright, 2021).

Previous studies have determined the mechanisms of σ displacement in σ^{70} , as either arising from RNA extension causing the 5' end of the RNA to push against the σ finger (Basu et al., 2014; Li et al., 2020), or DNA scrunching as a result of an enlarged transcription bubble (Kapanidis et al., 2006; Revyakin et al., 2006; Zuo et al., 2020). However, the mechanisms of σ^{54} displacement remain unclear. Based on our previous holoenzyme, RPo and RPitc structures, it is clear that region II and the core binding domain (CBD) must be displaced for RNA to clear the exit channel, and the upstream DNA binding domains extra-long helix-helix-turn-helix (ELH-HTH)

and RpoN must be dissociated for RNAP to progress downstream. In this study, we used single particle cryo-EM to determine the mechanisms of σ^{54} promoter escape.

5.2 RP5-9nt complexes

To determine the mechanism of σ^{54} displacement during promoter escape we designed DNA-RNA scaffolds with increasing transcription bubble and RNA lengths (Figure 5.2.2). We then sought out to determine the cryo-EM structures to determine the mechanism σ^{54} displacement.

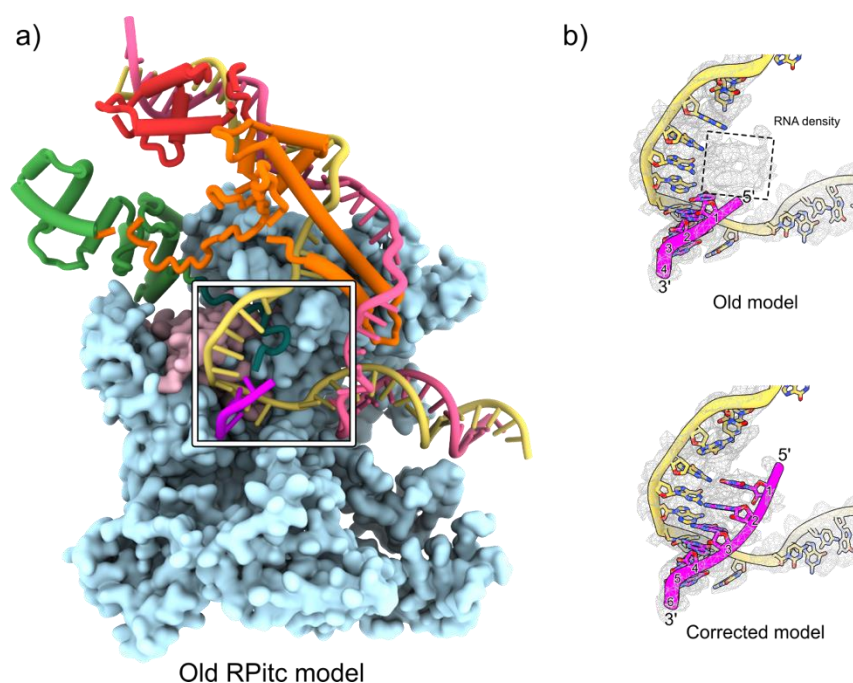


Figure 5.2.1: Re-assignment of RNA density in RPitc model. a) Overview of the cross-section old RPitc model. Zoomed in region is highlighted by white box. b) Zoomed in active site, showing only DNA template strand and RNA for clarity. Upon analysis of the RPitc density map, we found density for two additional nucleotides (5'-UpGGGGG-3'), despite the DNA scaffold only providing no complementary bases in on the template strand for the additional nucleotides, suggesting that σ^{54} initial transcription undergoes slippage or reiterative transcription.

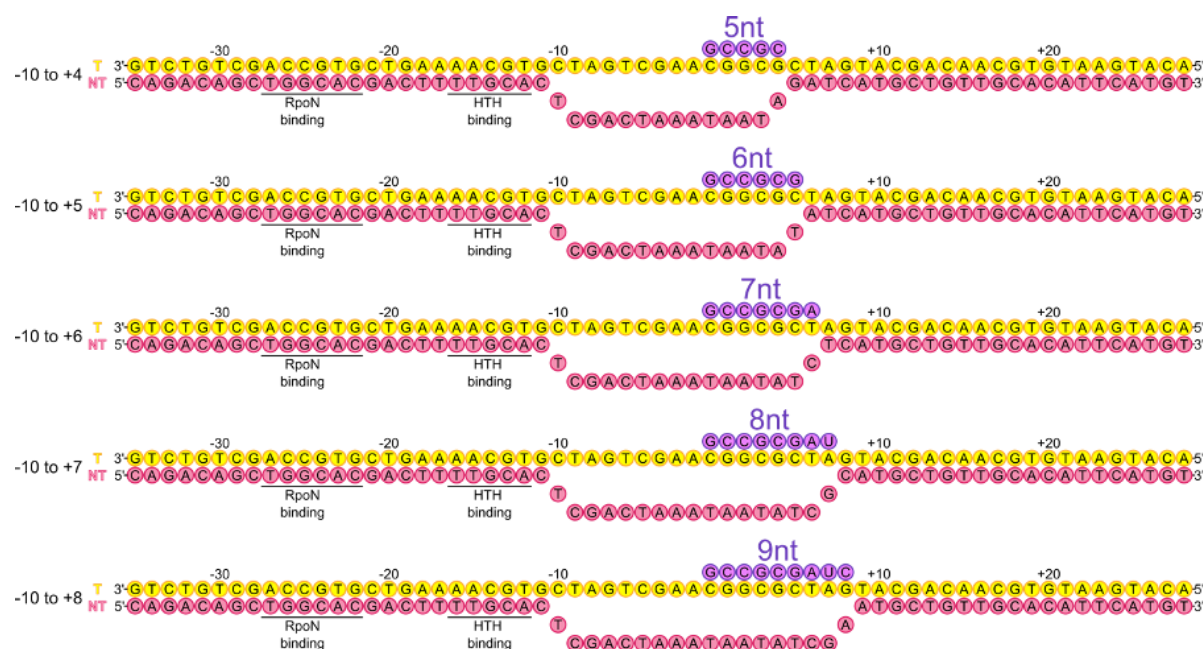


Figure 5.2.2: Design of the DNA-RNA scaffolds used in this investigation

We initially prepared initially transcribing complexes by incubating with specific rNTPs to produce *de novo* synthesised RNA; however, we found that complexes showed a large amount of RNA heterogeneity, and the active site suffered from mixed occupancy and different RNA lengths, preventing high resolution structure determination or RNA-bound complexes. Even in the previously published RPitc structure, which was designed to synthesise UpGGG, upon reassessment of the map, we found that it has density for an additional 2 bases, demonstrating that σ^{54} undergoes reiterative transcription or slippage (Figure 5.2.1a, b). To improve complex homogeneity, we designed scaffolds with pre-synthesised RNA (Figure 5.2.2) and determined 6 structures of RNA- and σ^{54} -bound initial transcription complexes.

5.2.1 5nt complexes reveal a novel passage of region II through the RNAP core

RP5nt

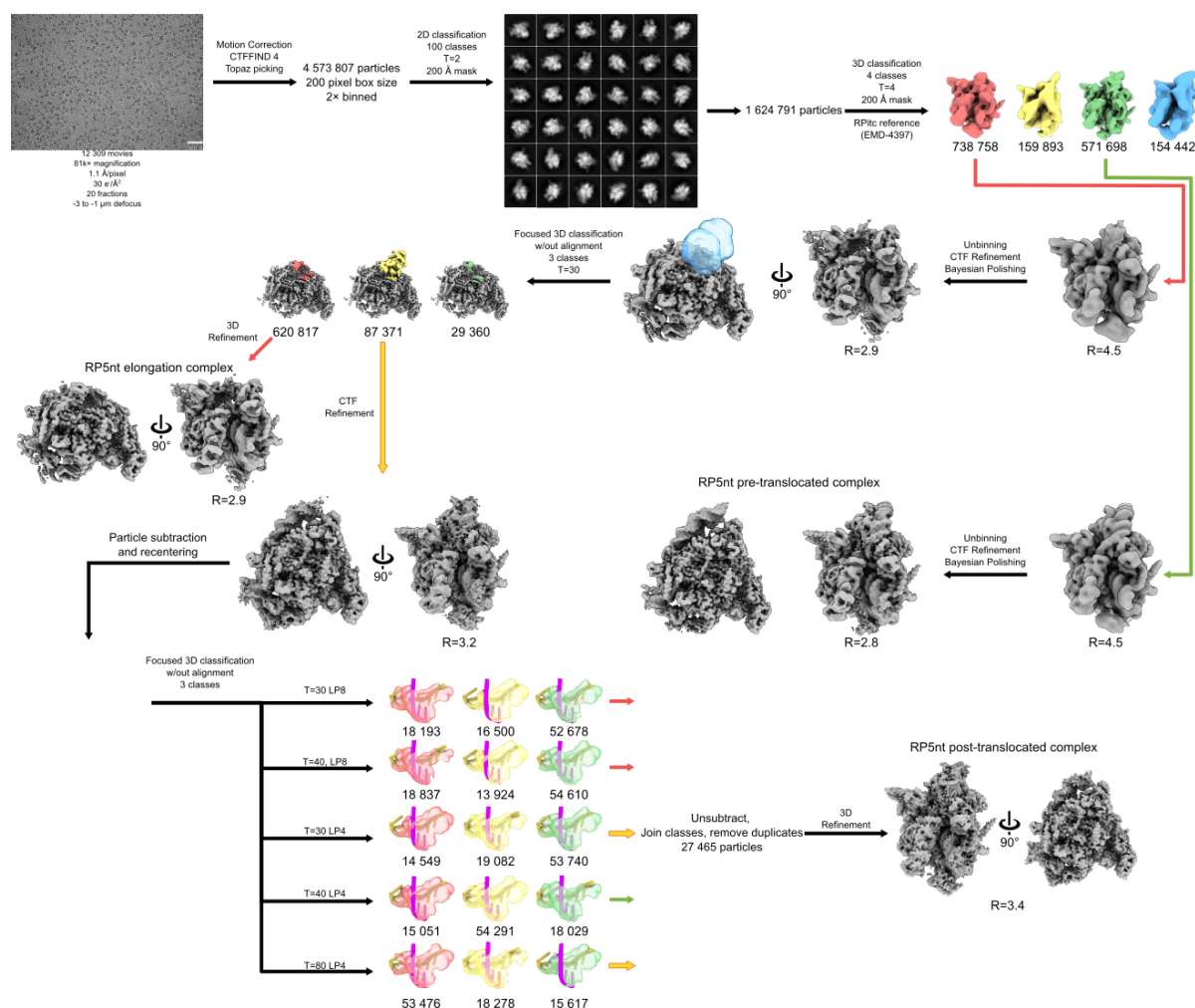


Figure 5.2.3: Processing pipeline for the RP5nt dataset

In the 5nt dataset, we used a combination focused classification, subtraction and recentring to determine a pre-, post- translocated σ^{54} -bound states, as well as a elongation states (Figure 5.2.3). The pre- and post- translocated states were resolved to a resolution of 2.8 and 3.4 Å respectively (Figure 5.2.5a, Figure 5.2.6).

Compared to RPitc, the σ^{54} and bridge helix, rudder and fork loop regions remained the same except for region II (RII). σ^{54} RII connects region I (RI) and the core binding domain (CBD) of the region III (RIII) that sits and blocks the RNA exit channel (Figure

5.2.4a). Much like the previous R336A σ^{54} mutants, we could only determine parts of region II (Figure 5.2.4a). For RPo, we could resolve up to D106, and in the revised model for RPitc, we could resolve D107. Residues C-terminal of Y112 are in the same positions but N-terminal residues are highly flexible, as shown by the different trajectories and lower resolution of C α backbone between residues 106-111 (Figure 5.2.4b, c).

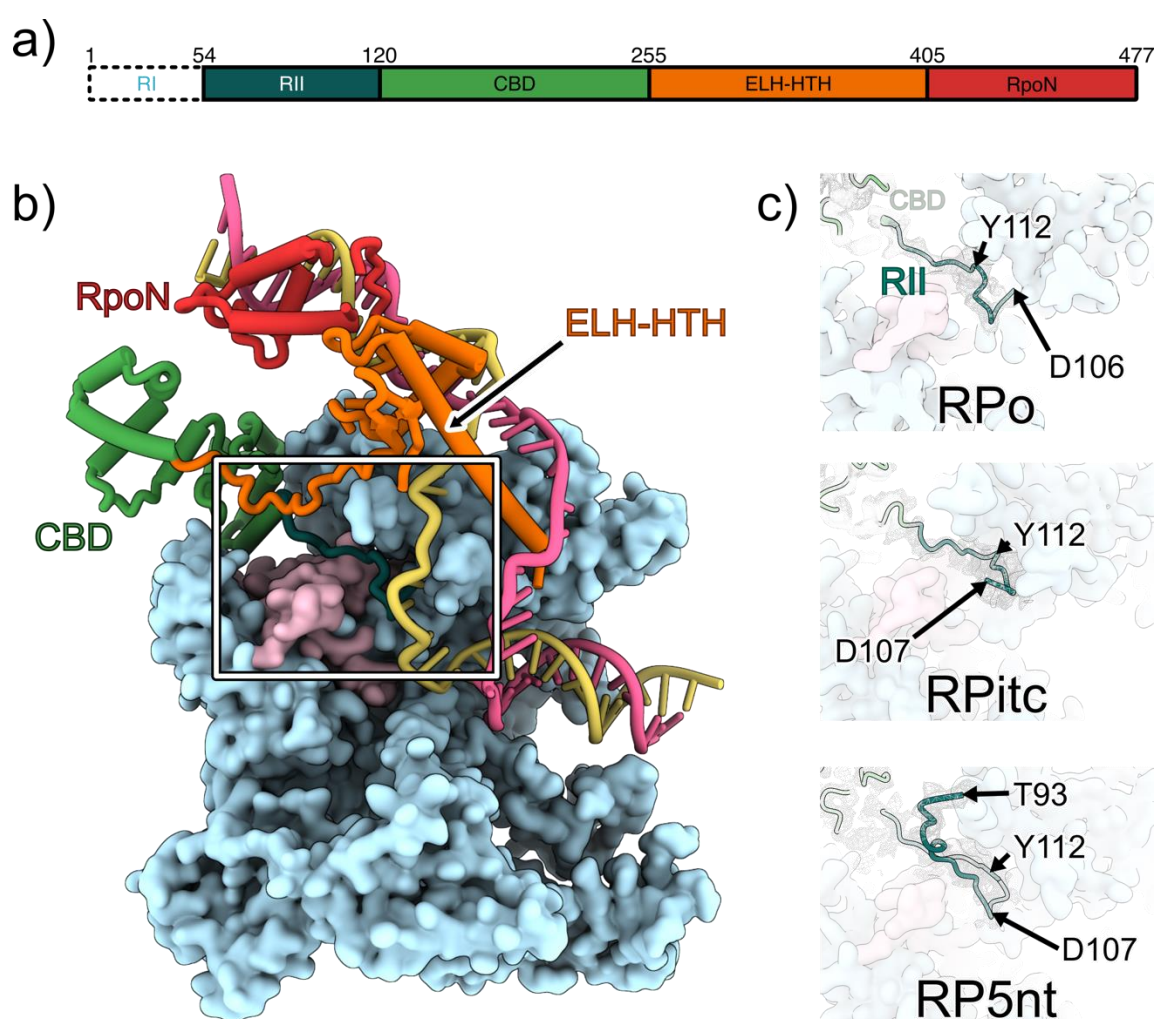


Figure 5.2.4: Comparison of region II in RPo, RPitc and RP5nt (pre-translocated) complexes. a) Overview of the σ^{54} domain structure. RI is shown in white, as it is unresolved. b) Cross section of the RPo complex. Zoomed in region highlighted by white box. c) Cartoon and density of region II in each structure

In the RP5nt σ^{54} -bound structures, we could resolve up to T93, enabling us to trace the path of the region II C α backbone as it exits the RNA cleft. Region II is in a constrained conformation and enters the catalytic site in a narrow tunnel formed by the β -flap, β i9 coiled coil, β -protrusion and an antiparallel β -sheet on extra-long-helix helix-turn-helix (ELH-HTH). It then contacts the template strand at the -5 position, and threads around the template DNA-RNA hybrid and forms a sharp kink at P110, before exiting via RNA exit channel and connecting to the CBD (Figure 5.2.7a-b). We refer to this kink (residues 104-113) as the RII-finger due to its similarity with the σ^{70} finger, also known as R3.2 .

The RII kink is stabilised by a hydrophobic interactions formed between P110 and Y112 on RII, P251, V253 and P254 on the β' lid and M330 on β' switch 2 region (Figure 5.2.7c).

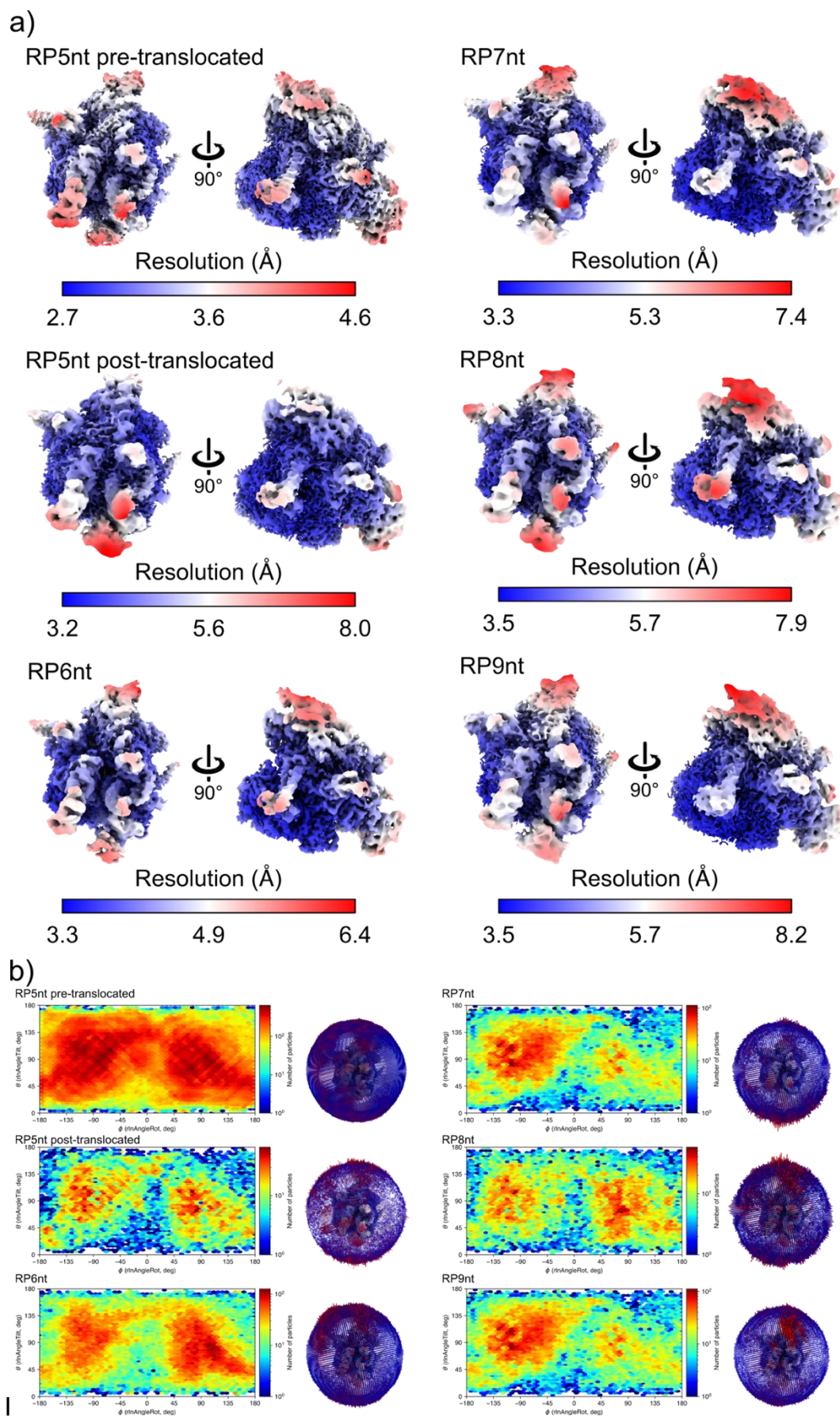


Figure 5.2.5: a) Local resolution maps. b) Angular distribution maps.

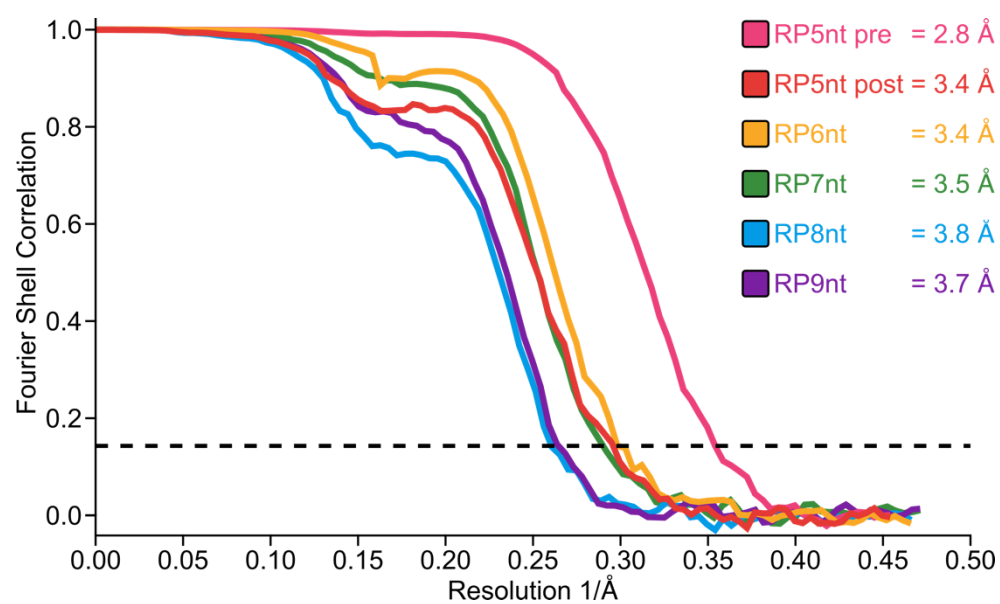


Figure 5.2.6: FSC maps.

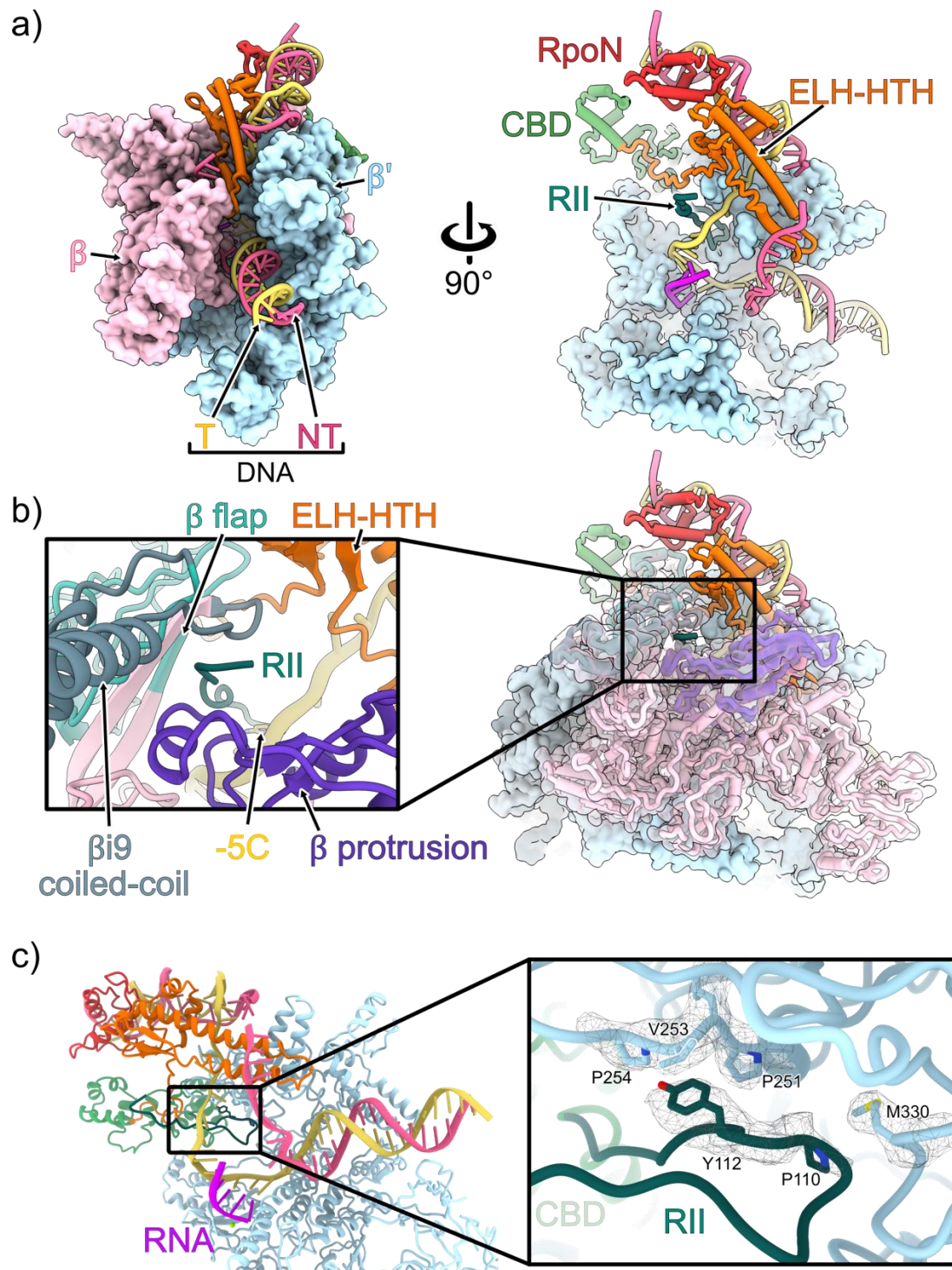


Figure 5.2.7: a) Left: overview of RP5nt structure downstream face. Right: cross-section of RP5nt, with RII highlighted, and α and β subunits hidden. b) RII entrance into the RNAP core. c) Left panel: cut through of the β' face of the 5nt pre-translocated elongation complex. β is not shown for clarity. Right panel: RII finger forms Van der Waals interactions with β' that stabilise the kinked structure.

5.2.2 Longer transcripts causes scrunching of the region II finger

RP6nt

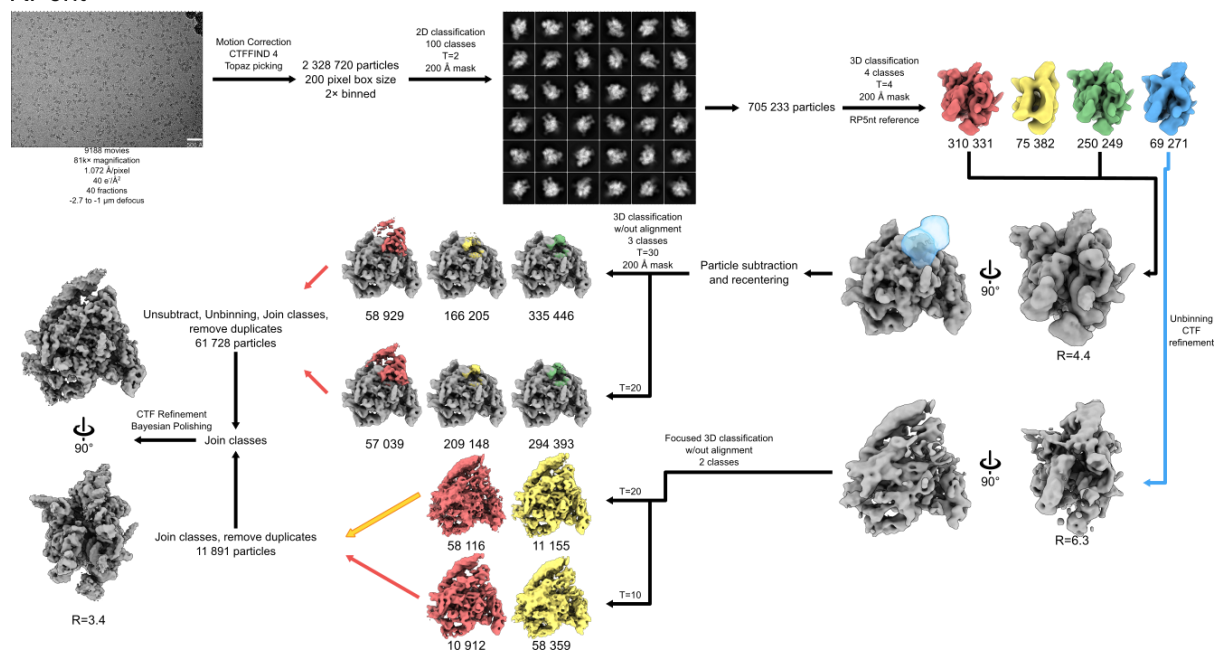


Figure 5.2.8: Processing pipeline for the RP6nt dataset

RP7nt

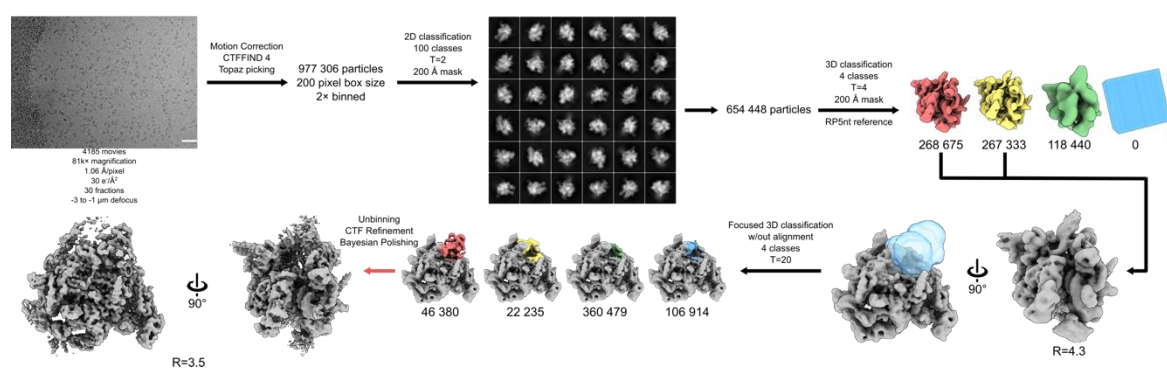


Figure 5.2.9: Processing pipeline for the RP7nt dataset

RP8nt

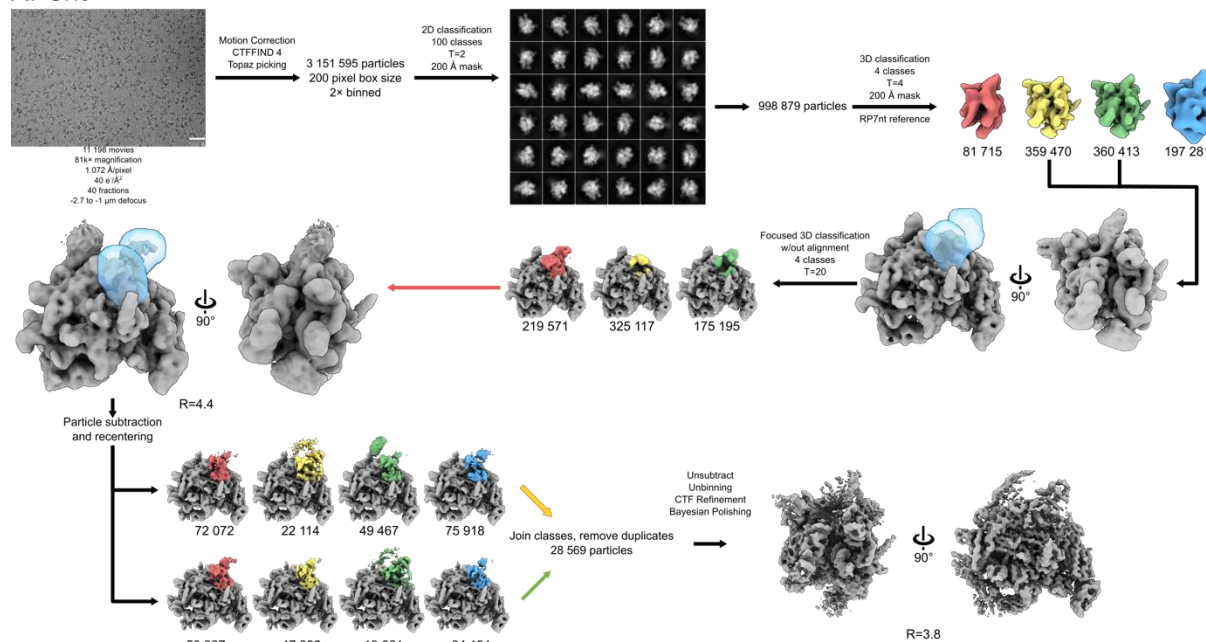


Figure 5.2.10: Processing pipeline for the RP8nt dataset.

RP9nt

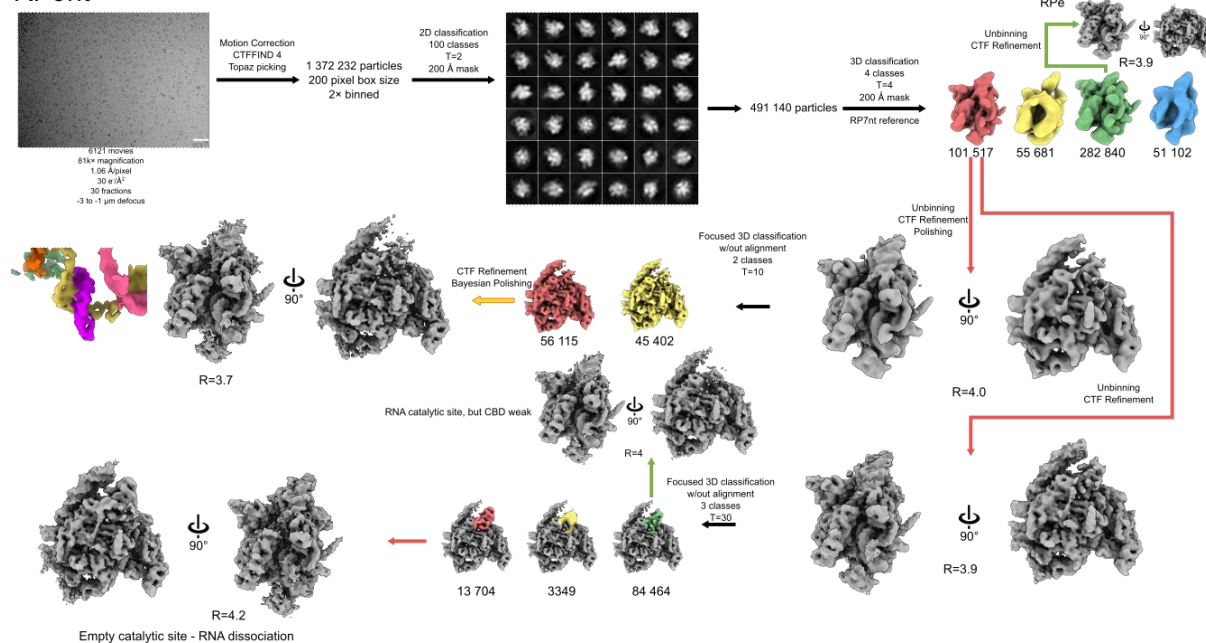


Figure 5.2.11: Processing pipeline for RP9nt dataset

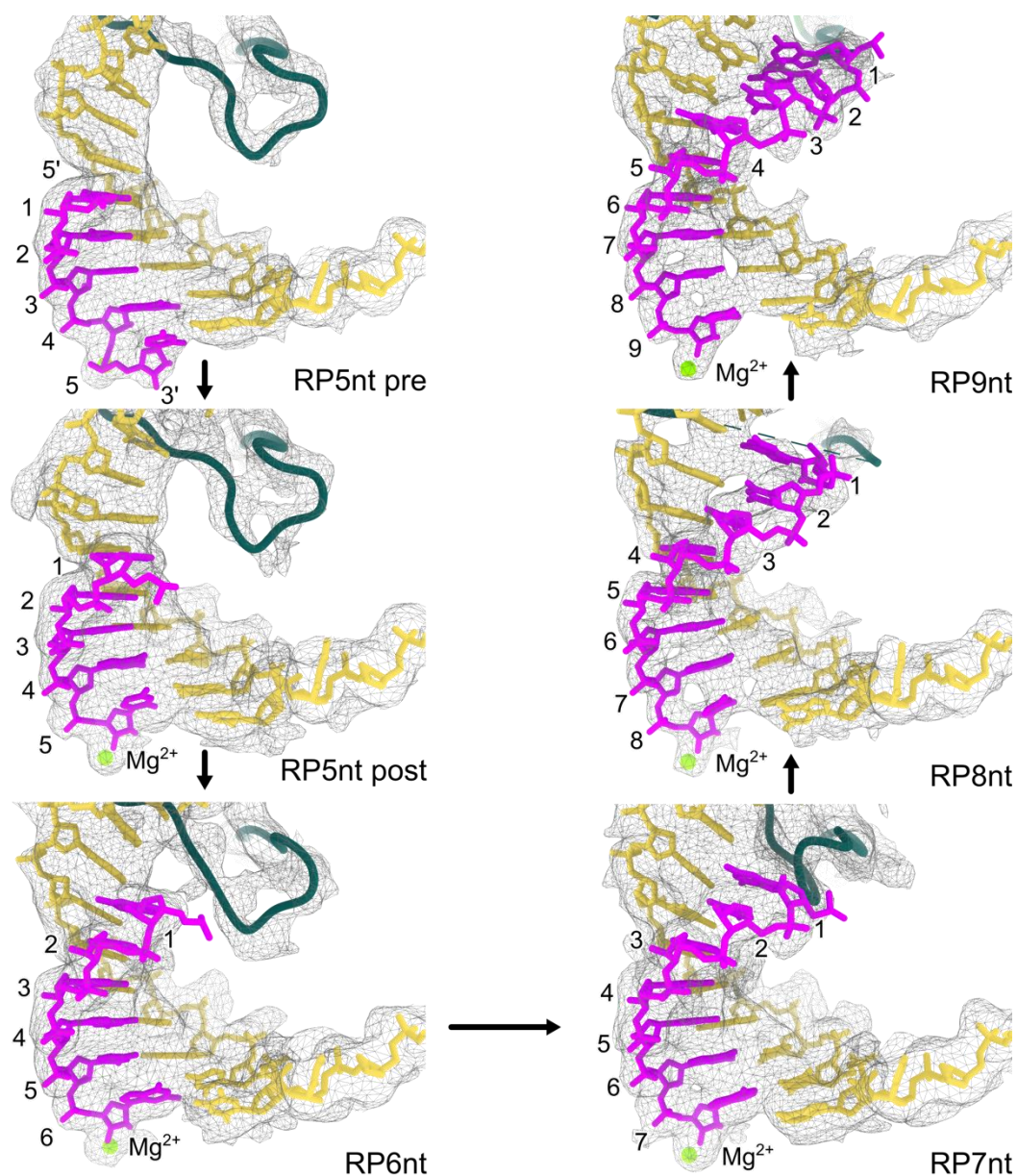


Figure 5.2.12: RNA, template DNA and RII densities and conformational changes at different transcript lengths.

To determine the mechanisms of σ^{54} displacement, we determined increasing lengths of RNA from 5-9nt (Supplementary Table 3). The processing strategy was identical to the RP5nt dataset. In all 6-9nt complex datasets we found a subset of σ^{54} -free elongation states, as well as RNA-bound post-translocated σ^{54} -bound states

(Figure 5.2.8-Figure 5.2.11). We found that longer transcripts caused folding back of the RII finger as the 5' end of the RNA contacts RII.

We found that RII at 7nt, the RII finger is coiled over to accommodate for the enlarged DNA-RNA bubble and prevent steric clashes; the tip of the RII finger is coiled backwards by 9.5 Å (as measured from the C α atoms), with the tip of the RII finger containing hydrophobic residues which form hydrophobic interactions with the 5' base. In 8nt and 9nt, the density for the RII-finger is partially lost, indicating this region is flexible and dynamic in nature. P110 and Y112 were also not resolved, indicating that these interactions were no longer present.

Taken together, this suggests that RNA extension between 7-9nt could enable pushing out of the RII-finger by breaking the hydrophobic interactions between the β' lid, β' switch region II and RII residues P110-Y112.

5.2.3 Transcription extension drives increased perturbation of transcription bubble and extra-long helix densities.

In addition to RII scrunching, we also see reduced ELH and transcription bubble densities as the transcription bubble and DNA-RNA hybrid size increased. Unlike RII-finger scrunching, ELH density is more smeared from 6nt of RNA, indicating that there is an increased flexibility of ELH.

ELH-HTH flexibility can in part be a result of the enlarged transcription bubble, and we also observe more broken transcription bubble density, suggesting that transcription bubble opening and scrunching allows ELH to slide and potentially

retract out from the transcription bubble. Similar observations of ELH retraction has been observed on the pathway to open complex formation in the partially loaded (RPip) state (PDB ID: 6GH6) (Glyde et al., 2018).

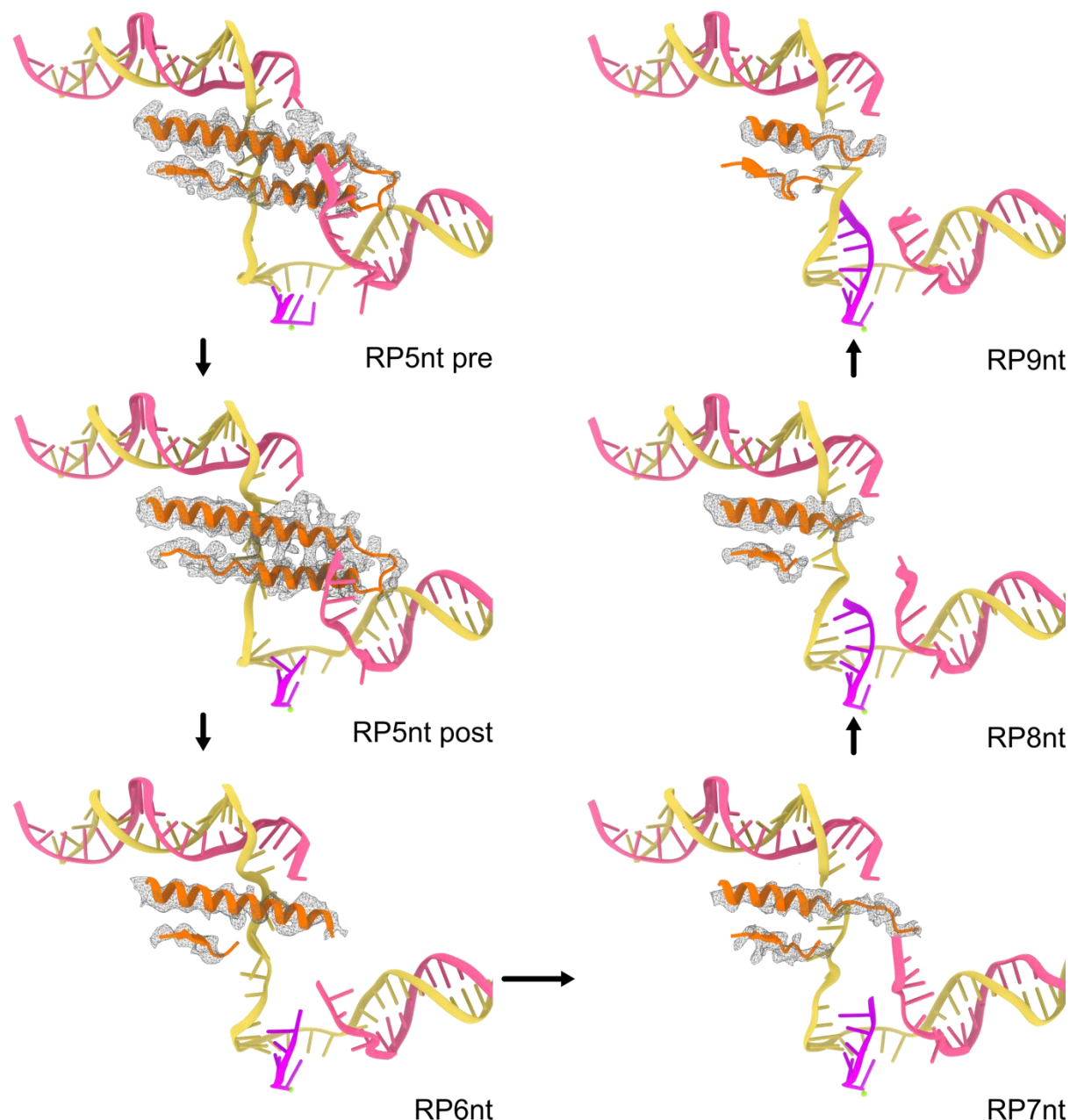


Figure 5.2.13: Increased RNA leads to loss of ELH density.

We also observe a significant loss of resolution of the upstream DNA and upstream DNA binding subdomains on σ^{54} (RpoN, ELH-HTH) in the local resolution maps,

suggesting that the binding of σ^{54} upstream DNA binding subdomains are less stable upon DNA scrunching (Figure 5.2.13).

Taken together, this suggests a potential path for σ^{54} dissociation, whereby DNA scrunching allows the upstream DNA binding RpoN and HTH become free, with the ELH being released from the enlarged DNA bubble. RNA extension then drives extrusion and folding back of RII from the RNA exit channel, likely pulling an flexible region I along with it, and finally dissociation of the CBD from the RNAP core.

5.3 Summary

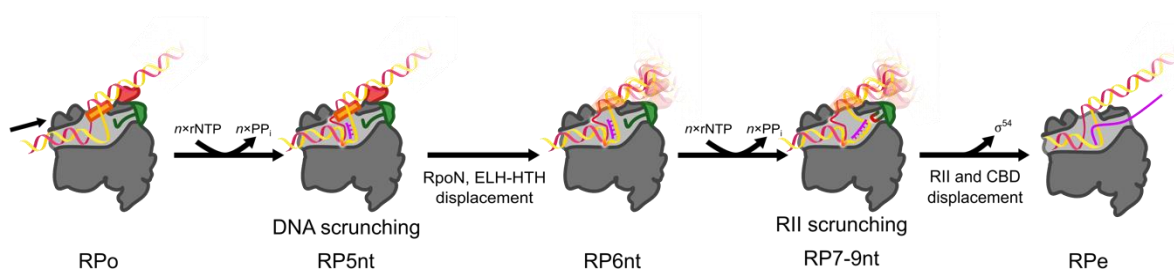


Figure 5.3.1: Proposed mechanism of σ^{54} displacement.

The structures presented above suggests that σ^{54} displacement occurs by both DNA scrunching, which releases the confinement of ELH, and RII displacement by RNA extension. Based on our RP5nt-6nt complexes, we suggest that early transcription leads to transcription bubble enlargement, as the downstream DNA is pushed into the catalytic site. This is shown by the 3d classes of both pre and post-translocated complexes within our 5nt dataset, and in accordance with previous σ^{70} FRET experiments demonstrating RNAP remains bound to the upstream fork of the transcription bubble whilst downstream DNA is pulled in (Kapanidis et al., 2006).

The ELH interacts with and stabilises the transcription bubble in the RPo structure (Glyde et al., 2018). During initial transcription, the increased transcription bubble releases the interactions with the hydrophobic bases in the transcription bubble and ELH. This increases flexibility of the ELH, as it is allowed to move out of the bubble (Supplementary Movie 7). The movement of ELH, and DNA supercoiling effects caused by scrunching leads to less stable binding of the HTH and RpoN subdomains as demonstrated by the reduced resolution of these subdomains.

At 7nt, we see that RII becomes folded back to accommodate the DNA-RNA bubble. As RNA length increases to 9nt, RII becomes more flexible and disordered (Supplementary Movie 7). This is similar to the σ^{70} finger mechanism seen in X-ray crystallographic work, however, this cryo-EM structure demonstrates that this is likely a universal mechanism for σ displacement and not a crystallographic artifact, with R3.2 of σ^{70} being functionally similar to σ^{54} RII, and the CBD playing similar roles to R4.1 (Chen et al., 2020; Li et al., 2020). As the RNA is extended, it is possible that as the RpoN and HTH subdomains dissociate, we propose that this enables RII to be pulled out as the CBD dissociates (Figure 5.3.1).

CBD does not bind to DNA, and we did not observe significant movement of the β' rudder, β' lid, β' coiled-coil, β' zipper, β fork loop or β flap. This suggests that transcription bubble enlargement and DNA scrunching is unlikely to significantly perturb the structure of the catalytic site or CBD binding site and must be carried out by RII movement. Therefore, RNA displacement is still likely to be essential for σ^{54} dissociation, as DNA scrunching cannot counteract CBD binding to the β' and α -CTD subunits that block the RNA exit channel. Whilst our conclusions must be

caveated due to the use of pre-synthesised RNA, both bubble enlargement and RNA scrunching have been observed *de novo* in previous FRET and crystallographic studies (Kapanidis et al., 2006; Zuo et al., 2020), although not using σ^{54} .

Chapter 6

Discussion

6.1 Mechanisms of σ^{54} mediated, PspF₁₋₂₇₅-driven open complex formation

The results described here provide insights into the DNA-melting stages of open complex formation. The high resolution structure of the RPc with fully duplexed promoter DNA, RPi(-12/-11) to RP(-10/-1) demonstrates that the initial melting of the transcription bubble occurs outside of the RNAP clamp, which is initiated by region I and ELH, and stabilised by RpoN, HTH and PspF₁₋₂₇₅ protomer 1 and 2 interactions with DNA. We further confirm that the residues of the L1 loops interact primarily with hydrophobic residues on the N-terminus region I and do not play direct a role in nucleation of DNA melting, as previously speculated (Dago et al., 2007; Danson et al., 2019; Zhang et al., 2009).

We show that DNA melting is coordinated with region I translocation through the central pore of PspF₁₋₂₇₅. This strongly suggests that bacterial enhancer binding proteins utilise ATP hydrolysis to thread and unfold region I through the nascent transcription bubble. This plays two roles; first, by unfolding region I, the hydrophobic interactions between the $\alpha 1$ helix and ELH, which forms a gate in the closed complex, is disrupted, removing the barrier blocking DNA loading. Secondly,

by tilting the PspF₁₋₂₇₅ ring as region I is threaded through the DNA bubble, the downstream DNA is bent, and the DNA bubble is further melted by at least 1 extra base pair as demonstrated by RP(-11/-8) conformation 3. We also determined a pre-RPo state, where the ELH is in a retracted state, similar to RPip, however, the clamp is closed. This suggests that the RPip state is likely a transient unstable intermediate and could explain why the R336A intermediate is required to stabilise this state.

Furthermore, the RP(-10/-1), as well as conformation 2, 3 and 5 of the RP(-11/-8) structure uncovers a new potential interaction interface between the RpoN subdomain of σ^{54} -region III and W56 of PspF₁₋₂₇₅. Previous crosslinking work has suggested interaction between PspF₁₋₂₇₅ and the -29-24 regions of the DNA (Zhang, Joly & Buck, 2012), however, no study to our knowledge has shown interaction between RpoN and the proposed W56 region on PspF₁₋₂₇₅. However, mutagenesis studies of the W56 residue has demonstrated that W56 plays an inhibitory role by binding to PspA in the absence of stimuli; W56A mutants of PspF₁₋₂₇₅ bypass PspA regulation and display over 2-fold increase in RPo formation efficiency. W56 is nearby to Y51, L52, S54 and Q57, all of which when mutated to cysteine display greater RPo formation efficiency. Taken together, this suggests that the RpoN interaction with the W56 region could not only stabilise the tilted conformation of PspF₁₋₂₇₅ but also play a role in preventing binding of inhibitory PspA and therefore help stimulate RPo formation (Joly, Burrows & Buck, 2008; Zhang et al., 2013).

6.2 σ^{70} mediated open complex formation comparison

These structures, along with structures from Darst and colleagues demonstrate conserved and diverging themes of open complex initiation (Chen et al., 2020) (Figure 6.2.1).

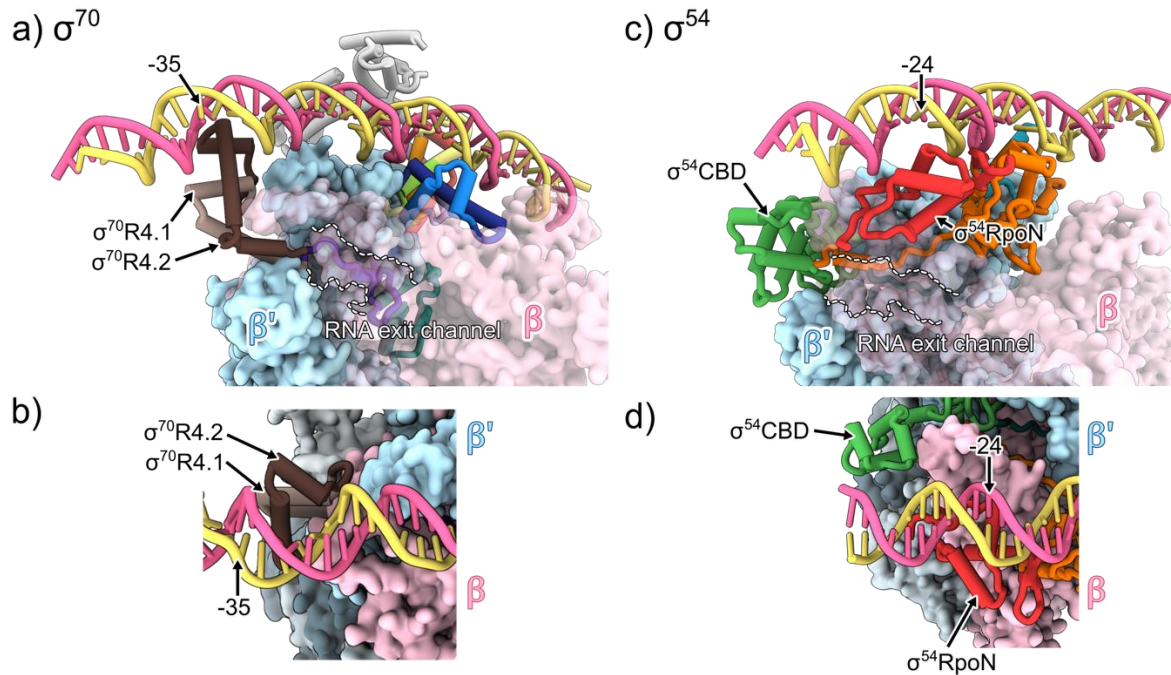


Figure 6.2.1: Comparison of between σ^{70} (a, b) and σ^{54} (c, d) upstream DNA binding elements. a, c) RNA exit channel indicated by a dotted white line. b, d) Top view of upstream DNA

In both closed complexes, both σ factors bind to similar regions of the DNA and RNAP. For example, σ^{70} region 4.2 and the RpoN domain both bind to upstream regions, σ^{70} regions 2.4 and 3 play analogous roles in binding to the -10 element, akin to ELH-HTH and region I in σ^{54} . However, the functionally analogous subdomains are located on different areas of RNAP; for example, σ^{70} region 4, responsible for recognition of the -35 region binds to the RNA exit channel on the β' side of RNAP, whereas the RpoN domain is on the β side of the RNAP (Figure 6.2.1b,d), and the CBD binds separately to the β' RNA exit channel without binding

to DNA (Figure 6.2.1a,c). σ^{70} region 2 interacts with the -10 element and forms the primary interface with the RNAP β' subunit, whereas σ^{54} separates these functions into RI-ELH for -12 binding and the CBD. σ^{54} region II enters the RNAP cleft and occupies the downstream dsDNA channel (RII.1) as well as the catalytic site and RNA exit channel (RII.2-3), preventing premature DNA loading; in σ^{70} this function is separated into σ^{70} R1.1, and the σ finger (region 3.2) (Figure 6.2.2).

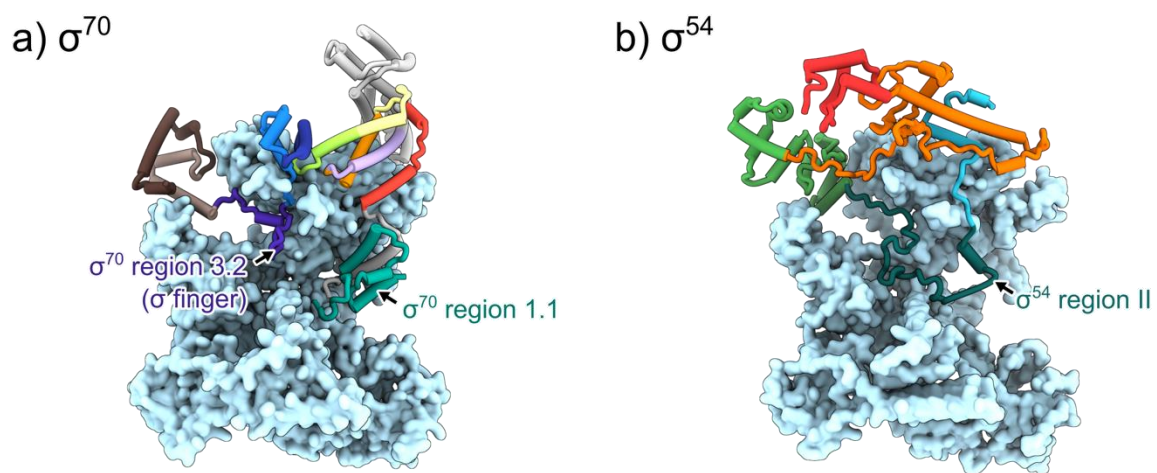


Figure 6.2.2: Comparison of subdomains binding to the internal cleft in σ^{70} (a, PDB ID:6PSQ) and σ^{54} (b, PDB ID: 5NWT)

Despite σ^{54} requiring active remodelling to form the open complex, initial bubble nucleation in σ^{70} and σ^{54} occurs outside of the RNAP clamp and irrespective of activator binding and clamp opening. However, the mechanisms of nucleation are very different; σ^{70} bubble nucleation involved isomerisation of conserved tryptophan dyad residues from “edge-on” to “chair” conformation to flip out the -11 base on the non-template and insert in between the -11 base pair to form stabilising aromatic stacking interactions with the flipped out base. On the other hand, σ^{54} utilises salt-bridge interactions between S335 on ELH and the -12 base, that is

flipped outwards. RI then inserts between the DNA and forms hydrophobic interactions via M15 and the -11 and -12 bases. Spontaneous melting by σ^{70} is enabled by the 3-5 base-specific pockets on σ^{70} , β lobe and β protrusion that allows for bubble extension that enable the flipping and stabilisation of -11 (non-template, conserved), -9 (template, conserved), -7 (non-template, conserved), -5 (non-template, non-conserved) and +2 (non-template, non-conserved) nucleotides. By contrast, σ^{54} does not utilise β subunit pockets for upstream DNA bases in the melted intermediate states. Consequently, σ^{54} requires active remodelling by bEBPs and insertion of ELH in between the DNA strands in the open and initially transcribing complexes. We have shown that the further melting from -12/-11 mismatch to -11/-8 partially open state is stabilised by interaction of the -12 and -11 bases with ELH, however, further stabilising interactions were not observed.

Recent work from the Darst group on *M. tuberculosis* and *E. coli* σ^{70} structures used antibiotic corallopyronin A to trap a partially melted bubble. The structures show suggest the that final loading also involves clamp opening to separate the β fork loop 2 and β' switch region 2 (Figure 6.2.3a, b) (Boyaci et al., 2019; Chen et al., 2020). However, the clamp opening is very minor, occurs solely between the corallopyronin A and RP2 states to enable loading of the ssDNA into the catalytic site, once a ~9nt bubble has already formed and the downstream DNA is already inside the cleft (Figure 6.2.3c). However, later studies on *E. coli* RNAP without corallopyronin A did not observe clamp, switch region 2 or fork loop 2 movement, suggesting that this may be *M. tuberculosis* specific conformation, a highly unstable intermediate or an off-pathway state caused by the antibiotic. In comparison, the

RPip structure shows no DNA strand separation, and the cleft is in a much wider open 39 Å state. This also makes sense, as our RPc structures show the clamp width is much less than the σ^{70} RPc structure, due to the larger, α helical $\sigma 1.1$ subdomain sitting inside the cleft.

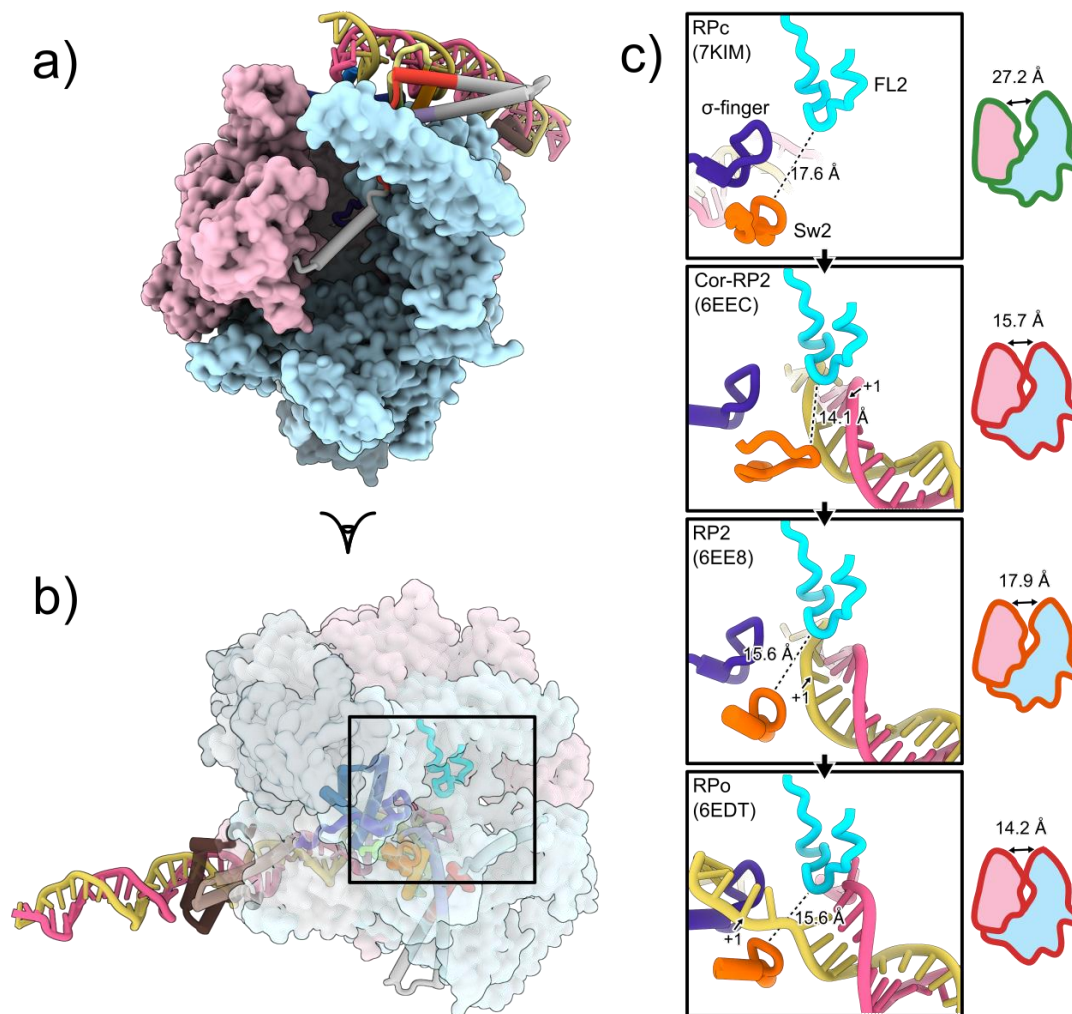


Figure 6.2.3: *M. tuberculosis* open complex formation involves open of the clamp to separate the switch region 2 (Sw2, orange) and fork loop 2 (FL2, cyan) in the final stages of DNA loading to enable template strand access to the catalytic site. a) *M. tuberculosis* closed complex structure as viewed from the downstream clamp. b) bottom view of the *M. tuberculosis* closed complex structure, showing the internal active site. c) zoomed in region showing the FL2 and Sw2 movements from RPc→Cor-RP2→RP2→RPo. Adapted from Boyaci et al. (2019); Lilic, Darst & Campbell (2021).

6.3 Mechanism of bEBP hydrolysis and coordination

Our results demonstrate that bEBPs hydrolyse ATP to translocate the N-terminus of σ^{54} region I through DNA and its central pore to enable DNA melting. The anticlockwise movement of the apo interfaces between RP(-11/-8) and RP(-10/-1) suggests that PspF₁₋₂₇₅ potentially undergoes a 1→6 ATP-hydrolysis-driven power stroke, as protomer 6 goes from ADP-bound to apo from RPi(-12/-11) to RP(-11/-8) conformation 3. In RP(-10/-1), the protomer states shift 1 subunit anticlockwise, with protomer previously in RP(-11/-8) being protomer 6 becoming nucleotide bound and moving to the protomer 5 position. Furthermore, protomer 1 binding site changes to an *apo* state, and moves to the protomer 6 position, suggesting nucleotide exchange occurring from position 6→position 5, similar to the ClpXP mechanism (Fei et al., 2019), except the N-terminus is only translocated ~1 residue rather than ≥6. This apparent lack of translocation could be due to the bonds between region I and L1 being mostly weaker hydrophobic interactions, as well as use of non-hydrolysable ADP.AIFx.

The ClpXP mechanism is in contrast to the classical mechanism, where the power stroke occurs from position 5 → position 6, with the nucleotide exchange occurring from position 6 → position 1. It is possible that the RP(-10/-1) could represent the RP(-11/-8) structure after 5 rounds of ATP hydrolysis, however, our high-resolution RPi(-12/-11) structure clearly shows that the protomer in position 5 is in the apo state and protomer 6 is ADP bound. This configuration conflicts with the classical AAA+ hydrolysis mechanism, where protomer 5 moves to position 6 before nucleotide exchange can eventually occur (Figure 6.3.1a). The RPi(-12/-11) conformation is more likely to fit into a ClpXP mechanism and could represent the

state before ATP binding. Taken together, these results indicate that bEBPs are likely to use a mechanism similar to ClpXP (Figure 6.3.1b).

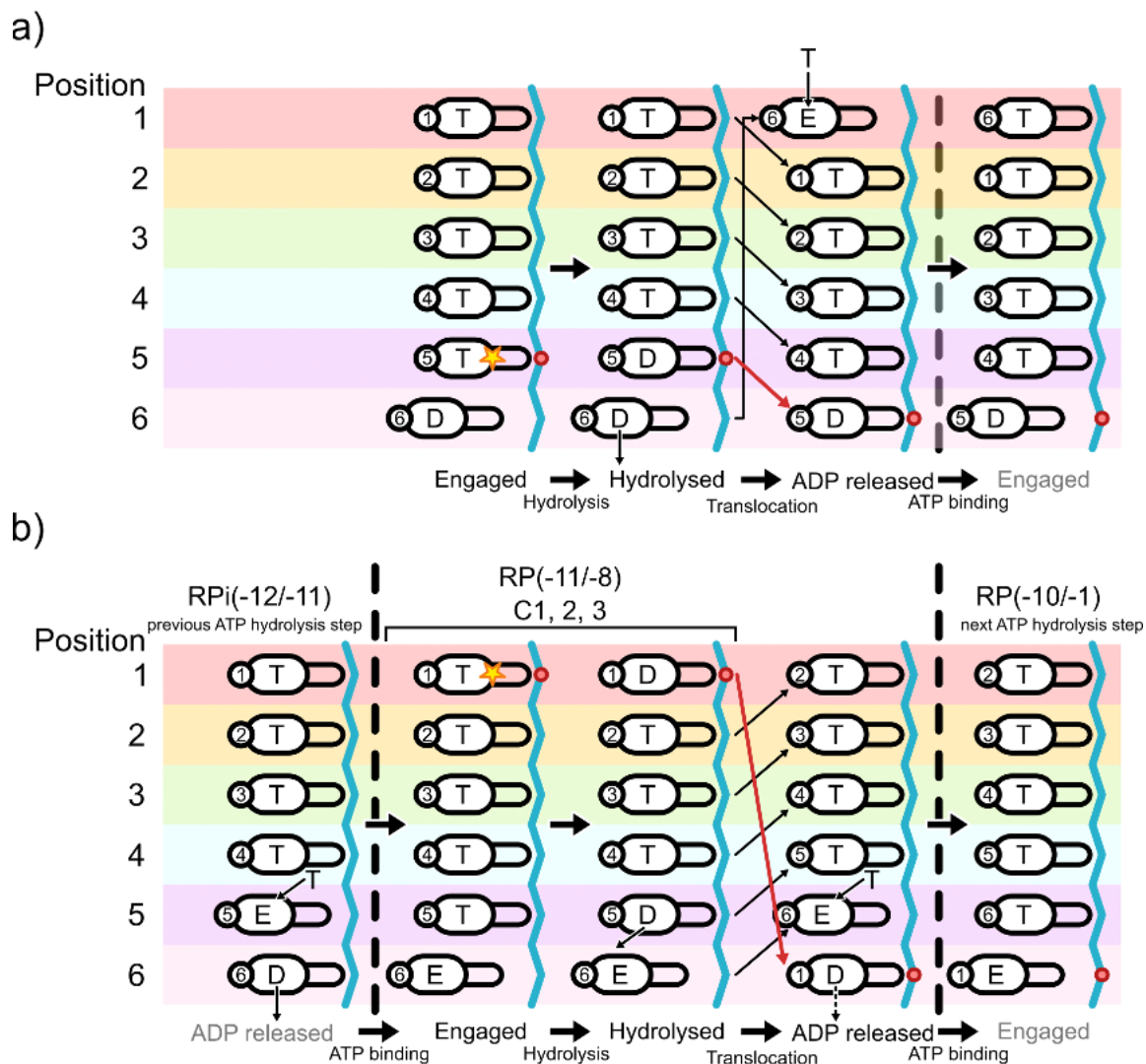


Figure 6.3.1: a) Classical mechanism. T= ATP-bound, D= ADP-bound, E=Empty. Star indicates the protomer undergoing the power stroke, with the red dot indicating the residue remaining bound to the protomer as it translocates (red arrow). The protomer number is indicated in the circle, as opposed to the position number, which is indicated by the colours and number on the left hand side. b) ClpXP mechanism, with the RPi(-12/-11), RP(-11/-8) and RP(-10/-1) conformations fitted into the pathway C1,2,3 referring to conformation 1, 2 and 3 respectively.

6.4 Comparison of bEBPs with other AAA+ proteins

bEBPs are functionally similar to other classical AAA+ proteins, containing conserved walker A and B, sensor and arginine finger motifs, that are used to couple chemical energy released from ATP hydrolysis with mechanical movement that is used to translocate substrate. We have shown that PspF₁₋₂₇₅, a bEBP adopts a spiral staircase arrangement, suggesting a sequential or partially sequential mechanism of action and opposing stochastic or concerted mechanisms.

However, bEBPs differ due to the additional L1 loop that is formed by an $\alpha 2$ insert. In bEBPs, the L1 loop forms the primary interacting loop with the region I substrate, whereas in classical AAA+ proteins the pore loop (also known as the L2 loop in bEBPs) are the primary interacting partner with the substrate.

Furthermore, bEBPs are unique in that they perform a binary switch like function by activating gene transcription in response to specific stimuli. This is in contrast to most AAA+ proteins that function to translocate, degrade, unwind or unfold their substrates in a more stepwise processive manner. bEBPs are the first AAA+ proteins to our knowledge that melt DNA by threading through a polypeptide chain, and it remains unknown how this sequential (or partially sequential) mechanism evolved.

6.5 σ^{54} -mediated promoter escape and comparison with σ^{70}

Our results demonstrate that promoter escape mechanisms in both σ^{54} and σ^{70} utilise similar mechanisms, with σ^{54} RII.3-CBD and σ^{70} sigma-finger (σ^{70} R3.2)- σ^{70} R4 binding to the RNA exit site and inserting into the RNA exit channel. Our results show that with increasing lengths of RNA, RII folds back to accommodate the enlarged DNA-RNA duplex. This is similar to the mechanism shown by Ebright and co-workers, demonstrating that the σ finger is displaced in a similar manner by the 5' end of the RNA. However, σ^{70} R3.2 enters much deeper into the catalytic site compared to σ^{54} RII.3 and consequently begins to fold back at a shorter RNA length of 5nt compared to 6-7nt in σ^{54} (Figure 6.5.1).

However, it is worth noting that the work by Li et al. (2020) did not include a full transcription bubble, and only the downstream fork. We also observe an increase in ELH flexibility probably as a result of transcription bubble scrunching. Scrunching of the DNA has been previously observed in FRET experiments with σ^{70} and has been shown to also play a role in the build-up of stress within the walls of the RNAP catalytic site. We propose that scrunching of DNA in σ^{54} plays a role primarily in the release of ELH-HTH and RpoN from the upstream DNA.

The CBD plays a critical role in retention of σ^{54} and is well conserved throughout all σ^{54} sequences, unlike RII which in some species are significantly shortened (for example in *Rhodobacter capsulatus*) (Buck et al., 2000). This suggests that in these species, σ^{54} displacement may occur at longer RNA lengths, and involve more DNA scrunching, as there is more space in the catalytic site for RNA to extend, as

observed in σ^H (ECF) factor in Li et al. (2020). It would be interesting to compare displacement mechanisms of shortened σ^{54} variant factors to see if the mechanisms are conserved.

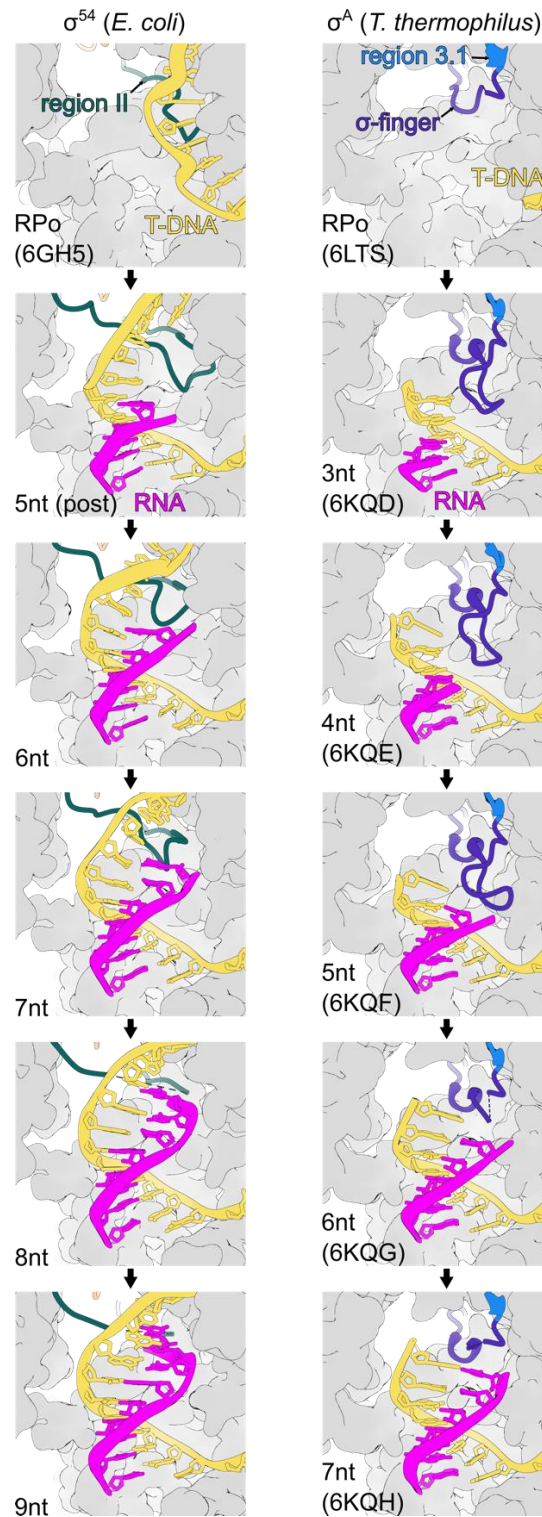


Figure 6.5.1: Comparison of σ^{54} and σ^{70} (σ^A) initially transcribing complexes demonstrates that both utilise a similar mechanism of displacement of subdomains blocking the RNA exit channel, in which the 5' end of the growing RNA displaces σ factor, causing it to fold backwards.

6.6 Eukaryotic perspectives

Whilst there are functionally analogous factors in eukaryotes, the mechanisms are very different. For example, with σ^{54} , both bEBPs and TFIIF are ATPases that are important for DNA melting in RNAPII, however, TFIIF is a ratcheting helicase that translocates along double-stranded DNA binds downstream regions of the transcription start site. TFIIF pushes DNA towards the active site, whilst remaining tethered to the RNAPII core complex, resulting in DNA unwinding. By comparison, bEBPs interact primarily with σ^{54} , and unfold σ^{54} to enable DNA opening rather than directly unwinding DNA.

The term bacterial enhancer binding proteins is derived from eukaryotic enhancer proteins; whilst enhancers such as the mediator complex play roles in promoting transition to processive elongation, orienting to enable open complex formation, the mechanisms are entirely different from bEBPs, or other processivity factors such as GreA/B or tethers such as RbpA. Furthermore, the Eukaryotic genome is orders of magnitude larger and has additional layers of nucleosomal regulation; enhancers in eukaryotes can be 100 kilobases upstream of the target promoter, and form much larger supramolecular structures compared to prokaryotic bEBPs, which bind 80-150 bp upstream of their target promoters.

RNAPI RRN7, RNAPII TFIIB and RNAPIII Rpc82/RPC3 in eukaryotic polymerases all share similarities with ELH in σ^{54} , in that they promote strand separation stabilise the transcription bubble in the open complex (Figure 6.6.1). TFIIF in RNAPII shares functional similarities with σ^{54} region I and σ^{70} region 2.3; they all function to bind stabilise the initial transcription bubble; however, recent work by Cramer and

colleagues has shown that the initial bubble retains base stacking and does not undergo base flipping (Schilbach et al., 2021). Unlike the W-dyad or region I M15, no insertion between 2 DNA strands of the nascent transcription bubble occurs in eukaryotes – instead, TFIIB forms charge-charge interactions with the outside of the nascent 6nt DNA bubble, aided by a “clamp-head-loop” which is in a similar location to the β' clamp tip in prokaryotes.

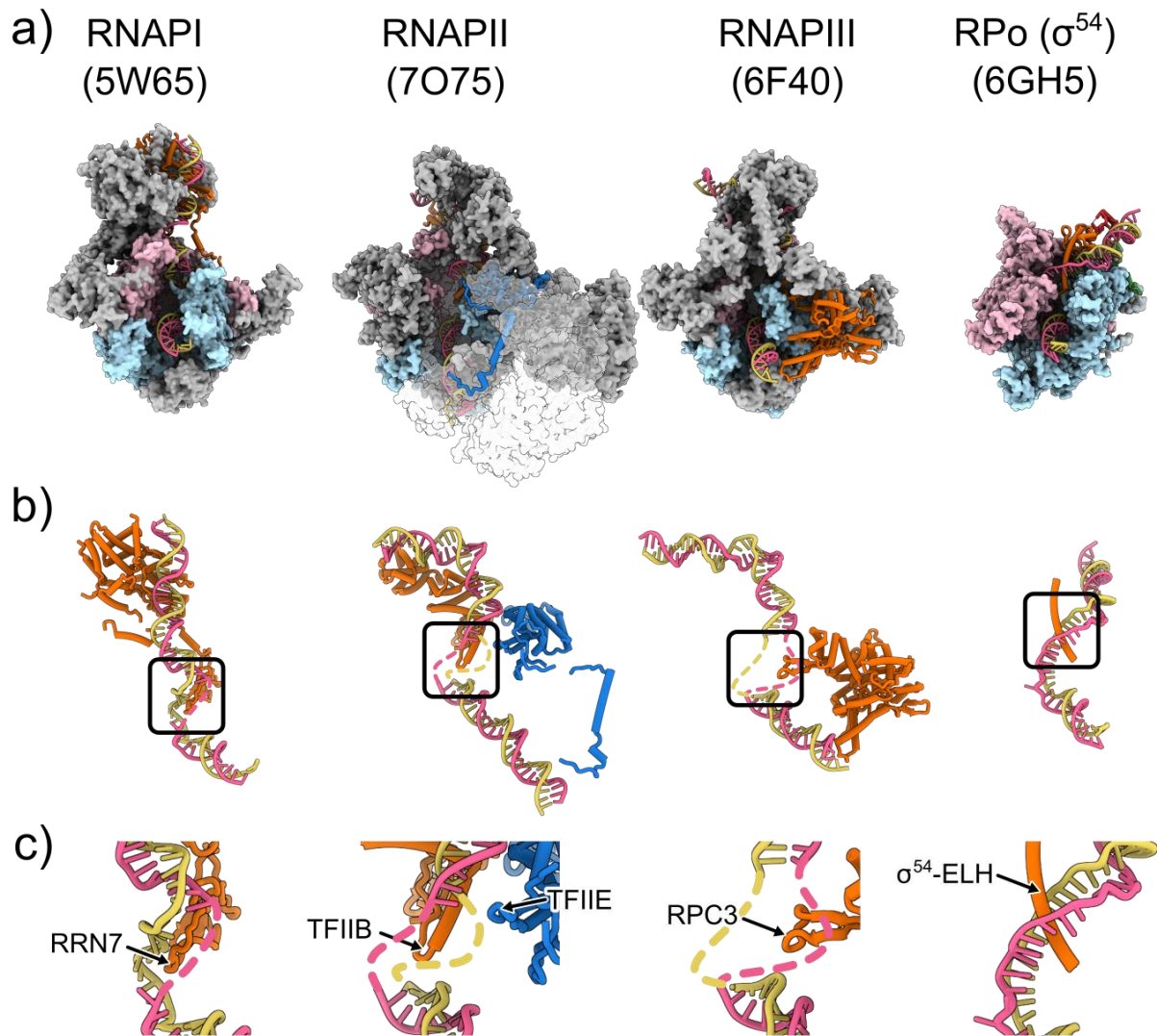


Figure 6.6.1: Comparison of eukaryotic and σ^{54} mechanisms in bubble stabilisation reveals common themes in bubble maintenance. TFIIH is shown in transparent for clarity. Coloration of RPB1 and RPB2 is the same as β' and β subunits, with the rest being coloured in grey for clarity.

Work in this thesis has highlighted similarities and differences in DNA melting and loading mechanisms between RNAPs. The recently determined high resolution *S. cerevisiae* RNAPII intermediate complex also demonstrates that early DNA melting occurs without clamp opening (Schilbach et al., 2021). Similar to σ^{70} , the new structures show that RNAPII clamp opening is not required for early DNA melting,

with the clamp remaining closed in both closed, intermediate and open states. However, comparison of the clamp widths of the closed complexes for both σ^{70} (PDB ID 7PSQ, 7KIM), and RNAPII (PDB ID 7O72) show that the clamps are wider than our wildtype RPc structure, suggesting that clamp opening in σ^{54} could still be required for open complex formation.

Promoter escape mechanisms are more complex and diverse in eukaryotes; previous reports have speculated that RNA extension would lead to displacement of TFIIB/Rrn7 inside the RNA exit channel, similar to region II or sigma finger or other promoter binding elements (Dexl et al., 2018; Sadian et al., 2019). However, recent work from the Murakami group showed that *S. cerevisiae* RNAPII can synthesise up to 49 nucleotides before transitioning to the elongation state and does not involve clashes between RNA and TFIIB. Instead, *in vitro* studies suggest that DSIF/Spt4/5 as well as capping enzymes Cet1-Ceg1 interaction with the 5' end of the RNA may promote escape. Capping is also enhanced by phosphorylation of Ser5 by the Cdk7, a kinase subunit part of the TFIIH complex (Fujiwara et al., 2019). Other mechanisms include involvement of the +1 nucleosome and Chd1 Nucleosome remodelers (Skene et al., 2014) or collision with other elongation complexes (Yang et al., 2022).

6.7 Summary and Future Perspectives

This work reveals new structures populating a near-complete pathway from closed complex to the elongation state. We show that DNA melting occurs in the closed complex with region I facilitating DNA melting outside of the cleft and before DNA loading. We show that σ^{54} -mediated open complex formation involves translocation

of region I N-terminus through the central pore of the bEBP, which likely unfolds region I, alleviating the inhibition imposed by the ELH-region I gate. Upon clamp loading, ELH transitions from a retracted (pre-RPo) to fully α -helical (RPo) state where it inserts between the DNA strands to stabilise the transcription bubble. Promoter escape occurs via a two-step process: first downstream DNA is fed into the active site, leading to transcription bubble scrunching due to the enlarged transcription bubble. This causes increased flexibility of the ELH, which begins to become free of the transcription bubble. At 7nt, the 5' end of the RNA begins to fold back region II. Upon further extension, the RNA pushes out the CBD blocking the RNA exit channel, leading to σ^{54} displacement, and transition to the elongation state. Knowing when σ^{54} is released from the transcription complex will likely require more extensive RNA synthesis to occur and new structures of these >9nt RNA states.

Our current work was limited by the inability of PspF₁₋₂₇₅ to form stable complexes with other ATP hydrolysis analogues or with walker B mutants. Future work would be to determine the mechanism of ATP hydrolysis coordination by bEBPs using ATP and time-resolved cryo-EM of wildtype on-pathway intermediate states or single molecule FRET studies. We would also like to explore optical tweezers and FRET experiments to be able to measure exact position of the N-terminus of region I during its ATP-hydrolysis driven translocation.

Similarly, we would also like to obtain *de novo* synthesised RPitc structures using longer transcript states. *In vitro* transcription assays have demonstrated that the initial transcripts from the *nifH* promoter DNA are not of uniform length, and this has hindered attempts at structure determination due to low occupancy and

heterogeneity due to abortive and reiterative transcription. In future work, we would like to test different promoter sequences, as well as studying potential roles of processivity factors to capture conclusive on-pathway stressed initially transcribing states prior to elongation.

Bibliography

- Abdelkareem, M., Saint-Andre, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J. & Weixlbaumer, A. (2019) Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell*. 75 (2), 298-309 e4. doi: 10.1016/j.molcel.2019.04.029.
- Abel, C.-B., Verónica, M.-F., I, G.-G. A. & Francisco, N. (2018) Subunits Common to RNA Polymerases. *The Yeast Role in Medical Applications*.
- Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K. & Walter, P. (2022) *Molecular biology of the cell*. W. W. Norton & Company.
- Bae, B., Chen, J., Davis, E., Leon, K., Darst, S. A. & Campbell, E. A. (2015) CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife*. 4. doi: 10.7554/eLife.08505.
- Basu, R. S., Warner, B. A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A. & Murakami, K. S. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem*. 289 (35), 24549-59. doi: 10.1074/jbc.M114.584037.

Baum, B. & Baum, D. A. (2020) The merger that made us. *BMC Biol.* 18 (1), 72. doi: 10.1186/s12915-020-00806-3.

Bellecourt, M. J., Ray-Soni, A., Harwig, A., Mooney, R. A. & Landick, R. (2019) RNA Polymerase Clamp Movement Aids Dissociation from DNA but Is Not Required for RNA Release at Intrinsic Terminators. *J Mol Biol.* 431 (4), 696-713. doi: 10.1016/j.jmb.2019.01.003.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. (2019a) Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods.* 16 (11), 1153-1160. doi: 10.1038/s41592-019-0575-8.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. (2019b) Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods.* 16 (11), 1153-1160. doi: 10.1038/s41592-019-0575-8.

Bernecky, C., Plitzko, J. M. & Cramer, P. (2017) Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol.* 24 (10), 809-815. doi: 10.1038/nsmb.3465.

Blaha, G. M. & Wade, J. T. (2022) Transcription-Translation Coupling in Bacteria. *Annu Rev Genet.* 56, 187-205. doi: 10.1146/annurev-genet-072220-033342.

Bordes, P., Wigneshweraraj, S. R., Schumacher, J., Zhang, X., Chaney, M. & Buck, M. (2003) The ATP hydrolyzing transcription activator phage shock protein F of *Escherichia*

coli: identifying a surface that binds sigma 54. *Proc Natl Acad Sci U S A*. 100 (5), 2278-83. doi: 10.1073/pnas.0537525100.

Borner, T., Aleynikova, A. Y., Zubo, Y. O. & Kusnetsov, V. V. (2015) Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim Biophys Acta*. 1847 (9), 761-9. doi: 10.1016/j.bbabbio.2015.02.004.

Borukhov, S., Sagitov, V. & Goldfarb, A. (1993) Transcript cleavage factors from *E. coli*. *Cell*. 72 (3), 459-466. doi: 10.1016/0092-8674(93)90121-6.

Boyaci, H., Chen, J., Jansen, R., Darst, S. A. & Campbell, E. A. (2019) Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature*. 1. doi: doi:10.1038/s41586-018-0840-5.

Brueckner, F., Armache, K. J., Cheung, A., Damsma, G. E., Kettenberger, H., Lehmann, E., Sydow, J. & Cramer, P. (2009) Structure-function studies of the RNA polymerase II elongation complex. *Acta Crystallogr D Biol Crystallogr*. 65 (Pt 2), 112-20. doi: 10.1107/S0907444908039875.

Brueckner, F. & Cramer, P. (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol*. 15 (8), 811-8. doi: 10.1038/nsmb.1458.

Buck, M., Gallegos, M.-T., Studholme, D. J., Guo, Y. & Gralla, J. D. (2000) The Bacterial Enhancer-Dependent $\sigma 54$ (σN) Transcription Factor. doi: 10.1128/JB.182.15.4129-4136.2000.

- Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. (1969) Factor stimulating transcription by RNA polymerase. *Nature*. 221 (5175), 43-6. doi: 10.1038/221043a0.
- Busby, S. & Ebright, R. H. (1999) Transcription activation by catabolite activator protein (CAP). *J Mol Biol*. 293 (2), 199-213. doi: 10.1006/jmbi.1999.3161.
- Bush, M. & Dixon, R. (2012) The Role of Bacterial Enhancer Binding Proteins as Specialized Activators of σ 54-Dependent Transcription. *Microbiol Mol Biol Rev: Vol. 3*. pp. 497-529.
- Carragher, B., Cheng, Y., Frost, A., Glaeser, R. M., Lander, G. C., Nogales, E. & Wang, H. W. (2019) Current outcomes when optimizing 'standard' sample preparation for single-particle cryo-EM. *J Microsc*. 276 (1), 39-45. doi: 10.1111/jmi.12834.
- Cartagena, A. J., Banta, A. B., Sathyan, N., Ross, W., Gourse, R. L., Campbell, E. A. & Darst, S. A. (2019) Structural basis for transcription activation by Crl through tethering of sigma(S) and RNA polymerase. *Proc Natl Acad Sci U S A*. 116 (38), 18923-18927. doi: 10.1073/pnas.1910827116.
- Cavaliere, P., Levi-Acobas, F., Mayer, C., Saul, F. A., England, P., Weber, P., Raynal, B., Monteil, V., Bellalou, J., Haouz, A. & Norel, F. (2014) Structural and functional features of Crl proteins and identification of conserved surface residues required for interaction with the RpoS/sigmaS subunit of RNA polymerase. *Biochem J*. 463 (2), 215-24. doi: 10.1042/BJ20140578.
- Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J.,

- Weiss, S. & Ebright, R. H. (2012) Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science*. 337 (6094), 591--595. doi: 10.1126/science.1218716.
- Chaney, M., Grande, R., Wigneshweraraj, S. R., Cannon, W., Casaz, P., Gallegos, M. T., Schumacher, J., Jones, S., Elderkin, S., Dago, A. E., Morett, E. & Buck, M. (2001) Binding of transcriptional activators to sigma 54 in the presence of the transition state analog ADP-aluminum fluoride: insights into activator mechanochemical action. *Genes Dev*. 15 (17), 2282-94. doi: 10.1101/gad.205501.
- Chen, B., Doucleff, M., Wemmer, D. E., De Carlo, S., Huang, H. H., Nogales, E., Hoover, T. R., Kondrashkina, E., Guo, L. & Nixon, B. T. (2007) ATP ground- and transition-states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, σ 54. *Structure*. 15 (4), 429-40. doi: 10.1016/j.str.2007.02.007.
- Chen, B., Sysoeva, T. A., Chowdhury, S., Guo, L., De Carlo, S., Hanson, J. A., Yang, H. & Nixon, B. T. (2010) Engagement of arginine finger to ATP triggers large conformational changes in NtrC1 AAA+ ATPase for remodeling bacterial RNA polymerase. *Structure*. 18 (11), 1420-30. doi: 10.1016/j.str.2010.08.018.
- Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A. Y., Olinares, P. D. B., Saecker, R. M., Winkelman, J. T., Maloney, M. F., Chait, B. T., Ross, W., Gourse, R. L., Campbell, E. A. & Darst, S. A. (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol Cell*. 78 (2), 275-288 e6. doi: 10.1016/j.molcel.2020.02.017.
- Cheung, A. C. & Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*. 471 (7337), 249-53. doi: 10.1038/nature09785.

Cheung, A. C. & Cramer, P. (2012) A movie of RNA polymerase II transcription. *Cell*. 149 (7), 1431-7. doi: 10.1016/j.cell.2012.06.006.

Cramer, P. (2002) Multisubunit RNA polymerases. *Current Opinion in Structural Biology*. 12 (1), 89-97. doi: 10.1016/s0959-440x(02)00294-4.

Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C. D., Lehmann, E., Leike, K., Sydow, J. F. & Vannini, A. (2008) Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*. 37, 337-52. doi: 10.1146/annurev.biophys.37.032807.130008.

Crewe, A. V., Eggenberger, D. N., Wall, J. & Welter, L. M. (1968) Electron Gun Using a Field Emission Source. *Review of Scientific Instruments*. 39 (4), 576-583. doi: 10.1063/1.1683435.

Crick, F. (1970) Central Dogma of Molecular Biology. *Nature*. 227 (5258), 561. doi: doi:10.1038/227561a0.

Croll, T. (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D*. 74 (6), 519-530. doi: doi:10.1107/S2059798318002425.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res*. 14 (6), 1188-90. doi: 10.1101/gr.849004.

Cvetkovic, M. A., Ortega, E., Bellelli, R. & Costa, A. (2022) Multiple roles of Pol epsilon in eukaryotic chromosome replication. *Biochem Soc Trans.* 50 (1), 309-320. doi: 10.1042/BST20210082.

Dago, A. E., Wigneshweraraj, S. R., Buck, M. & Morett, E. (2007) A role for the conserved GAFTGA motif of AAA+ transcription activators in sensing promoter DNA conformation. *J Biol Chem.* 282 (2), 1087-97. doi: 10.1074/jbc.M608715200.

Danev, R., Yanagisawa, H. & Kikkawa, M. (2021) Cryo-EM performance testing of hardware and data acquisition strategies. *Microscopy (Oxf).* 70 (6), 487-497. doi: 10.1093/jmicro/dfab016.

Danson, A. E., Jovanovic, M., Buck, M. & Zhang, X. (2019) Mechanisms of sigma(54)-Dependent Transcription Initiation and Regulation. *J Mol Biol.* doi: 10.1016/j.jmb.2019.04.022.

Darst, S. A., Kubalek, E. W. & Kornberg, R. D. (1989) Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. *Nature.* 340 (6236), 730-2. doi: 10.1038/340730a0.

Darst, S. A., Ribí, H. O., Pierce, D. W. & Kornberg, R. D. (1988) Two-dimensional crystals of *Escherichia coli* RNA polymerase holoenzyme on positively charged lipid layers. *J Mol Biol.* 203 (1), 269-73. doi: 10.1016/0022-2836(88)90107-6.

Delbeau, M., Omollo, E. O., Froom, R., Koh, S., Mooney, R. A., Lilic, M., Brewer, J. J., Rock, J., Darst, S. A., Campbell, E. A. & Landick, R. (2022) Structural and functional basis of the universal transcription factor NusG pro-pausing activity in *Mycobacterium tuberculosis*. doi: 10.1101/2022.10.21.513233.

Dexl, S., Reichelt, R., Kraatz, K., Schulz, S., Grohmann, D., Bartlett, M. & Thomm, M. (2018) Displacement of the transcription factor B reader domain during transcription initiation. *Nucleic Acids Res.* 46 (19), 10066-10081. doi: 10.1093/nar/gky699.

Donnat, C., Levy, A., Poitevin, F., Zhong, E. D. & Miolane, N. (2022) Deep generative modeling for volume reconstruction in cryo-electron microscopy. *J Struct Biol.* 214 (4), 107920. doi: 10.1016/j.jsb.2022.107920.

Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, A. W. & Schultz, P. (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys.* 21 (2), 129-228. doi: 10.1017/s0033583500004297.

Dulin, D., Bauer, D. L. V., Malinen, A. M., Bakermans, J. J. W., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., Brodolin, K., Kulbachinskiy, A. & Kapanidis, A. N. (2018) Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nat Commun.* 9 (1), 1478. doi: 10.1038/s41467-018-03902-9.

Eduljee, A. D. (2021) *Structural and functional studies on transcriptional proofreading by GreA*

Étude structurale et fonctionnelle de l'activité de relecture par GreA au cours de la transcription. Université de Strasbourg,

Egerton, R. F. (2016) The Transmission Electron Microscope. *Physical Principles of Electron Microscopy.* pp. 55-88.

- Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H. & Sekine, S. I. (2019) Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science*. 363 (6428), 744-747. doi: 10.1126/science.aav8912.
- Ehara, H., Kujirai, T., Shirouzu, M., Kurumizaka, H. & Sekine, S. I. (2022) Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. *Science*. 377 (6611), eabp9466. doi: 10.1126/science.abp9466.
- Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M. & Sekine, S. I. (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science*. 357 (6354), 921-924. doi: 10.1126/science.aan8552.
- Elmlund, D., Le, S. N. & Elmlund, H. (2017) High-resolution cryo-EM: the nuts and bolts. *Curr Opin Struct Biol*. 46, 1-6. doi: 10.1016/j.sbi.2017.03.003.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 66 (Pt 4), 486-501. doi: 10.1107/S0907444910007493.
- Erickson, G. J. & Smith, C. R. (1988) *Maximum-Entropy and Bayesian Methods in Science and Engineering*.
- Fang, C., Li, L., Zhao, Y., Wu, X., Philips, S. J., You, L., Zhong, M., Shi, X., O'Halloran, T. V., Li, Q. & Zhang, Y. (2020) The bacterial multidrug resistance regulator BmrR distorts promoter DNA to activate transcription. *Nat Commun*. 11 (1), 6284. doi: 10.1038/s41467-020-20134-y.

Fang, C., Philips, S. J., Wu, X., Chen, K., Shi, J., Shen, L., Xu, J., Feng, Y., O'Halloran, T. V. & Zhang, Y. (2021) CueR activates transcription through a DNA distortion mechanism. *Nat Chem Biol.* 17 (1), 57-64. doi: 10.1038/s41589-020-00653-x.

Farnung, L., Ochmann, M., Engholm, M. & Cramer, P. (2021) Structural basis of nucleosome transcription mediated by Chd1 and FACT. *Nat Struct Mol Biol.* 28 (4), 382-387. doi: 10.1038/s41594-021-00578-6.

Faruqi, A. R. & McMullan, G. (2018) Direct imaging detectors for electron microscopy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment.* 878, 180-190. doi: 10.1016/j.nima.2017.07.037.

Fei, X., Bell, T. A., Jenni, S., Stinson, B. M., Baker, T. A., Harrison, S. C. & Sauer, R. T. (2019) Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. doi: 10.1101/704999.

Feng, X., Noguchi, Y., Barbon, M., Stillman, B., Speck, C. & Li, H. (2021) The structure of ORC-Cdc6 on an origin DNA reveals the mechanism of ORC activation by the replication initiator Cdc6. *Nat Commun.* 12 (1), 3883. doi: 10.1038/s41467-021-24199-1.

Filipovski, M., Soffers, J. H. M., Vos, S. M. & Farnung, L. (2022) Structural basis of nucleosome retention during transcription elongation. *Science.* 376 (6599), 1313-1316. doi: 10.1126/science.abo3851.

Franklin, R. E. & Gosling, R. G. (1953) Molecular configuration in sodium thymonucleate. *Nature.* 171 (4356), 740-1. doi: 10.1038/171740a0.

Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I. & Oren, M. (2014) 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* 15 (5), R69. doi: 10.1186/gb-2014-15-5-r69.

Fujiwara, R., Damodaren, N., Wilusz, J. E. & Murakami, K. (2019) The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. *Proc Natl Acad Sci U S A.* 116 (45), 22573-22582. doi: 10.1073/pnas.1905449116.

Furth, J. J., Hurwitz, J. & Anders, M. (1962) The Role of Deoxyribonucleic Acid in Ribonucleic Acid Synthesis. *Journal of Biological Chemistry.* 237 (8), 2611-2619. doi: 10.1016/s0021-9258(19)73796-x.

Gajos, M., Jasnovidova, O., van Bommel, A., Freier, S., Vingron, M. & Mayer, A. (2021) Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Res.* 49 (8), 4402-4420. doi: 10.1093/nar/gkab208.

Gallegos, M. T. & Buck, M. (2000) Sequences in sigma(54) region I required for binding to early melted DNA and their involvement in sigma-DNA isomerisation. *J Mol Biol.* 297 (4), 849-59. doi: 10.1006/jmbi.2000.3608.

Gao, F., Danson, A. E., Ye, F., Jovanovic, M., Buck, M. & Zhang, X. (2020) Bacterial Enhancer Binding Proteins—AAA+ Proteins in Transcription Activation. *Biomolecules.* 10 (3), 351. doi: 10.3390/biom10030351.

Geszvain, K. & Landick, R. (2004) The Structure of Bacterial RNA Polymerase. *The Bacterial Chromosome.* pp. 283-296.

Girbig, M., Xie, J., Grotsch, H., Libri, D., Porrua, O. & Muller, C. W. (2022) Architecture of the yeast Pol III pre-termination complex and pausing mechanism on poly(dT) termination signals. *Cell Rep.* 40 (10), 111316. doi: 10.1016/j.celrep.2022.111316.

Glaeser, R. M., Nogales, E. & Chiu, W. (2021) *Single-particle Cryo-EM of Biological Macromolecules*. IOP Publishing. <https://dx.doi.org/10.1088/978-0-7503-3039-8>.

Glyde, R., Ye, F., Darbari, V. C., Zhang, N., Buck, M. & Zhang, X. (2017) Structures of RNA Polymerase Closed and Intermediate Complexes Reveal Mechanisms of DNA Opening and Transcription Initiation. *Molecular Cell*. 67 (1), 106--116.e4. doi: 10.1016/j.molcel.2017.05.010.

Glyde, R., Ye, F., Jovanovic, M., Kotta-Loizou, I., Buck, M. & Zhang, X. (2018) Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. *Mol Cell*. 70 (6), 1111-1120 e3. doi: 10.1016/j.molcel.2018.05.021.

Guo, X., Myasnikov, A. G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P. & Weixlbaumer, A. (2018) Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell*. 69 (5), 816-827 e4. doi: 10.1016/j.molcel.2018.02.008.

Haag, J. R. & Pikaard, C. S. (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol*. 12 (8), 483-92. doi: 10.1038/nrm3152.

Han, H., Monroe, N., Sundquist, W. I., Shen, P. S. & Hill, C. P. (2017) The AAA ATPase Vps4 binds ESCRT-III substrates through a repeating array of dipeptide-binding pockets. *Elife*. 6. doi: 10.7554/eLife.31324.

Hao, M., Ye, F., Jovanovic, M., Kotta-Loizou, I., Xu, Q., Qin, X., Buck, M., Zhang, X. & Wang, M. (2022) Structures of Class I and Class II Transcription Complexes Reveal the Molecular Basis of RamA-Dependent Transcription Activation. *Adv Sci (Weinh)*. 9 (4), e2103669. doi: 10.1002/advs.202103669.

Harden, T. T., Herlambang, K. S., Chamberlain, M., Lalanne, J.-B., Wells, C. D., Li, G.-W., Landick, R., Hochschild, A., Kondev, J. & Gelles, J. (2020) Alternative transcription cycle for bacterial RNA polymerase. *Nature Communications*. 11 (1), 1-11. doi: doi:10.1038/s41467-019-14208-9.

Harden, T. T., Wells, C. D., Friedman, L. J., Landick, R., Hochschild, A., Kondev, J. & Gelles, J. (2016) Bacterial RNA polymerase can retain sigma70 throughout transcription. *Proc Natl Acad Sci U S A*. 113 (3), 602-7. doi: 10.1073/pnas.1513899113.

Heel, M. v. (2014) *Principles of Phase Contrast (Electron) Microscopy*.

Herbert, K. M., Greenleaf, W. J. & Block, S. M. (2008) Single-Molecule Studies of RNA Polymerase: Motoring Along RNAP: RNA polymerase. *Annu Rev Biochem*. doi: 10.1146/annurev.biochem.77.073106.100741.

Hirschman, J., Wong, P. K., Sei, K., Keener, J. & Kustu, S. (1985) Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc Natl Acad Sci U S A*. 82 (22), 7525-9. doi: 10.1073/pnas.82.22.7525.

Hsu, L. M. (2009) Monitoring abortive initiation. *Methods*. 47 (1), 25-36. doi: 10.1016/j.ymeth.2008.10.010.

- Hsu, L. M., Vo, N. V. & Chamberlin, M. J. (1995) *Escherichia coli* transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proc Natl Acad Sci U S A*. 92 (25), 11588-92. doi: 10.1073/pnas.92.25.11588.
- Huang, R. C., Maheshwari, N. & Bonner, J. (1960) Enzymatic synthesis of RNA. *Biochem Biophys Res Commun*. 3, 689-94. doi: 10.1016/0006-291x(60)90088-7.
- Hubin, E. A., Fay, A., Xu, C., Bean, J. M., Saecker, R. M., Glickman, M. S., Darst, S. A. & Campbell, E. A. (2017) Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife*. 6. doi: 10.7554/eLife.22520.
- Hunt, T. P. & Magasanik, B. (1985) Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. *Proc Natl Acad Sci U S A*. 82 (24), 8453-7. doi: 10.1073/pnas.82.24.8453.
- Hurwitz, J., Bresler, A. & Diring, R. (1960) The enzymic incorporation of ribonucleotides into polyribonucleotides and the effect of DNA. *Biochemical and Biophysical Research Communications*. 3 (1), 15-19. doi: 10.1016/0006-291x(60)90094-2.
- Imashimizu, M., Tanaka, K. & Shimamoto, N. (2011) Comparative Study of Cyanobacterial and *E. coli* RNA Polymerases: Misincorporation, Abortive Transcription, and Dependence on Divalent Cations. *Genet Res Int*. 2011, 572689. doi: 10.4061/2011/572689.
- Jessop, M., Felix, J. & Gutsche, I. (2021) AAA+ ATPases: structural insertions under the magnifying glass. *Curr Opin Struct Biol*. 66, 119-128. doi: 10.1016/j.sbi.2020.10.027.

Jishage, M., Iwata, A., Ueda, S. & Ishihama, A. (1996) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol.* 178 (18), 5447-51. doi: 10.1128/jb.178.18.5447-5451.1996.

Joly, N. & Buck, M. (2011) Single Chain Forms of the Enhancer Binding Protein PspF Provide Insights into Geometric Requirements for Gene Activation*. *J Biol Chem: Vol. 14.* pp. 12734-42.

Joly, N., Burrows, P. C. & Buck, M. (2008) An intramolecular route for coupling ATPase activity in AAA+ proteins for transcription activation. *J Biol Chem.* 283 (20), 13725-35. doi: 10.1074/jbc.M800801200.

Joly, N., Schumacher, J. & Buck, M. (2006) Heterogeneous nucleotide occupancy stimulates functionality of phage shock protein F, an AAA+ transcriptional activator. *J Biol Chem.* 281 (46), 34997-5007. doi: 10.1074/jbc.M606628200.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature.* 596 (7873), 583-589. doi: 10.1038/s41586-021-03819-2.

- Jun, S. H., Hirata, A., Kanai, T., Santangelo, T. J., Imanaka, T. & Murakami, K. S. (2014) The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat Commun.* 5, 5132. doi: 10.1038/ncomms6132.
- Kang, J. Y., Mishanina, T. V., Bao, Y., Chen, J., Llewellyn, E., Liu, J., Darst, S. A. & Landick, R. (2022) An ensemble of interconverting conformations of the elemental paused transcription complex creates regulatory options. *bioRxiv*. doi: 10.1101/2022.09.11.507475.
- Kang, J. Y., Mishanina, T. V., Bellecourt, M. J., Mooney, R. A., Darst, S. A. & Landick, R. (2018a) RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell*. 69 (5), 802-815 e5. doi: 10.1016/j.molcel.2018.01.018.
- Kang, J. Y., Mishanina, T. V., Landick, R. & Darst, S. A. (2019) Mechanisms of Transcriptional Pausing in Bacteria. *J Mol Biol*. 431 (20), 4007-4029. doi: 10.1016/j.jmb.2019.07.017.
- Kang, J. Y., Mooney, R. A., Nedialkov, Y., Saba, J., Mishanina, T. V., Artsimovitch, I., Landick, R. & Darst, S. A. (2018b) Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell*. 173 (7), 1650--1662.e14. doi: 10.1016/j.cell.2018.05.017.
- Kang, J. Y., Olinares, P. D. B., Chen, J., Campbell, E. A., Mustaev, A., Chait, B. T., Gottesman, M. E. & Darst, S. A. (2017) Structural basis of transcription arrest by coliphage HK022 N_{un} in an Escherichia coli RNA polymerase elongation complex. doi:10.7554/eLife.25478.

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. & Ebright, R. H.

(2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*. 314 (5802), 1144-7. doi: 10.1126/science.1131399.

Kater, L., Wagener, N., Berninghausen, O., Becker, T., Neupert, W. & Beckmann, R.

(2020) Structure of the Bcs1 AAA-ATPase suggests an airlock-like translocation mechanism for folded proteins. *Nat Struct Mol Biol*. 27 (2), 142-149. doi: 10.1038/s41594-019-0364-1.

Kettenberger, H., Armache, K. -J. & Cramer, P. (2003) Architecture of the RNA

Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. *Cell*. 114 (3), 347-357. doi: 10.1016/s0092-8674(03)00598-1.

Kettenberger, H., Armache, K. J. & Cramer, P. (2004) Complete RNA polymerase II

elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*. 16 (6), 955-65. doi: 10.1016/j.molcel.2004.11.040.

Khan, Y. A., White, K. I. & Brunger, A. T. (2022) The AAA+ superfamily: a review of the

structural and mechanistic principles of these molecular machines. *Crit Rev Biochem Mol Biol*. 57 (2), 156-187. doi: 10.1080/10409238.2021.1979460.

Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. (2021) New tools

for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J*. 478 (24), 4169-4185. doi: 10.1042/BCJ20210708.

Kohl, H. & Reimer, L. (2008) *Transmission Electron Microscopy. Springer Series in*

Optical Sciences. 5 edn., Springer New York, NY.

Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. (2014) Quantifying the local resolution of cryo-EM density maps. *Nat Methods*. 11 (1), 63-5. doi: 10.1038/nmeth.2727.

Kurkela, J., Fredman, J., Salminen, T. A. & Tyystjarvi, T. (2021) Revealing secrets of the enigmatic omega subunit of bacterial RNA polymerase. *Mol Microbiol*. 115 (1), 1-11. doi: 10.1111/mmi.14603.

Lamour, V., Hogan, B. P., Erie, D. A. & Darst, S. A. (2006) Crystal structure of *Thermus aquaticus* Gfh1, a Gre-factor paralog that inhibits rather than stimulates transcript cleavage. *J Mol Biol*. 356 (1), 179-88. doi: 10.1016/j.jmb.2005.10.083.

Landick, R. (2021) Transcriptional Pausing as a Mediator of Bacterial Gene Regulation. *Annu Rev Microbiol*. 75, 291-314. doi: 10.1146/annurev-micro-051721-043826.

Lane, W. J. & Darst, S. A. (2010) Molecular evolution of multisubunit RNA polymerases: structural analysis. *J Mol Biol*. 395 (4), 686-704. doi: 10.1016/j.jmb.2009.10.063.

Larson, M. H., Mooney, R. A., Peters, J. M., Windgassen, T., Nayak, D., Gross, C. A., Block, S. M., Greenleaf, W. J., Landick, R. & Weissman, J. S. (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*. 344 (6187), 1042-7. doi: 10.1126/science.1251871.

Lee, J. & Borukhov, S. (2016) Bacterial RNA Polymerase-DNA Interaction-The Driving Force of Gene Expression and the Target for Drug Action. *Front Mol Biosci*. 3, 73. doi: 10.3389/fmolb.2016.00073.

Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*. 157 (3), 624-35. doi: 10.1016/j.cell.2014.02.033.

Li, L., Fang, C., Zhuang, N., Wang, T. & Zhang, Y. (2019) Structural basis for transcription initiation by bacterial ECF sigma factors. *Nat Commun*. 10 (1), 1153. doi: 10.1038/s41467-019-09096-y.

Li, L., Molodtsov, V., Lin, W., Ebright, R. H. & Zhang, Y. (2020) RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. doi: 10.1073/pnas.1920747117.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A. & Cheng, Y. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods*. 10 (6), 584-90. doi: 10.1038/nmeth.2472.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*. 75 (Pt 10), 861-877. doi: 10.1107/S2059798319011471.

Lilic, M., Darst, S. A. & Campbell, E. A. (2021) Structural basis of transcriptional activation by the *Mycobacterium tuberculosis* intrinsic antibiotic-resistance transcription factor WhiB7. *Mol Cell*. 81 (14), 2875-2886 e5. doi: 10.1016/j.molcel.2021.05.017.

Lin, H., Deng, E. Z., Ding, H., Chen, W. & Chou, K. C. (2014) iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition. *Nucleic Acids Res*. 42 (21), 12961-72. doi: 10.1093/nar/gku1019.

Lin, J., Shorter, J. & Lucius, A. L. (2022) AAA+ proteins: one motor, multiple ways to work. *Biochem Soc Trans*. 50 (2), 895-906. doi: 10.1042/BST20200350.

Liu, B., Hong, C., Huang, R. K., Yu, Z. & Steitz, T. A. (2017) Structural basis of bacterial transcription activation. *Science*. 358 (6365), 947-951. doi: 10.1126/science.aao1923.

Liu, B., Zuo, Y. & Steitz, T. A. (2016) Structures of *E. coli* sigmaS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc Natl Acad Sci U S A*. 113 (15), 4051-6. doi: 10.1073/pnas.1520555113.

Liu, Y., Yu, L., Pukhrambam, C., Winkelman, J. T., Firlar, E., Kaelber, J. T., Zhang, Y., Nickels, B. E. & Ebright, R. H. (2022) Structural and mechanistic basis of reiterative transcription initiation. *Proc Natl Acad Sci U S A*. 119 (5). doi: 10.1073/pnas.2115746119.

Lo, Y. H., Sobhany, M., Hsu, A. L., Ford, B. L., Krahn, J. M., Borgnia, M. J. & Stanley, R. E. (2019) Cryo-EM structure of the essential ribosome assembly AAA-ATPase Rix7. *Nat Commun*. 10 (1), 513. doi: 10.1038/s41467-019-08373-0.

Maeda, H., Fujita, N. & Ishihama, A. (2000) Competition among seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res: Vol. 18*. pp. 3497-503.

Majumder, P., Rudack, T., Beck, F., Danev, R., Pfeifer, G., Nagy, I. & Baumeister, W. (2019) Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proc Natl Acad Sci U S A*. 116 (2), 534-539. doi: 10.1073/pnas.1817752116.

Malinen, A. M., Bakermans, J., Aalto-Setälä, E., Blessing, M., Bauer, D. L. V., Parilova, O., Belogurov, G. A., Dulin, D. & Kapanidis, A. N. (2022) Real-Time Single-Molecule Studies of RNA Polymerase-Promoter Open Complex Formation Reveal Substantial Heterogeneity Along the Promoter-Opening Pathway. *J Mol Biol*. 434 (2), 167383. doi: 10.1016/j.jmb.2021.167383.

Martinez-Fernandez, V., Garrido-Godino, A. I., Miron-Garcia, M. C., Begley, V., Fernandez-Pevida, A., de la Cruz, J., Chavez, S. & Navarro, F. (2018) Rpb5 modulates the RNA polymerase II transition from initiation to elongation by influencing Spt5 association and backtracking. *Biochim Biophys Acta Gene Regul Mech*. 1861 (1), 1-13. doi: 10.1016/j.bbagrm.2017.11.002.

Meka, H., Werner, F., Cordell, S. C., Onesti, S. & Brick, P. (2005) Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res*. 33 (19), 6435-44. doi: 10.1093/nar/gki945.

Milo, R. & Phillips, R. (2015) *Cell Biology by the Numbers*. CRC Press.

- Mohamed, A. A., Vazquez Nunez, R. & Vos, S. M. (2022) Structural advances in transcription elongation. *Curr Opin Struct Biol.* 75, 102422. doi: 10.1016/j.sbi.2022.102422.
- Mooney, R. A. & Landick, R. (2003) Tethering sigma70 to RNA polymerase reveals high in vivo activity of sigma factors and sigma70-dependent pausing at promoter-distal locations. *Genes Dev.* 17 (22), 2839-51. doi: 10.1101/gad.1142203.
- Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., Sarafianos, S., Tuske, S., Patel, J., Jansen, R., Irschik, H., Arnold, E. & Ebright, R. H. (2008) The RNA polymerase "switch region" is a target for inhibitors. *Cell.* 135 (2), 295-307. doi: 10.1016/j.cell.2008.09.033.
- Murakami, K. S. (2013) X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J Biol Chem.* 288 (13), 9126-34. doi: 10.1074/jbc.M112.430900.
- Murayama, S., Ishikawa, S., Chumsakul, O., Ogasawara, N. & Oshima, T. (2015) The Role of alpha-CTD in the Genome-Wide Transcriptional Regulation of the Bacillus subtilis Cells. *PLoS One.* 10 (7), e0131588. doi: 10.1371/journal.pone.0131588.
- Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehl, J., Uchanski, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R. & Scheres, S. H. W. (2020) Single-particle cryo-EM at atomic resolution. *Nature.* 587 (7832), 152-156. doi: 10.1038/s41586-020-2829-0.

- Naryshkina, T., Kuznedelov, K. & Severinov, K. (2006) The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA-DNA hybrid. *J Mol Biol.* 361 (4), 634-43. doi: 10.1016/j.jmb.2006.05.034.
- Niu, W., Kim, Y., Tau, G., Heyduk, T. & Ebright, R. H. (1996) Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell.* 87 (6), 1123-34. doi: 10.1016/s0092-8674(00)81806-1.
- Noe Gonzalez, M., Blears, D. & Svejstrup, J. Q. (2021) Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat Rev Mol Cell Biol.* 22 (1), 3-21. doi: 10.1038/s41580-020-00308-8.
- Nudler, E. (2009) RNA polymerase active center: the molecular engine of transcription. *Annu Rev Biochem.* 78, 335-61. doi: 10.1146/annurev.biochem.76.052705.164655.
- Orlova, E. V. & Saibil, H. R. (2011) Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev.* 111 (12), 7710-48. doi: 10.1021/cr100353t.
- Paget, M. S. (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules: Vol. 3.* pp. 1245-65.
- Park, J., Kang, M. & Kim, M. (2015) Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic Acids Res.* 43 (5), 2625-37. doi: 10.1093/nar/gkv133.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021) UCSF ChimeraX: Structure visualization for

researchers, educators, and developers. *Protein Sci.* 30 (1), 70-82. doi: 10.1002/pro.3943.

Puchades, C., Sandate, C. R. & Lander, G. C. (2020) The molecular principles governing the activity and functional diversity of AAA+ proteins. *Nat Rev Mol Cell Biol.* 21 (1), 43-58. doi: 10.1038/s41580-019-0183-6.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* 14 (3), 290-296. doi: 10.1038/nmeth.4169.

Rappas, M., Schumacher, J., Niwa, H., Buck, M. & Zhang, X. (2006) Structural basis of the nucleotide driven conformational changes in the AAA+ domain of transcription activator PspF. *J Mol Biol.* 357 (2), 481-92. doi: 10.1016/j.jmb.2005.12.052.

Ray-Soni, A., Bellecourt, M. J. & Landick, R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem.* 85, 319-47. doi: 10.1146/annurev-biochem-060815-014844.

Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. (2006) Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. doi: 10.1126/science.1131398.

Riley, P. A. (1970) A suggested mechanism for DNA transcription. *Nature.* 228 (5271), 522-5. doi: 10.1038/228522a0.

Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P. & Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase. *Nature*. 478 (7368), 269-73. doi: 10.1038/nature10435.

Roberts, M., Rowley, G., Kormanec, J. & Zalm, M. E. J. (2017) The Role of Alternative Sigma Factors in Pathogen Virulence. In: Gurtler, J.B., Doyle, M.P. & Kornacki, J.L. (eds.) *Foodborne Pathogens: Virulence Factors and Host Susceptibility*. Cham, Springer International Publishing, pp. 229-303.

Rohou, A. & Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*. 192 (2), 216-21. doi: 10.1016/j.jsb.2015.08.008.

Rosenthal, P. B. & Henderson, R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol*. 333 (4), 721-45. doi: 10.1016/j.jmb.2003.07.013.

Roston, D., Demapan, D. & Cui, Q. (2019) Extensive free-energy simulations identify water as the base in nucleotide addition by DNA polymerase. *Proceedings of the National Academy of Sciences*. 116 (50), 25048-25056. doi: 10.1073/pnas.1914613116.

Russo, C. J., Scotcher, S. & Kyte, M. (2016) A precision cryostat design for manual and semi-automated cryo-plunge instruments. *Rev Sci Instrum*. 87 (11), 114302. doi: 10.1063/1.4967864.

Saba, J., Chua, X. Y., Mishanina, T. V., Nayak, D., Windgassen, T. A., Mooney, R. A. & Landick, R. (2019) The elemental mechanism of transcriptional pausing. *Elife*. 8. doi: 10.7554/eLife.40981.

Sadian, Y., Baudin, F., Tafur, L., Murciano, B., Wetzel, R., Weis, F. & Muller, C. W. (2019) Molecular insight into RNA polymerase I promoter recognition and promoter melting. *Nat Commun*. 10 (1), 5543. doi: 10.1038/s41467-019-13510-w.

Scheres, S. H. (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*. Vol. 3. pp. 519-30.

Scheres, S. H. (2016) Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol*. 579, 125-57. doi: 10.1016/bs.mie.2016.04.012.

Schillbach, S., Aibara, S., Dienemann, C., Grabbe, F. & Cramer, P. (2021) Structure of RNA polymerase II pre-initiation complex at 2.9 Å defines initial DNA opening. *Cell*. 184 (15), 4064-4072 e28. doi: 10.1016/j.cell.2021.05.012.

Schumacher, J., Zhang, X., Jones, S., Bordes, P. & Buck, M. (2004) ATP-dependent transcriptional activation by bacterial PspF AAA+protein. *J Mol Biol*. 338 (5), 863-75. doi: 10.1016/j.jmb.2004.02.071.

Seraphim, T. V. & Houry, W. A. (2020) AAA+ proteins. *Curr Biol*. 30 (6), R251-R257. doi: 10.1016/j.cub.2020.01.044.

Shi, W., Jiang, Y., Deng, Y., Dong, Z. & Liu, B. (2020a) Visualization of two architectures in class-II CAP-dependent transcription activation. *PLoS Biol*. 18 (4), e3000706. doi: 10.1371/journal.pbio.3000706.

Shi, W., Zhou, W., Zhang, B., Huang, S., Jiang, Y., Schammel, A., Hu, Y. & Liu, B.

(2020b) Structural basis of bacterial sigma(28) -mediated transcription reveals roles of the RNA polymerase zinc-binding domain. *EMBO J.* 39 (14), e104389. doi:

10.15252/embj.2020104389.

Shin, Y., Qayyum, M. Z., Pupov, D., Esyunina, D., Kulbachinskiy, A. & Murakami, K. S.

(2021) Structural basis of ribosomal RNA transcription regulation. *Nat Commun.* 12 (1), 528. doi: 10.1038/s41467-020-20776-y.

Skene, P. J., Hernandez, A. E., Groudine, M. & Henikoff, S. (2014) The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife.* 3, e02042. doi: 10.7554/eLife.02042.

Spitz, F. & Furlong, E. E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet.* 13 (9), 613-26. doi: 10.1038/nrg3207.

Steitz, T. A. (1998) A mechanism for all polymerases. *Nature.* 391 (6664), 231-2. doi: 10.1038/34542.

Stepanova, E. V., Shevelev, A. B., Borukhov, S. I. & Severinov, K. V. (2009)

Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA. *Biofizika.* 54 (5), 773-90.

Stevens, A. (1960) Incorporation of the adenine ribonucleotide into RNA by cell

fractions from *E. coli* B. *Biochemical and Biophysical Research Communications.* 3 (1), 92-96. doi: 10.1016/0006-291x(60)90110-8.

Sutherland, C. & Murakami, K. S. (2018) *An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. EcoSal/Plus.* 8 (1). doi: 10.1128/ecosalplus.ESP-0004-2018.

Svetlov, V. & Nudler, E. (2009) *Macromolecular micromovements: how RNA polymerase translocates. Curr Opin Struct Biol.* 19 (6), 701-7. doi: 10.1016/j.sbi.2009.10.002.

Sydow, J. F., Brueckner, F., Cheung, A. C., Damsma, G. E., Dengl, S., Lehmann, E., Vassylyev, D. & Cramer, P. (2009) *Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol Cell.* 34 (6), 710-21. doi: 10.1016/j.molcel.2009.06.002.

Sydow, J. F. & Cramer, P. (2009) *RNA polymerase fidelity and transcriptional proofreading. Curr Opin Struct Biol.* 19 (6), 732-9. doi: 10.1016/j.sbi.2009.10.009.

Syed, A. & Gralla, J. D. (1997) *Isolation and properties of enhancer-bypass mutants of sigma 54. Mol Microbiol.* 23 (5), 987-95. doi: 10.1046/j.1365-2958.1997.2851651.x.

Sysoeva, T. A., Chowdhury, S., Guo, L. & Nixon, B. T. (2013) *Nucleotide-induced asymmetry within ATPase activator ring drives sigma54-RNAP interaction and ATP hydrolysis. Genes Dev.* 27 (22), 2500-11. doi: 10.1101/gad.229385.113.

Tan, L., Wiesler, S., Trzaska, D., Carney, H. C. & Weinzierl, R. O. (2008) *Bridge helix and trigger loop perturbations generate superactive RNA polymerases. J Biol.* 7 (10), 40. doi: 10.1186/jbiol98.

Tiwari, P., Kaila, P. & Guptasarma, P. (2019) Understanding anomalous mobility of proteins on SDS-PAGE with special reference to the highly acidic extracellular domains of human E- and N-cadherins. *Electrophoresis*. 40 (9), 1273-1281. doi:

10.1002/elps.201800219.

Toulme, F., Mosrin-Huaman, C., Sparkowski, J., Das, A., Leng, M. & Rahmouni, A. R. (2000) GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* 19 (24), 6853-9. doi:

10.1093/emboj/19.24.6853.

Turowski, T. W. & Boguta, M. (2021) Specific Features of RNA Polymerases I and III: Structure and Assembly. *Front Mol Biosci.* 8, 680090. doi:

10.3389/fmolb.2021.680090.

van Eeuwen, T., Li, T., Kim, H. J., Gorbea Colon, J. J., Parker, M. I., Dunbrack, R. L., Garcia, B. A., Tsai, K. L. & Murakami, K. (2021) Structure of TFIIK for phosphorylation of CTD of RNA polymerase II. *Sci Adv.* 7 (15). doi: 10.1126/sciadv.abd4420.

Vassilyev, D. G., Vassilyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. (2007) Structural basis for transcription elongation by bacterial RNA polymerase.

Nature. 448 (7150), 157-62. doi: 10.1038/nature05932.

Vassilyeva, M. N., Svetlov, V., Dearborn, A. D., Klyuyev, S., Artsimovitch, I. & Vassilyev, D. G. (2007) The carboxy-terminal coiled-coil of the RNA polymerase beta'-subunit is the main binding site for Gre factors. *EMBO Rep.* 8 (11), 1038-43. doi:

10.1038/sj.embor.7401079.

Vishwakarma, R. K., Qayyum, M. Z., Babitzke, P. & Murakami, K. S. (2022) Allosteric mechanism of transcription inhibition by NusG-dependent pausing of RNA polymerase. *bioRxiv*. doi: 10.1101/2022.10.28.514324.

von Loeffelholz, O. & Klaholz, B. P. (2021) Setup and Troubleshooting of Volta Phase Plate Cryo-EM Data Collection. *Methods Mol Biol*. 2305, 291-299. doi: 10.1007/978-1-0716-1406-8_14.

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H. & Cramer, P. (2018a) Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature*. 560 (7720), 607-612. doi: 10.1038/s41586-018-0440-4.

Vos, S. M., Farnung, L., Linden, A., Urlaub, H. & Cramer, P. (2020) Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat Struct Mol Biol*. 27 (7), 668-677. doi: 10.1038/s41594-020-0437-1.

Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. (2018b) Structure of paused transcription complex Pol II-DSIF-NELF. *Nature*. 560 (7720), 601-606. doi: 10.1038/s41586-018-0442-2.

Voutsina, A., Riva, M., Carles, C. & Alexandraki, D. (1999) Sequence divergence of the RNA polymerase shared subunit ABC14.5 (Rpb8) selectively affects RNA polymerase III assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 27 (4), 1047-55. doi: 10.1093/nar/27.4.1047.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C. & Raunser, S. (2019) SPHERE-

crYOLO is a fast and accurate fully automated particle picker for cryo-EM.

Communications Biology. 2 (1), 218. doi: 10.1038/s42003-019-0437-z.

Wald, J., Fahrenkamp, D., Goessweiner-Mohr, N., Lugmayr, W., Ciccarelli, L., Vesper, O. & Marlovits, T. C. (2022) Mechanism of AAA+ ATPase-mediated RuvAB-Holliday junction branch migration. *Nature.* 609 (7927), 630-639. doi: 10.1038/s41586-022-05121-1.

Wang, B. & Artsimovitch, I. (2020) NusG, an Ancient Yet Rapidly Evolving Transcription Factor. *Front Microbiol.* 11, 619618. doi: 10.3389/fmicb.2020.619618.

Wang, B., Sexton, R. E. & Feig, M. (2017) Kinetics of nucleotide entry into RNA polymerase active site provides mechanism for efficiency and fidelity. *Biochim Biophys Acta Gene Regul Mech.* 1860 (4), 482-490. doi: 10.1016/j.bbagr.2017.02.008.

Wang, D., Bushnell, D. A., Huang, X., Westover, K. D., Levitt, M. & Kornberg, R. D. (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science.* 324 (5931), 1203-6. doi: 10.1126/science.1168729.

Wang, F., Redding, S., Finkelstein, I. J., Gorman, J., Reichman, D. R. & Greene, E. C. (2013) The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion. *Nat Struct Mol Biol.* 20 (2), 174-81. doi: 10.1038/nsmb.2472.

Wang, H., Schilbach, S., Ninov, M., Urlaub, H. & Cramer, P. (2022) Structures of transcription preinitiation complex engaged with the +1 nucleosome. *Nat Struct Mol Biol.* doi: 10.1038/s41594-022-00865-w.

Wang, Y., Severinov, K., Loizos, N., Fenyő, D., Heyduk, E., Heyduk, T., Chait, B. T. & Darst, S. A. (1997) Determinants for *Escherichia coli* RNA polymerase assembly within the β subunit. *Journal of Molecular Biology*. 270 (5), 648-662. doi:

10.1006/jmbi.1997.1139.

Wang, Z., Zhao, S., Li, Y., Zhang, K., Mo, F., Zhang, J., Hou, Y., He, L., Liu, Z., Wang, Y., Xu, Y., Wang, H., Buck, M., Matthews, S. J. & Liu, B. (2021) RssB-mediated sigma(S) Activation is Regulated by a Two-Tier Mechanism via Phosphorylation and Adaptor Protein - IraD. *J Mol Biol*. 433 (3), 166757. doi: 10.1016/j.jmb.2020.166757.

Watson, J. D. & Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 171 (4356), 737-8. doi: 10.1038/171737a0.

Webster, M. W. & Weixlbaumer, A. (2021) Macromolecular assemblies supporting transcription-translation coupling. *Transcription*. 12 (4), 103-125. doi:

10.1080/21541264.2021.1981713.

Weiss, S. B. & Gladstone, L. (1959) A Mammalian System for the Incorporation of Cytidine Triphosphate into Ribonucleic Acid¹. *Journal of the American Chemical Society*. 81 (15), 4118-4119. doi: 10.1021/ja01524a087.

Weissenberger, G., Henderikx, R. J. M. & Peters, P. J. (2021) Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods*. 18 (5), 463-471. doi:

10.1038/s41592-021-01130-6.

Werner, F. & Grohmann, D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology*. 9 (2), 85. doi:

doi:10.1038/nrmicro2507.

Wilkins, M. H., Stokes, A. R. & Wilson, H. R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature*. 171 (4356), 738-40. doi: 10.1038/171738a0.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*. 27 (1), 293-315. doi: 10.1002/pro.3330.

Williams, D. B. & Carter, C. B. (2009) Pumps and Holders. *Transmission Electron Microscopy*. pp. 127-140.

Winkelman, J. T., Nickels, B. E. & Ebright, R. H. (2021) CHAPTER 1. The Transition from Transcription Initiation to Transcription Elongation: Start-site Selection, Initial Transcription, and Promoter Escape. *RNA Polymerases as Molecular Motors: Chemical Biology*. pp. 1-24.

Wong, C., Tintut, Y. & Gralla, J. D. (1994) The domain structure of sigma 54 as determined by analysis of a set of deletion mutants. *J Mol Biol*. 236 (1), 81-90. doi: 10.1006/jmbi.1994.1120.

Wong, H. C., Chen, J., Mouche, F., Rouiller, I. & Bern, M. (2004) Model-based particle picking for cryo-electron microscopy. *J Struct Biol*. 145 (1-2), 157-67. doi: 10.1016/j.jsb.2003.05.001.

Wu, J. G., Yan, Y., Zhang, D. X., Liu, B. W., Zheng, Q. B., Xie, X. L., Liu, S. Q., Ge, S. X., Hou, Z. G. & Xia, N. S. (2022) Machine Learning for Structure Determination in

Single-Particle Cryo-Electron Microscopy: A Systematic Review. *IEEE Trans Neural Netw Learn Syst.* 33 (2), 452-472. doi: 10.1109/TNNLS.2021.3131325.

Xie, J., Libri, D. & Porrua, O. (2023) Mechanisms of eukaryotic transcription termination at a glance. *J Cell Sci.* 136 (1). doi: 10.1242/jcs.259873.

Yakhnin, A. V., FitzGerald, P. C., McIntosh, C., Yakhnin, H., Kireeva, M., Turek-Herman, J., Mandell, Z. F., Kashlev, M. & Babitzke, P. (2020) NusG controls transcription pausing and RNA polymerase translocation throughout the *Bacillus subtilis* genome. *Proc Natl Acad Sci U S A.* 117 (35), 21628-21636. doi: 10.1073/pnas.2006873117.

Yamamoto, K., Yamanaka, Y., Shimada, T., Sarkar, P., Yoshida, M., Bhardwaj, N., Watanabe, H., Taira, Y., Chatterji, D. & Ishihama, A. (2018) Altered Distribution of RNA Polymerase Lacking the Omega Subunit within the Prophages along the *Escherichia coli* K-12 Genome. *mSystems: Vol. 1.*

Yang, C., Fujiwara, R., Kim, H. J., Basnet, P., Zhu, Y., Gorbea Colon, J. J., Steimle, S., Garcia, B. A., Kaplan, C. D. & Murakami, K. (2022) Structural visualization of de novo transcription initiation by *Saccharomyces cerevisiae* RNA polymerase II. *Mol Cell.* 82 (3), 660-676 e9. doi: 10.1016/j.molcel.2021.12.020.

Yang, Y., Darbari, V. C., Zhang, N., Lu, D., Glyde, R., Wang, Y.-P., Winkelman, J. T., Gourse, R. L., Murakami, K. S., Buck, M. & Zhang, X. (2015) Structures of the RNA polymerase-s54 reveal new and conserved regulatory strategies. *Science.* 349 (6250), 882--886. doi: 10.1126/science.aab1478.Structures.

Yang, Y., Liu, C., Zhou, W., Shi, W., Chen, M., Zhang, B., Schatz, D. G., Hu, Y. & Liu, B. (2021) Structural visualization of transcription activated by a multidrug-sensing MerR family regulator. *Nat Commun.* 12 (1), 2702. doi: 10.1038/s41467-021-22990-8.

Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. (2012) Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure.* 20 (2), 237-47. doi: 10.1016/j.str.2011.12.007.

Yao, R. W., Wang, Y. & Chen, L. L. (2019) Cellular functions of long noncoding RNAs. *Nat Cell Biol.* 21 (5), 542-551. doi: 10.1038/s41556-019-0311-8.

Ye, F., Gao, F., Liu, X., Buck, M. & Zhang, X. (2022) Mechanisms of DNA opening revealed in AAA+ transcription complex structures. *Sci Adv.* 8 (51), eadd3479. doi: 10.1126/sciadv.add3479.

Ye, F., Kotta-Loizou, I., Jovanovic, M., Liu, X., Dryden, D. T., Buck, M. & Zhang, X. (2020) Structural basis of transcription inhibition by the DNA mimic protein Ocr of bacteriophage T7. *Elife.* 9. doi: 10.7554/eLife.52125.

Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature.* 587 (7832), 157-161. doi: 10.1038/s41586-020-2833-4.

Zaros, C., Briand, J. F., Boulard, Y., Labarre-Mariotte, S., Garcia-Lopez, M. C., Thuriaux, P. & Navarro, F. (2007) Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucleic Acids Res.* 35 (2), 634-47. doi: 10.1093/nar/gkl686.

Zenkin, N., Yuzenkova, Y. & Severinov, K. (2006) Transcript-assisted transcriptional proofreading. *Science*. 313 (5786), 518-20. doi: 10.1126/science.1127422.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999) Crystal Structure of *Thermus aquaticus* Core RNA Polymerase at 3.3 Å Resolution. *Cell*. 98 (6), 811-824. doi: 10.1016/s0092-8674(00)81515-9.

Zhang, K., Pintilie, G. D., Li, S., Schmid, M. F. & Chiu, W. (2020) Resolving individual atoms of protein complex by cryo-electron microscopy. *Cell Res*. 30 (12), 1136-1139. doi: 10.1038/s41422-020-00432-2.

Zhang, N. & Buck, M. (2015) A perspective on the enhancer dependent bacterial RNA polymerase. *Biomolecules*. 5 (2), 1012--1019. doi: 10.3390/biom5021012.

Zhang, N., Joly, N. & Buck, M. (2012) A common feature from different subunits of a homomeric AAA+ protein contacts three spatially distinct transcription elements. *Nucleic Acids Res*. 40 (18), 9139-52. doi: 10.1093/nar/gks661.

Zhang, N., Joly, N., Burrows, P. C., Jovanovic, M., Wigneshweraraj, S. R. & Buck, M. (2009) The role of the conserved phenylalanine in the sigma54-interacting GAFTGA motif of bacterial enhancer binding proteins. *Nucleic Acids Res*. 37 (18), 5981-92. doi: 10.1093/nar/gkp658.

Zhang, N., Simpson, T., Lawton, E., Uzdavinyas, P., Joly, N., Burrows, P. & Buck, M. (2013) A Key Hydrophobic Patch Identified in an AAA+ Protein Essential for Its In Trans Inhibitory Regulation. *J Mol Biol*. 425 (15), 2656-69. doi: 10.1016/j.jmb.2013.04.024.

- Zhang, S., Aibara, S., Vos, S. M., Agafonov, D. E., Luhrmann, R. & Cramer, P. (2021) Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science*. 371 (6526), 305-309. doi: 10.1126/science.abf1870.
- Zhang, X. & Wigley, D. B. (2008) The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat Struct Mol Biol*. 15 (11), 1223-7. doi: 10.1038/nsmb.1501.
- Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E. & Ebright, R. H. (2012) Structural basis of transcription initiation. *Science*. 338 (6110), 1076-80. doi: 10.1126/science.1227786.
- Zhao, K., Liu, M. & Burgess, R. R. (2010) Promoter and regulon analysis of nitrogen assimilation factor, sigma54, reveal alternative strategy for *E. coli* MG1655 flagellar biosynthesis. *Nucleic Acids Res*. 38 (4), 1273-83. doi: 10.1093/nar/gkp1123.
- Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y. & Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*. 14 (4), 331-332. doi: 10.1038/nmeth.4193.
- Zhu, C., Guo, X., Dumas, P., Takacs, M., Abdelkareem, M., Vanden Broeck, A., Saint-Andre, C., Papai, G., Crucifix, C., Ortiz, J. & Weixlbaumer, A. (2022) Transcription factors modulate RNA polymerase conformational equilibrium. *Nat Commun*. 13 (1), 1546. doi: 10.1038/s41467-022-29148-0.
- Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. & Scheres, S. H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. 7. doi: 10.7554/eLife.42166.

Zuo, Y., De, S., Feng, Y. & Steitz, T. A. (2020) Structural Insights into Transcription Initiation from De Novo RNA Synthesis to Transitioning into Elongation. *iScience*. 23 (9), 101445. doi: 10.1016/j.isci.2020.101445.

Appendix

Supplementary Table 1: Chromatography methods used in purification procedures

Column Name	Type	Maximum Pre-column Pressure (mPa)	Column Volume (ml)	Sample Application type	Flow Rate (ml/min)				Column Volumes for each method phase					Elution Gradient (%)		
					Equilibration	Sample Application	Column Wash	Elution	Equilibration	Column Wash	Elution			Start	End	Type
HisTrap™ HP	Nickel Affinity	0.5	5	Direct	5	2	3	4	5	12	5	15	5	10	100	Linear
HiTrap™ Heparin HP	Heparin Affinity	0.5	5	Direct	4	2	2	2	5	6	3	6	5	4	100	Linear
Bio-Rex™ 70	Cation Exchange	0.3	30	Direct	1	1	1	1	2	3	1.5	3.5	2	0	100	Linear
HiTrap™ Q HP	Anion Exchange	0.5	5	Direct	4	2	2	2	5	6	3	6	5	4	100	Linear
HiPrep™ 26/10 Desalting	Desalting/Size Exclusion	0.5	53	Direct	1	5	N/A	5	1.5	N/A	N/A	1.5	N/A	N/A	N/A	Isocratic
Superdex Hiload® 75 16/600	Size Exclusion	0.5	120	Injection (10ml loop)	1	1	N/A	1	1.5	N/A	N/A	1.1	N/A	N/A	N/A	Isocratic
Superose 6 10/300 Increase	Size Exclusion	2	24	Injection (500µl loop)	0.35	0.35	N/A	0.35	3	N/A	N/A	1.2	N/A	N/A	N/A	Isocratic

Supplementary Table 2: Model validation statistics for intermediate complexes

Sample	RP(-11/-8) conformation 1	RP(-11/-8) conformation 2	RP(-11/-8) conformation 3	RP(-11/-8) conformation 4	RP(-11/-8) conformation 5	RP(-11/-8) conformation 6	RP(- 10/-1)	pre-RPo state
Non-hydrogen atoms	26960	26819	26953	26928	26832	26903	26931	19635
Protein residues	5154	5125	5153	5148	5128	5143	5165	3581
Nucleic acid residues	68	68	68	68	68	68	64	87
Ligands	5 Mg ²⁺	5 Mg ²⁺	5 Mg ²⁺	5 Mg ²⁺	5 Mg ²⁺	5 Mg ²⁺	5 Mg ²⁺	1 Mg ²⁺
	5 ADP	5 ADP	5 ADP	5 ADP	5 ADP	5 ADP	5 ADP	
	5 AF3	5 AF3	5 AF3	5 AF3	5 AF3	5 AF3	5 AF3	
B factors (Å ²)								
Protein	191.59	273.75	225.37	397.12	595.25	882.62	349.09	105.54
Nucleic acid	461.16	289.8	438.31	640.15	235.94	284.5	385.72	296.67
Ligands	121.29	563.98	171.92	585.45	307.84	706.08	344.07	326.69
R.m.s. deviations								
Bond lengths (Å)	0.007	0.002	0.003	0.01	0.002	0.011	0.002	0.004
Bond angles (°)	1.032	0.641	0.655	0.758	0.653	0.904	0.543	0.857
MolProbity score	2.52	1.52	1.68	2.11	2.09	2.42	1.83	2.09
Clashscore	23.55	3.53	4.7	11.71	10.91	19.19	6.79	8.4
Poor rotamers (%)	-	-	-	-	-	-	-	-
Ramachandran plot								
Favored (%)	85	93.84	93.31	90.73	90.57	0.2	92.78	86.1
Disallowed (%)	0.43	0.04	0	0.12	0.08	86	0.02	0.84

Supplementary Table 3: Model validation statistics for longer transcript complexes

Sample	RP5nt pre-translocated	RP5nt post-translocated	RP6nt	RP7nt	RP8nt	RP9nt
Refinement						
Model composition						
Non-hydrogen atoms	29164	28821	28681	28965	28878	28605
Protein residues	3668	3668	3638	3635	3617	3617
Nucleic acid residues	102	100	101	104	105	103
Ligands	2 Zn ²⁺	2 Zn ²⁺	2 Zn ²⁺	2 Zn ²⁺	2 Zn ²⁺	2 Zn ²⁺
	1 Mg ²⁺	1 Mg ²⁺	1 Mg ²⁺	1 Mg ²⁺	1 Mg ²⁺	1 Mg ²⁺
B factors (Å ²)						
Protein	47.31	40.32	63.05	50.02	42	37.68
Nucleic acid	97.46	95.31	101.33	101.62	156.55	180.21
Ligands	58.23	27.99	58.31	54.39	33.57	46.53
R.m.s. deviations						
Bond lengths (Å)	0.004	0.003	0.005	0.005	0.005	0.003
Bond angles (°)	0.672	0.604	0.71	0.765	0.681	0.643
Validation						
MolProbity score	2.29	1.85	1.99	1.96	1.95	1.95
Clashscore	8.32	8.71	10.84	11.06	10.12	11.03
Poor rotamers (%)	1.66	0.27	0.42	0.37	0.37	0.34
Ramachandran plot						
Favored (%)	94.21	94.37	93.3	94.13	93.59	94.28
Disallowed (%)	0.05	0.03	0.03	0.03	0.03	0.08

Supplementary Movie 1: A morph movie showing the monomer conformational changes between protomers

Supplementary Movie 2-3: A morph movie showing the proposed conformational changes within the hexameric ring, top and side views

Supplementary Movie 4-6: A morph movie showing the conformational changes from RP(-12/-11) to RP(-11-8) conformation 3 to RP(-10-1) within the hexameric ring and DNA, top and side views

Supplementary Movie 7: A morph movie showing the conformational changes in region II and ELH upon extension of RNA from 5nt pretranslocated to 9nt.