NORTHWESTERN UNIVERSITY

Cryogenic-Electron Microscopy Structures of Viral and Eukaryotic Transcription Complexes

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences

By

Ryan Howard Abdella

EVANSTON, ILLINOIS

March 2021

ABSTRACT

Cryogenic-Electron Microscopy Structures of Viral and Eukaryotic Transcription Complexes

Ryan Howard Abdella

Transcription of nucleic acids (DNA and RNA) is a hallmark of life, taking the information stored within genomic nucleic acids and converting it into a form that is useful for producing the proteins necessary for cellular and organismal function. In eukaryotes, transcription of DNA into messenger RNA (mRNA) requires the formation of a 56-subunit pre-initiation complex comprised of RNA Polymerase II (Pol II), the co-activator Mediator, and a group of general transcription factors. Mediator facilitates the assembly of this complex at gene promoters and stimulates phosphorylation of the Pol II C-terminal domain (CTD) by CDK7, a subunit of the cyclin-activated kinase (CAK) module of TFIIH. To understand the mechanism of this process, I have used cryo-electron microscopy to solve the structure of the human Mediator-bound PIC to sub-4 Å. Transcription factor binding sites within Mediator are primarily flexibly tethered to the tail module. CDK7 is stabilized by multiple contacts with Mediator. Two binding sites exist for the Pol II CTD, one between the head and middle modules of Mediator, and the other in the active site of CDK7, suggesting the former helps position the latter.

Paramyxoviruses are enveloped, non-segmented, negative-strand (NNS) RNA viruses that cause a broad spectrum of human and animal diseases. The viral genome, packaged by the nucleoprotein (N), serves as a template for the polymerase complex, composed of the large protein (L) and the homo-tetrameric phosphoprotein (P). The ~250 kDa L possesses all enzymatic activities necessary for its function but requires P *in vivo*. Structural information is available for individual P domains from different paramyxoviruses, but how P interacts with L and how that affects the activity of L is mostly unknown due to the lack of high-resolution structures of this

complex in this viral family. In this study, I determined the structure of the L-P complex from parainfluenza virus 5 (PIV5) at 4.3 Å resolution using cryo-electron microscopy. P-OD binds to the RNA-dependent RNA polymerase (RdRp) domain of L and protrudes away from it, while the X domain (XD) of one chain of P is bound near the L nucleotide entry site. The methyltransferase (MTase) domain and the C-terminal domain (CTD) of L adopt a novel conformation, positioning the MTase active site immediately above the poly-ribonucleotidyltransferase (PRNTase) domain and near the likely exit site for the product RNA 5' end. Our study reveals a potential mechanism that mononegavirus polymerases may employ to switch between transcription and genome replication. This knowledge will assist in the design and development of antivirals against paramyxoviruses.

ACKNOWLEDGEMENTS

Science does not occur in a vacuum, and many people helped in aspects of my work that I describe here. Takashi Okura, a post-doc in Robert Lamb's lab, started the L-P project, and Megha Aggarwal, another post-doc in the Lamb lab, helped with the final steps of the project. The Med-PIC project is a collaboration with the Tjian lab at the University of California, Berkeley. I want to thank Carla Inouye for all her work purifying proteins that I got to assemble into large complexes and for training me when I was fortunate enough to travel to California during the ill-fated month of November 2016. I also want to thank Anna Talyzina and Siyu Chen for their collaboration on the Med-PIC project that got a surprise boost very late in the process and led to the work described here.

To the entire lab, past and present, thank you for your friendship and advice over the past six years. It's those around us that help us to do the work we do, and you all made my years in graduate school a great experience. I especially thank Yan Han and Sue Fishbain for their many conversations over coffee, lunch, and tea. Your guidance helped me become the scientist and person I am today. I also want to thank my advisor, Yuan He, for taking a chance on a first-year graduate student with minimal wet lab experience and letting me work on a project that I found fascinating. Even though this project turned out to be more complicated than we imagined, you never lost patience and allowed me to keep pushing.

I want to thank the Lamb and Andersen labs for their presence as mentors during my first year of graduate school. You helped me to find my passion for science, and many of you became good friends. Thank you to my thesis committee, Alfonso Mondragón, Heather Pinkett, and Ishwar Radhakrishnan, for all your advice over the years.

Thank you to Jonathan Remis, Jason Pattie, and the rest of the SBF staff for your assistance and technical support. Thank you to Zhiheng Yu, Chuan Hong, and Rick Huang at the HHMI

Janelia CryoEM Facility for help in microscope operation and data collection of initial Med-PIC data sets. Thank you, Cathy, Deborah, and the staff of the MBTP and MolBio for your support.

To all my friends on the 4th floor of Cook Hall and scattered throughout Pancoe, Silverman, Ryan, and Tech, thanks for being an ear when I needed one. Grad school isn't easy, and it was great to experience it with others.

To my Pokémon Go community, thank you for helping to make Evanston and Chicago home for us. It's brought a real sense of satisfaction to watch this community grow, and I'm so proud of what we accomplished. I've truly enjoyed seeing all of your faces from the coldest days out on the lakefill to beers and PvP.

And lastly, thank you to my family and friends who have accompanied me on this journey. I'm sad that we won't be able to celebrate in person this year, but we'll make up for it ☺

To Shannon and Emma, it seems like so long ago that we started this grad school experience. Thank you for your thoughtfulness and friendship over all these years.

To Andy and Jenny, thank you for welcoming us into your Sojo family and becoming such good friends. We really value your friendship and can't wait to watch your family grow!

To mom and dad, you're the reason I fell in love with math and science. Thank you for all your support; I couldn't have done this without you. We can't wait until we can move closer and see you and the rest of the family more often!

To Alicia, thank you. You've been there for all the ups and downs, the excitement and disappointment, the venting, the hopefulness, and all the good times we've had here. Thank you for taking the chance and deciding to move to Chicago with me. Thank you for listening to all my science rants, I know you didn't understand, but sometimes we just need someone to listen. It hasn't been easy, but you've made this adventure so much fun. I love you so much.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1:

Introduction

1.1. Transcription – The Central Dogma of Biology

Information stored in nucleic acid polymers is utilized by all life forms to create new progeny faithfully. Deoxyribonucleic acid (DNA) is the genomic storage material for all known life except for some viruses due to its increased stability relative to ribonucleic acid (RNA). The vast majority of DNA exists in a double-stranded (ds) state, while RNA is usually single-stranded (ss) but can exist in a double-stranded state. DNA is composed of a deoxyribose molecule bound to one of four different bases, the purines adenine and guanine, and the pyrimidines thymine and cytosine. Adenine forms a base pair with thymine, and guanine does the same with cytosine, allowing one strand of DNA to provide all the information necessary to create the complementary strand. RNA is very similar to DNA, except for an extra hydroxyl group on ribose and uracil replacing thymine as one of the pyrimidine bases. DNA and RNA are complementary with each other; an adenine deoxyribonucleotide can base pair with a uracil ribonucleotide. The genomes of an organism contain many genes, continuous series of bases that encode for a protein. In all known living organisms, this process involves the transcription of the DNA sequence of a gene into a complementary strand of RNA called messenger (m)RNA, which is used by the ribosome, an enzyme common to all organisms, as the template to make a protein.

Viruses utilize either DNA or RNA, but rarely both, to store their genetic information. The nucleic acid that is used is the first step in classifying viruses. The Baltimore Classification of viruses contains seven groups, dsDNA viruses, ssDNA viruses, dsRNA viruses, positive sense (+)ssRNA viruses, negative-sense (-)ssRNA viruses, ssRNA-reverse transcriptase (RT) viruses, and dsDNA-RT viruses. The RT viral groups are the two exceptions to the above rule that viruses utilize either DNA or RNA because they use replicate by using the other nucleic acid as an intermediate step. The end goal of this process is always to produce mRNA that is used by the host organism's ribosomes to make the virus's proteins. The diversity in the mechanisms by which

viruses store their genetic material and make mRNA is fascinating, and knowledge of these mechanisms is critical as it represents a way to inhibit the synthesis of a specific group of viruses' genetic material using drugs without affecting the host organism.

1.2. Mononegaviridae

The order *Mononegaviridae* consists of enveloped viruses with non-segmented, negative-strand (NNS) RNA genomes and includes many families of viruses whose members pose significant burdens to human and livestock health. Example members of this order include measles virus (MeV) and mumps virus (MuV) from the *Paramyxoviridae* family, human metapneumovirus (hMPV) and respiratory syncytial virus (RSV) from the *Pneumoviridae* family, Ebola virus and Marburg virus from the *Filoviridae* family, and *Rabies lyssavirus* and vesicular stomatitis virus (VSV) from the *Rhabdoviridae* family. Given the high pathogenicity of members of these families, many viruses from this order are heavily studied to develop vaccines and anti-viral therapeutics. Within the *Paramyxoviridae* family, PIV5 has proven to be an excellent model system for the study of viral proteins [1-5].

1.2.1. Paramyxovirus transcription

Paramyxoviruses express between six and ten proteins encoded in their approximately 15 kilobase (kb) genomes. Four of the virus's genes are found in all mononegaviruses, the nucleo- (N) protein, phospho- (P) protein, matrix (M) protein, and large (L) protein [6]. N encapsidates the viral genome into a massive helical nucleoprotein complex, a common mechanism to protect the virus's genome from cellular RNases. Paramyxoviruses follow a standard "rule of six" where each monomer of N binds to 6 bases [7]. Replication and transcription of the genome are carried out by the polymerase complex consisting of L, P, and N [8]. Following infection, the genomic

ribonucleoprotein complex serves as a template for transcription by L-P complexes that were packaged into the viral particle. Transcription begins at the 3' end of the genome with the synthesis of a short leader sequence (Le) before the beginning of the first gene is encountered. The Le RNA is released before the start of the first gene, and the polymerase complex recognizes gene start (GS) sequence signals found at the beginning of each gene. These GS signals serve as internal promoters for the synthesis of mRNA from each gene. The polymerase complex adds a 5' guanosine cap, methylates the cap (m^7G), and methylates the transcript at the 2' oxygen of the first base's ribose group. Gene end (GE) signals located at the end of each gene are responsible for signaling the polymerase to poly-adenylate and release the transcript [9, 10]. After the polymerase complex has transcribed each gene, it scans through the intergenic region until it encounters a new GS signal. The polymerase complex can fall off the genome in these intergenic regions resulting in the first gene, encoding N, being transcribed to the greatest extent, followed by the second gene, encoding P, resulting in the fewest transcripts being produced for the last and largest gene, L. Transcription of the viral genome requires separation of the RNA from each monomer of N within the large helical genome assembly. Because no additional N monomers are present before transcription and translation of the virus's genes have occurred, a mechanism to retain the disengaged N monomers to re-coat the genome once it emerges from the polymerase complex is required [11].

1.2.2. Paramyxovirus genome replication

Replication of paramyxovirus genomes is carried out by the same L-P-N-RNA complex as transcription, although instead of producing separate capped, polyadenylated transcripts, L transcribes the entire genome creating an intact, N-coated anti-genome. The anti-genome serves as the template for a second round of genome replication to produce a new genome. Replication

initiates at the same position as transcription in paramyxoviruses. However, in this situation, the Le RNA is not released, GS and GE signals are ignored, and the polymerase complex transcribes through the entire genome without stopping [12]. This final N-coated genome is then packaged into new viral particles and can go on to infect new cells. Paramyxoviruses are unusual within NNS viruses because they have a bipartite promoter formed by the first 12 bases of the 3' end of the genome and additional cytosine residues located downstream [9]. These cytosine residues are positioned on the same face of the N-RNA helix, placing them close to the 3' end of the genome. Although the exact position of these cytosine residues differs by species, they are all solvent facing, suggesting a role in recognition by L.

1.2.3. Genome replication versus transcription

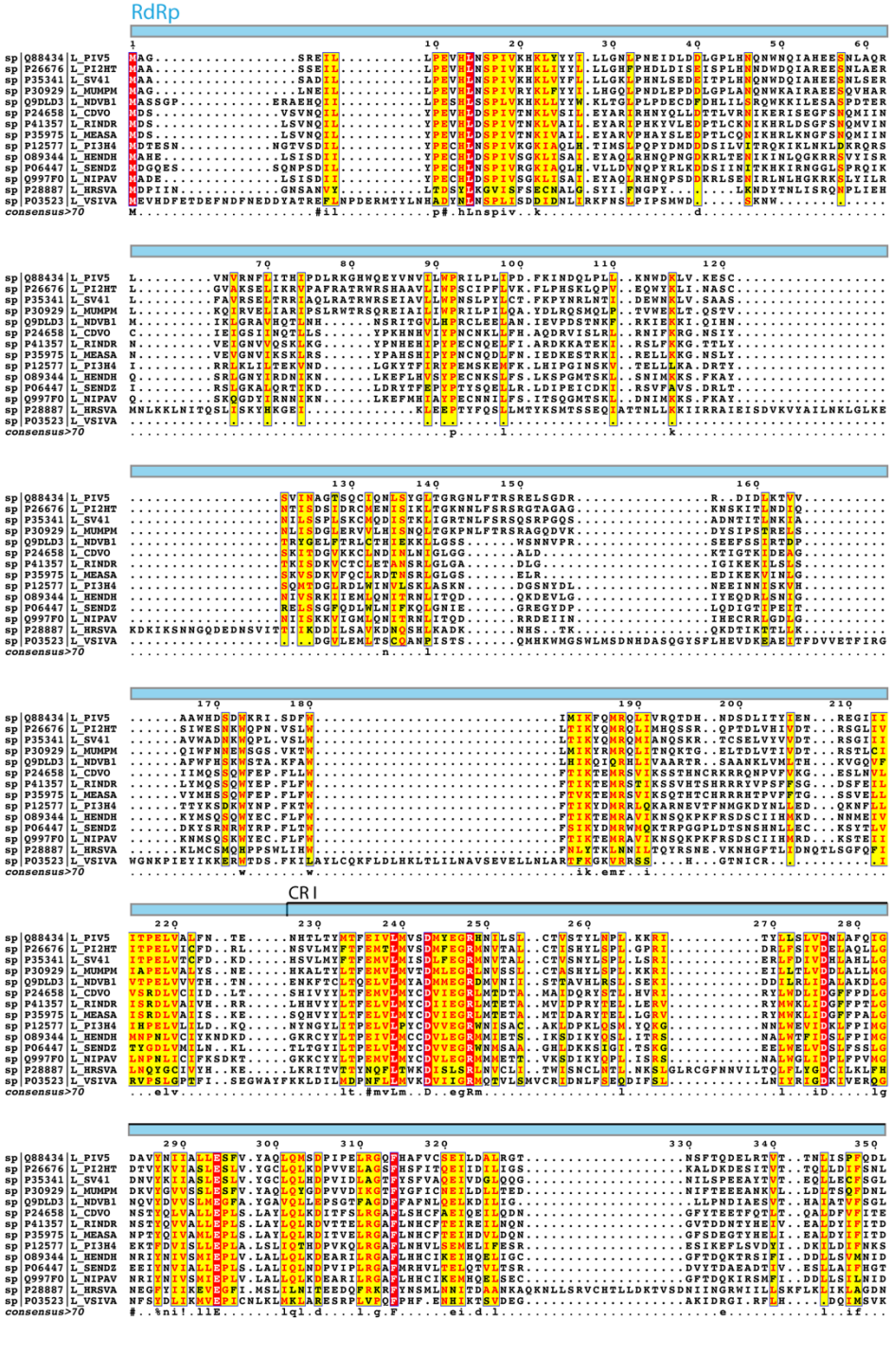
The difference in the release of the Le RNA and recognition of the first GS signal or lack thereof between transcription and genome replication and the necessary processing of the RNA required during transcription is poorly understood. Evidence from the pneumovirus RSV suggests that the cellular concentrations of ATP, GTP, and CTP are responsible for this decision due to different starting bases for the initiation of these two processes [12, 13]. However, no evidence for two initiation sites exists for paramyxoviruses. Instead, the amount of N protein is implicated in the transition from transcription to genome replication for paramyxoviruses and many other NNS virus families, suggesting a common regulatory mechanism [14-16]. This follows logically with the requirement of encapsidating the newly synthesized (anti-)genome with N, a requirement not necessary when transcribing the viral genome. This pool of free N is not present during the early stages of infection and is only formed after transcription and translation of the viral genome. One question that has yet to be answered is whether a single polymerase can transcribe and replicate the genome. If possible, only a single pool of polymerase complexes would be

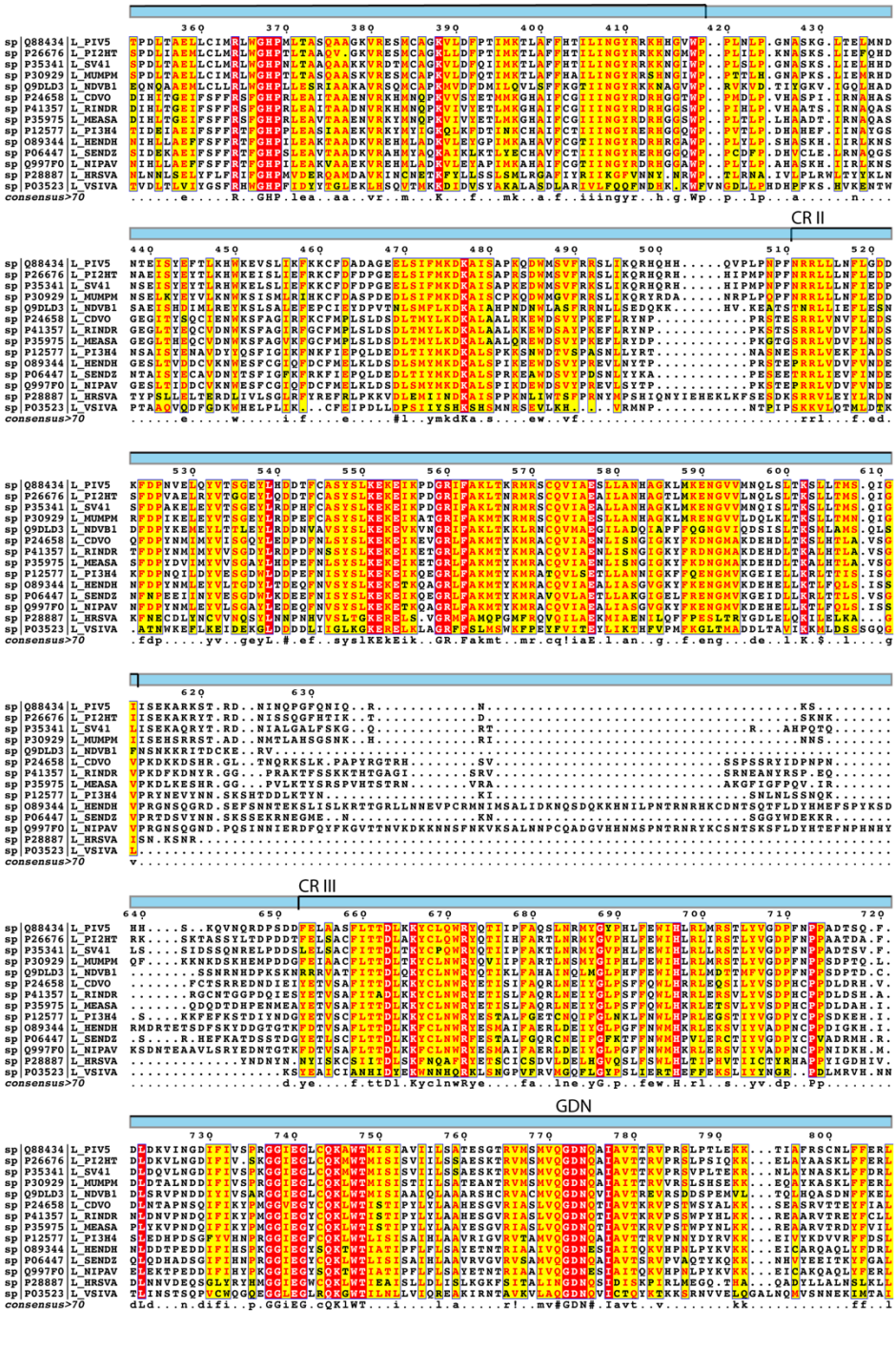
necessary. However, if not possible, at least two different pools of polymerase complexes within an infected cell would be necessary, an initial transcription-competent pool and a later, genome replication-competent pool [17, 18]. Due to the inclusion of approximately 50 L-P complexes in viral particles, in the case of Sendai virus, the possibility of an initial pool of transcriptionally-competent polymerase complexes from the infecting viral particle and a second pool of newly translated, replication-competent polymerase complex exists [19].

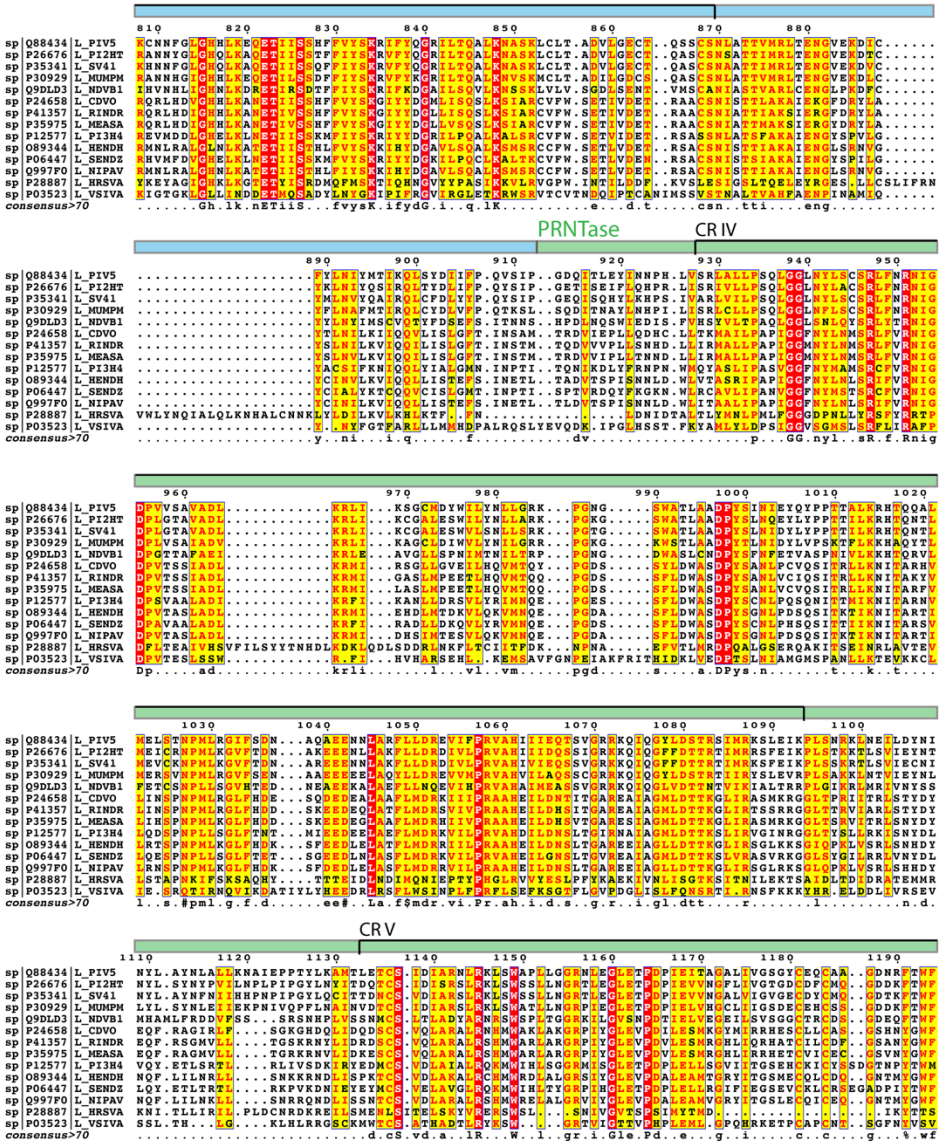
1.2.4. The L protein of Mononegaviruses

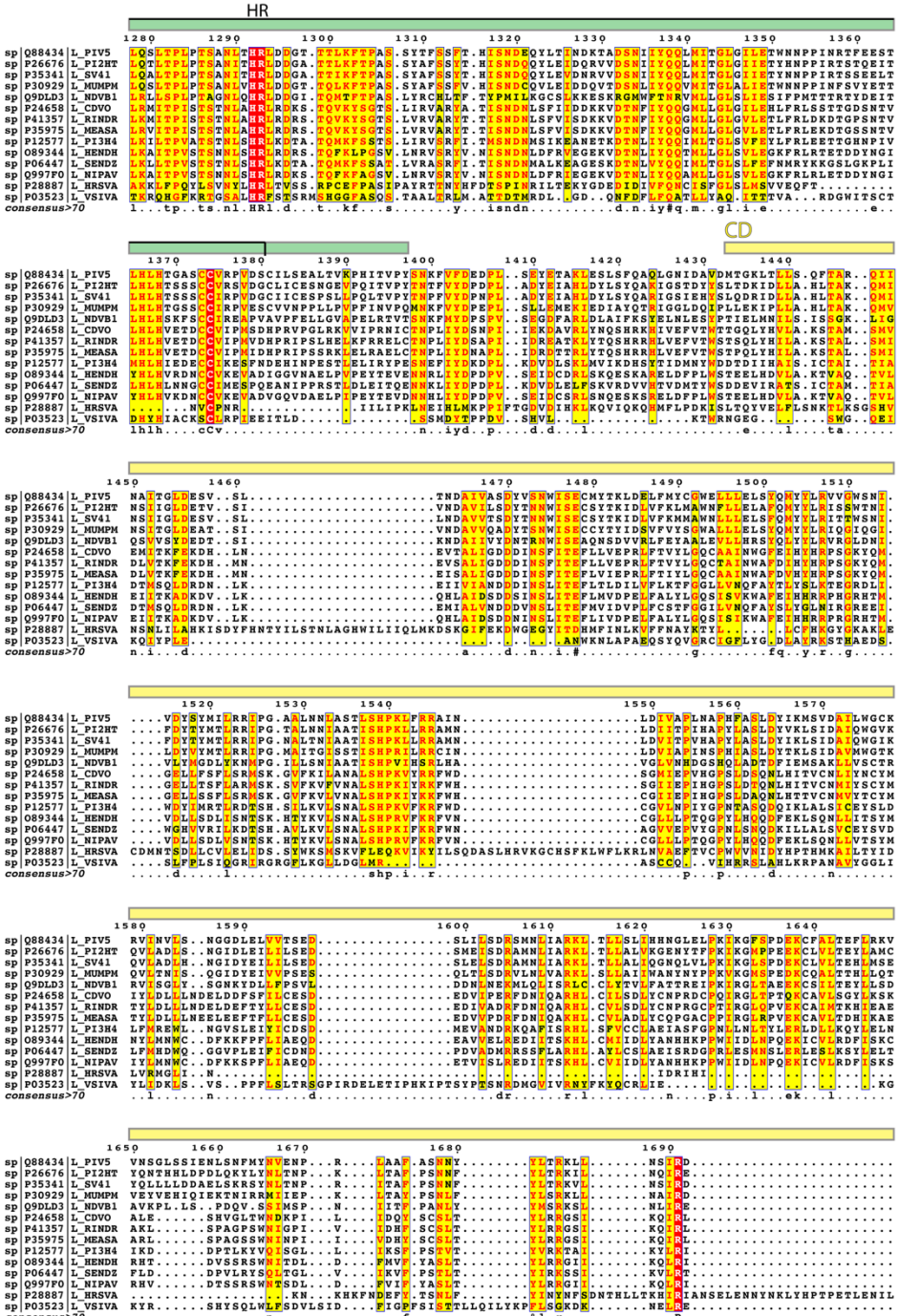
The main component of the polymerase complex, L, is a multifunctional enzyme that is more than 2,000 amino acids in length and over 240 kDa in size, except for Bornaviruses whose L is slightly smaller. Early studies on NNS viruses identified six conserved regions in L (CR I-VI) that form crucial motifs involved in its function and interactions with its cofactor P (Figure 1.1) [20, 21]. The functions of these particular regions have been revealed from structural and biochemical studies [22-24]. Structures of L from other NNS viruses have identified five conserved domains: RNA-dependent RNA polymerase (RdRp), poly-ribonucleotidyltransferase (PRNTase), connecting domain (CD), methyltransferase (MTase), and the C-terminal domain (CTD) (Figure 1.2) [11, 24-29]. The conserved regions are all located within the enzymatic domains of L.

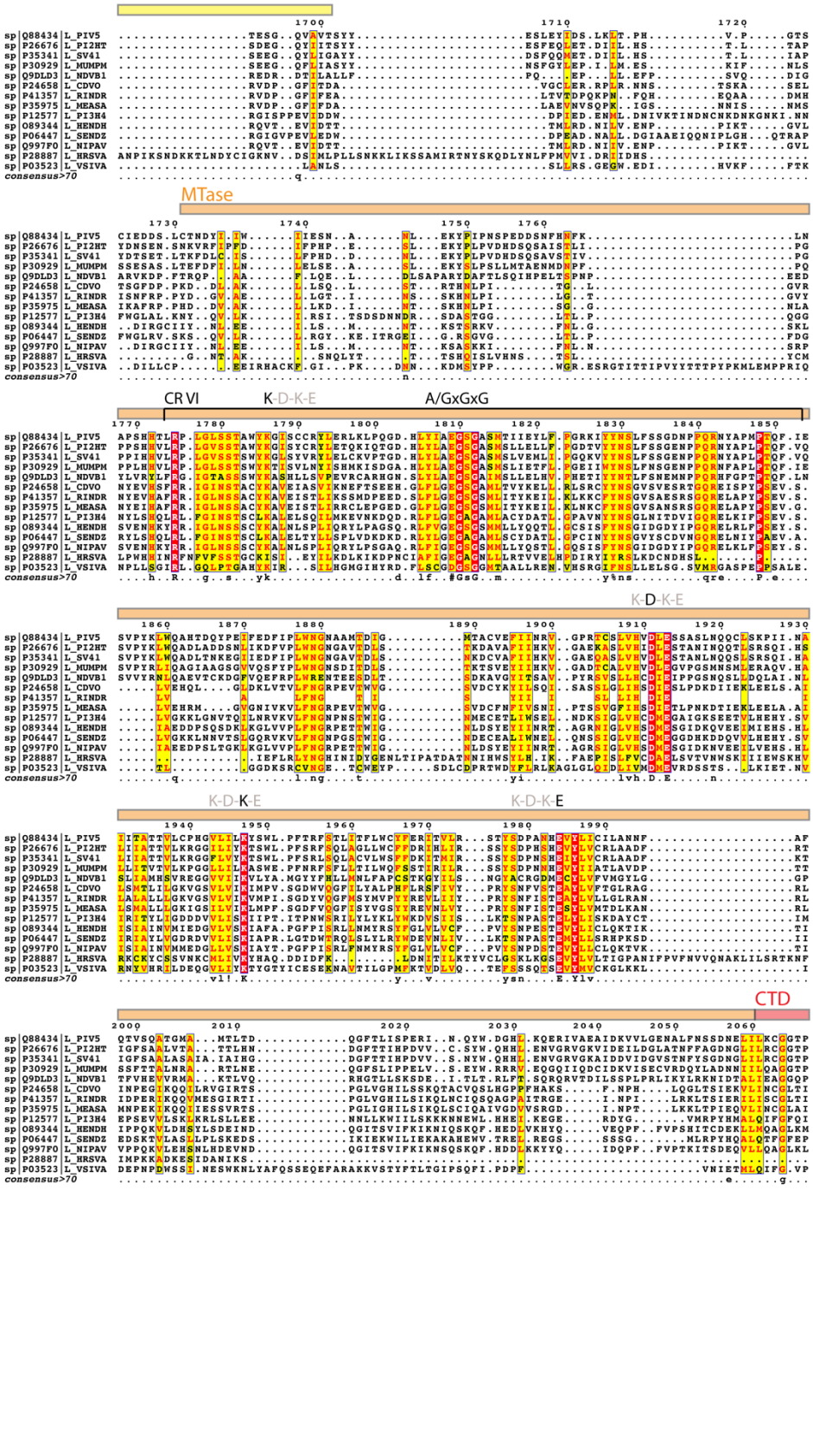
Figure 1.1. **Sequence alignment of NNS L proteins.** Sequences for PIV5, PIV2, Simian virus 41, Mumps virus, Newcastle disease virus, Canine distemper virus, Rinderpest virus, Measles virus, PIV3, Hendra virus, Sendai virus, Nipah virus, human respiratory syncytial virus, and vesicular stomatitis virus. The colored bar above sequence alignment shows domain boundaries. Conserved regions (CR) and motifs are labeled.


RdRp
CRI


CR II
CR III
GDN


PRNTase
CR IV
CR V
GxxT


HR
CD


MTase
CR VI
K-D-K-E
A/GxGxG
K-D-K-E
K-D-K-E
K-D-K-E
CTD

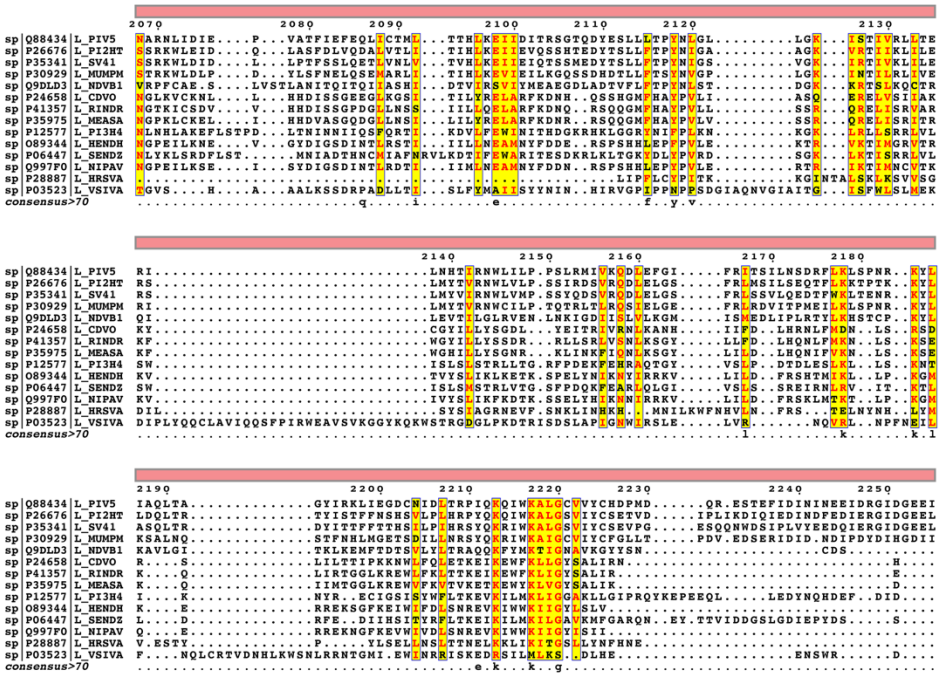
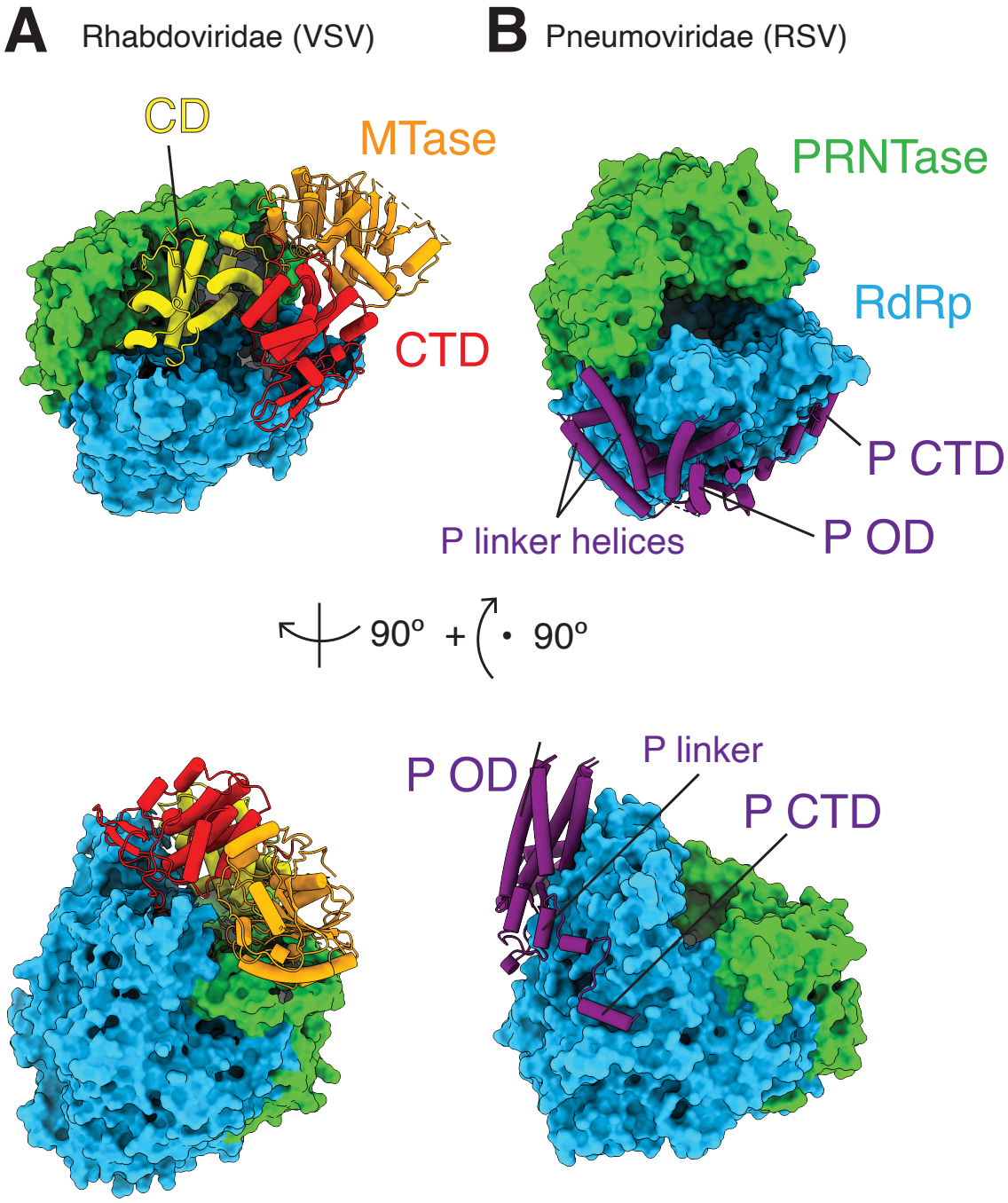

2070 2080 2090 2100 2110 2120 2130
2140 2150 2160 2170 2180
2190 2200 2210 2220 2230 2240 2250

Figure 1.2. **Existing structures of NNS virus polymerase complexes.** A) Structure of a typical rhabdovirus L protein from VSV with the CD-MTase-CTD module bound on top of the RdRp-PRNTase module. A small fragment of P (35-106, not shown) stabilizes the CD-MTase-CTD module. B) Structure of a typical pneumovirus L-P complex from RSV. The CD-MTase-CTD module does not stably associate with the RdRp-PRNTase module. P forms a tetramer due to its oligomerization domain with unique folds for each linker-CTD region on different surfaces of L.


A
Rhabdoviridae (VSV)
B
Pneumoviridae (RSV)
CD
MTase
CTD
PRNTase
RdRp
P CTD
P linker helices
P OD
90°
+
90°
P OD
P linker
P CTD

As predicted based on sequence similarity, the RdRp domains of mononegavirus L proteins share a conserved architecture with most RNA and DNA polymerases, appearing like a right-handed “fingers-thumb-palm” structure [11]. The RdRp and PRNTase domains combine to form a stable ring-like structure that is seen in other NNS viruses [11, 25, 27-32]. CRs I-III are located in the RdRp domain, the first of which is responsible for binding the cofactor P. The fingers subdomain contains CR II, which is involved in RNA binding [21, 33]. Flexible loops connect the sub-domains, allowing subtle mobility among them. A conserved Gly840 at the intersection of the palm and thumb sub-domains provides additional flexibility. CR III forms the palm subdomain, which folds into a small anti-parallel β-sheet. The RdRp active site is formed by the conserved GDN (772-774) motif located between two of the β-strands. This active site motif is surrounded by several loop regions that form the tunnel that the template RNA travels through towards the catalytic site. The palm sub-domain also contains a conserved aromatic residue, Tyr667, which is positioned away from the active site, contributing to L stability. The fingers and thumb sub-domains are primarily helical except for a few small β-sheets. These two sub-domains form most of the interface with the PRNTase domain.

The PRNTase domain utilizes an entirely different mechanism to cap the nascent RNA strand than eukaryotes. In all NNS viruses, the first transcribed nucleotide forms a covalent bond to a histidine residue located in a conserved histidine-arginine (HR) motif found in CR V [34]. Cap addition occurs via nucleophilic attack by guanosine diphosphate on the covalent pRNA-His bond. In eukaryotes, a nucleophilic 5’ ppRNA attacks a GTP [35]. The PRNTase domain is located above the RdRp, creating an interior central cavity where the active site for RNA transcription is located. A second conserved motif in CR V, GxxT, binds the capping guanosine nucleotide [11]. Reoviruses have a priming loop that projects down into the central cavity [36]. However, the homologous loop in mononegaviruses is much shorter and likely does not serve that same

function. Instead, a different loop located in CR V of the PRNTase domain projects down into the central cavity, occupying the same position as a priming loop in the reovirus polymerase [11]. C-terminal to this priming loop is the conserved HR motif responsible for covalently binding the 5' end of the RNA. The location of the priming loop and PRNTase domain would prevent continued elongation of RNA synthesis beyond just a few nucleotides, likely required a domain rearrangement to accommodate the growing product RNA strand and create its exit tunnel from the complex. Experiments on VSV indicate that capping and methylation of transcripts occur after exactly 31 nucleotides have been transcribed [37]. Based on our current structural knowledge of the complex, this means that the active sites for capping and methylation are either the same or very close together, which is not the case in existing structures [11, 25, 27-29, 32].

A long flexible loop connects the RdRp domain to the CD, a poorly conserved domain with no known function. Though the CD has not been observed to perform any enzymatic activity, it is required for proper polymerase function and cannot tolerate insertions in its sequence [38]. A second loop connects the CD and MTase domain, but the role of these flexible loops on either side of the CD is not yet understood.

The MTase domain performs the essential catalytic reactions of methylating the 2'O of the first product nucleotide, followed by methylating N7 of the capping guanosine [24]. Its active site is highly conserved, containing a catalytic K-D-K-E motif common to 2'O methyltransferases [39]. In PIV5, those residues are K1786-D1911-K1947-E1984. K1786 and an AxGxG motif, conserved in paramyxoviruses but slightly atypical from the GxGxG motif of most NNS viruses, are located within CR VI. The AxGxG motif forms part of the SAM binding site, the methyl group donor [24].

Like the CD, the CTD contains very little sequence conservation and is structurally divergent among NNS viruses. A structure of the MTase-CTD module from hMPV reveals that a conserved KxxxKxxG motif in paramyxoviruses, pneumoviruses, and filoviruses, is located in α6 of the CTD.

This motif is a canonical eukaryotic guanylyltransferase (GTase) motif located in a β-sheet [35], suggesting that it does not play a role in cap formation in these viruses [24]. Instead, it is positioned so that both lysine residues point into the active site and contribute to positioning the RNA molecule for proper methylation of the cap.

Structures of L from the *Rhabdoviridae* family place the CD-MTase-CTD module on top of the RdRp-PRNTase ring (Figure 1.2) [11, 27, 28]. This conformation of the CD-MTase-CTD module is stabilized by a fragment of P [40]. In the absence of this fragment of P, the CD-MTase-CTD module does not stability associate with the RdRp-PRNTase module and, instead, is flexibly tethered like beads-on-a-string. Structures of full-length L from the *Pneumoviridae* family solved in complex with full-length P had no density for the CD-MTase-CTD module, showing that this beads-on-a-string structure is likely shared among NNS viruses [25, 29, 32]. The fragment of P from the rhabdoviruses that interacts with the CD and MTase is not conserved among NNS viruses, leaving questions about how conserved the role of this peptide in stabilizing L is. The crystal structure of a substrate-bound MTase-CTD dimer from hMPV shows that these domains stably interact and likely do not separate upon dissociation of the CD-MTase-CTD module from the RdRp-PRNTase ring [24].

1.2.5. The phosphoprotein is a coactivator of L

P performs several vital functions in the viral life cycle and is composed of three domains: N-terminal (P-ND), central oligomerization (P-OD), and C-terminal domain (P-CTD) (Figure 1.2, 1.3). These domains are mostly conserved in NNS viruses, but their folds can differ significantly [41, 42]. The ND of P is generally unstructured, not highly conserved, and binds a nascent N monomer (N^0), forming the N^0P complex, which prevents premature N oligomerization and non-specific RNA binding [2, 43].

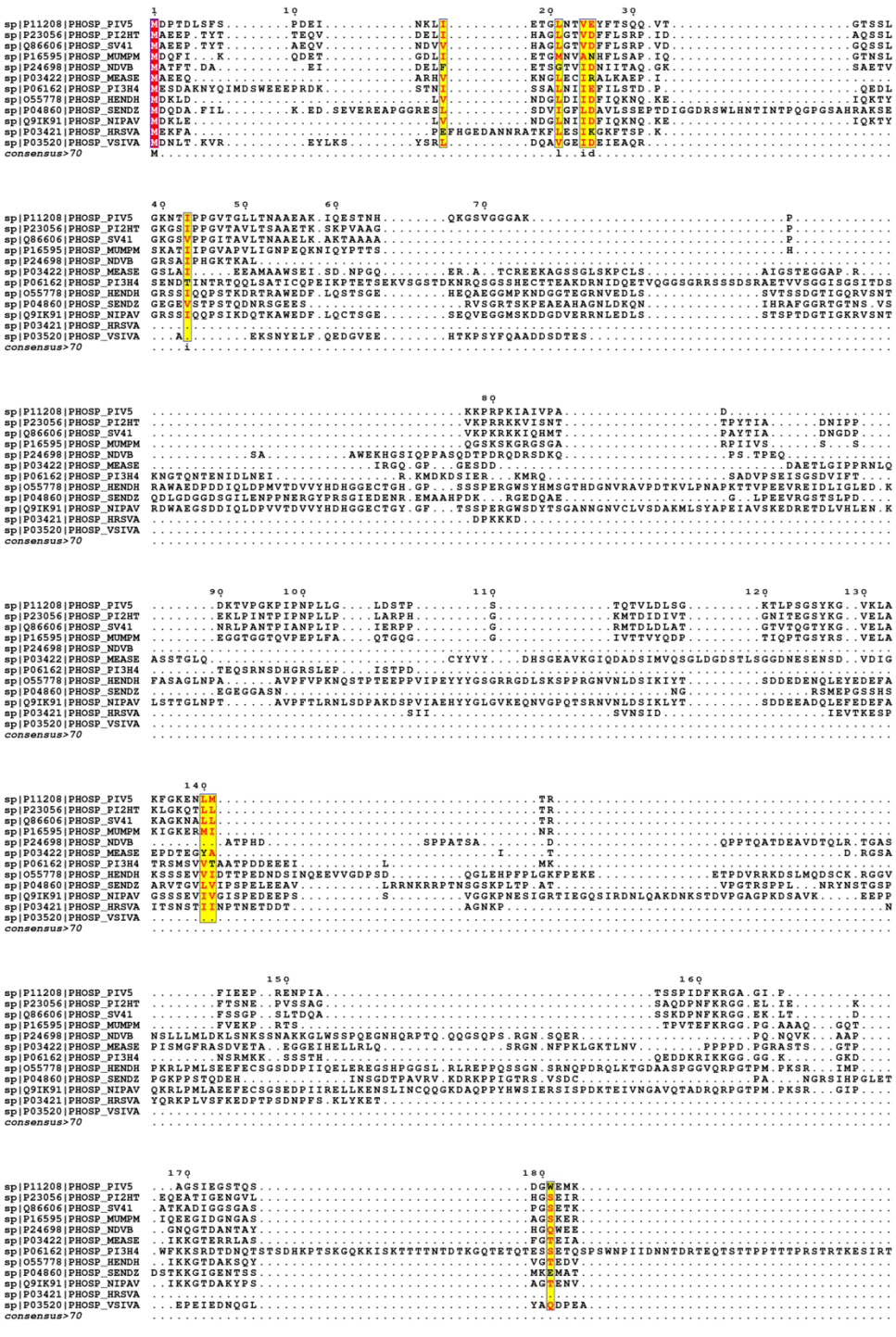
In paramyxoviruses, the CTD is referred to as the C-terminal X domain (P-XD). P-XD is also intrinsically disordered in solution but forms a three-helix bundle under crystallizing conditions and associates with the molecular recognition element (MoRE) of the C-terminus of N (N-tail) [42, 44, 45]. In VSV, P-CTD forms a four-helix bundle that binds directly to the globule domains of two adjacent N monomers assembled into a nucleocapsid [41]. This suggests the same role for P-CTD in binding to N but completely different mechanisms for doing so.

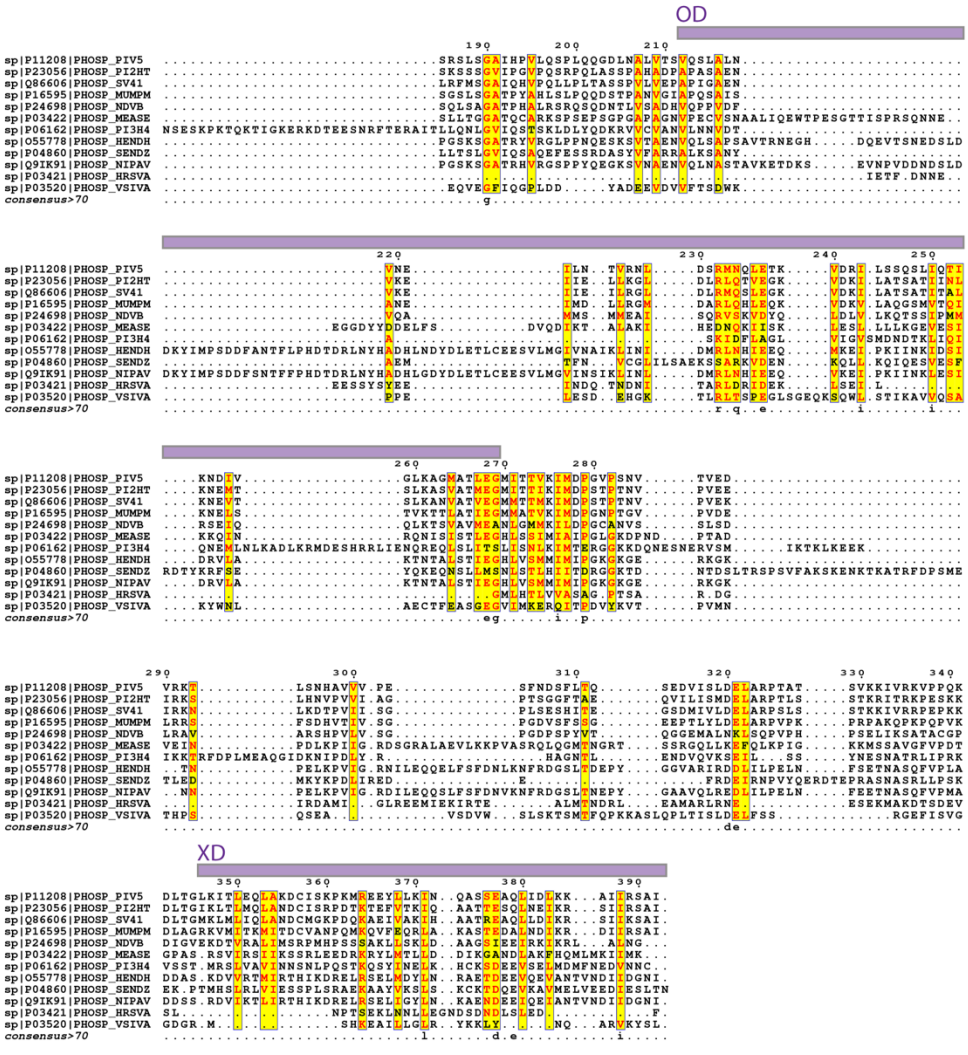
P-OD self-oligomerizes to form either dimers or tetramers and directly interacts with the L protein, tethering the polymerase to the N-coated RNA template [43]. To date, crystal structures of P-OD from Sendai virus (SeV), MeV, NiV, MuV, and VSV are available [46-50]. The paramyxoviruses (SeV, MeV, and NiV) share a tetrameric coiled-coil structure with all chains parallel to each other. The MuV P-OD, the one exception to this rule, contains two parallel chains of helices arranged in an antiparallel orientation to the other parallel pair. The VSV P-OD forms a significantly shorter dimer with flanking β-sheets to stabilize the structure.

The interaction between L and P is essential for paramyxovirus replication and transcription. Biochemical characterization of the interaction between PIV5 L and P reveals the importance of the N-terminal half of the L protein [51]. In contrast, the C-terminal half is dispensable for the interaction with P. Similar results have been observed in MeV, SeV, and HPIV3 in which the N-

terminal region of L is responsible for L-P interaction [52-55]. However, in *Rabies lyssavirus* (RaV), the C-terminus of L and the first 19 residues at the N-terminus of P are essential for the interaction between L and P [56]. The C-terminus of SeV P has been identified as an L binding region by mutational analysis [57]. A bipartite interaction between L and P has been observed in MeV, where the P-OD and P-XD bind at two separate surfaces of L [58]. The recently published cryo-EM structures of L-P from RSV and hMPV demonstrate a tentacular arrangement of P with significant interfaces between L and P [25, 26, 29]. Each of the four chains of P has a distinct fold in the region C-terminal to the P-OD tetramer and interact with unique locations on L, suggesting structural plasticity in this interaction. Whether this is a common feature of L-P complexes among NNS RNA viruses or limited to viruses within the *Pneumoviridae* family remains to be determined. The fragment of P that stabilizes the CD-MTase-CTD module in rhabdoviruses appears unique to that family given the lack of sequence conservation among NNS viruses and the differences in the lengths of P-OD, which results in that segment being much further from L in other virus families.

Figure 1.3. **Sequence alignment of NNS P proteins.** Sequences for PIV5, PIV2, Simian virus 41, Mumps virus, Newcastle disease virus, Canine distemper virus, Rinderpest virus, Measles virus, PIV3, Hendra virus, Sendai virus, Nipah virus, human respiratory syncytial virus, and vesicular stomatitis virus. The colored bar above sequence alignment shows domain boundaries.

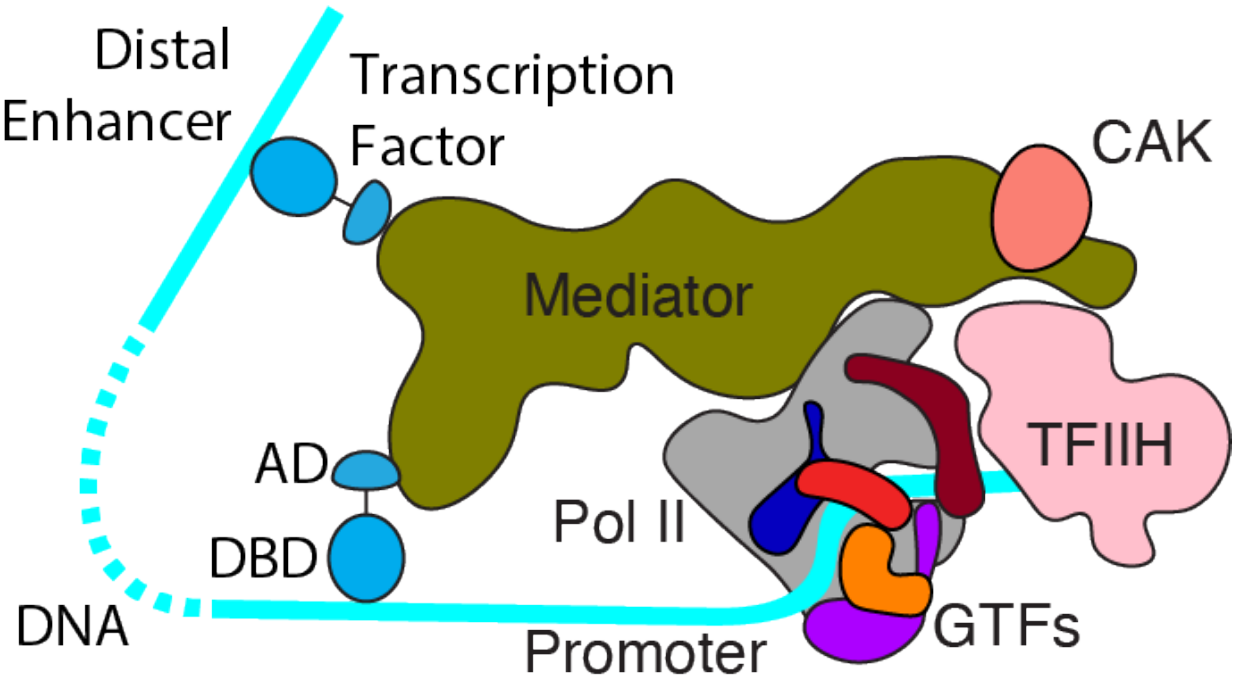



OD
XD
sp|P11208|PHOSP_PIV5
sp|P23056|PHOSP_PI2HT
sp|Q86606|PHOSP_SV41
sp|P16595|PHOSP_MUMPM
sp|P24698|PHOSP_NDVB
sp|P03422|PHOSP_MEASE
sp|P06162|PHOSP_PI3H4
sp|O55778|PHOSP_HENDH
sp|P04860|PHOSP_SENDZ
sp|Q9IK91|PHOSP_NIPAV
sp|P03421|PHOSP_HRSVA
sp|P03520|PHOSP_VSIVA
consensus>70

1.3. Eukaryotic mRNA Transcription

Transcription of all mRNA in eukaryotes is carried out by RNA polymerase II (Pol II) [59, 60]. Pol II cannot by itself locate the transcription start site (TSS), open a transcription bubble to expose the template strand, and transition to an elongation state. Instead, DNA-bound transcription factors position the co-activator complex Mediator to facilitate the assembly of the pre-initiation complex (PIC), composed of Pol II and the general transcription factors (GTF) TFIIA, TFIIB, TFIID (TBP), TFIIE, TFIIF, and TFIIH (Figure 1.4) [61-63]. These GTFs help position Pol II to initiate transcription at the correct genomic locus, and TFIIH feeds DNA into the active site of Pol II, generating force against the TBP/TFIIB/TFIIA lobe to unwind the DNA and expose the template strand [64]. The entire Mediator-bound PIC (Med-PIC) is 2.7 megadaltons (MDa) in size, contains 56 polypeptides, and represents a unique challenge for structural characterization due to difficulties in obtaining and assembling these complexes, as well as to their inherent flexibility.

Figure 1.4. **Eukaryotic transcription initiation.** Transcription factors bind to specific DNA sequences using their DNA-binding domains (DBD) located at distal enhancers or in direct proximity to gene promoters and recruit the co-activator complex Mediator through interactions between the activation domain (AD) of the transcription factor and activator-binding sites within Mediator. The resulting pre-initiation complex, composed of RNA Polymerase II, the general transcription factors TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH, and Mediator is capable of initiating transcription of the target gene.



1.3.1. Structures of PICs provide atomic details on promoter opening

Unraveling the DNA duplex to expose the template strand of DNA for RNA transcription requires the coordinated effort of Pol II and the GTFs. Structures of PICs from yeast and humans have elucidated this process by comparing their structure on different DNA templates [65-67]. Transitions from the closed complex (CC) to the open complex (OC) confirm that the XPB subunit of TFIIH acts as a DNA translocase, hydrolyzing ATP to force DNA into the active site of Pol II. Because of the stable engagement of the TBP-TFIIB-TFIIA-TFIIF submodule with the TATA box and surrounding DNA, including TFIIB recognition elements, this causes negative supercoiling of the DNA within the active site, which contributes to the melting of a DNA bubble at the TSS. XPB must translocate 12 bp to transition between the CC and OC states. Even though an 11 bp bubble was engineered into the OC template through forced mismatches in the DNA sequence, the formation of the OC bubble resulted in the opening of an additional two bp of DNA on the upstream end of the DNA bubble [66]. Comparison of OC structures between yeast and human show that the TBP-TFIIB-TFIIA-TFIIF submodule can shift relative to Pol II, suggesting some plasticity in this interface [66, 67]. However, both structures suggest that TFIIE plays a role in stabilizing the transition from CC to OC by engaging with the coiled-coiled clamp of Pol II, DNA, or possibly both.

1.3.2. Phosphorylation of the RPB1 CTD is a crucial step in transcription initiation

The largest subunit of Pol II, RPB1, contains a long, repetitive CTD connected by a flexible linker region to the rest of the subunit [68]. Neither the CTD nor the 80-residue linker is visible in structural studies due to their mobility [69-71]. The CTD consists of 26 repeats in yeast and 52 repeats in humans of the consensus YSPTSPS heptamer sequence [72, 73]. Every residue within this sequence can exist in multiple states, either by post-translational modification (Y, S, and T) or isomerization (P), leading to a CTD code that allows it to interact with different partners at

specific times [74]. The use of antibodies specific to different combinations of modifications has elucidated the general principles of their roles in transcription [72]. However, the distribution of modifications within the CTD is just beginning to be understood [75, 76]. Unlike the multifunctional viral polymerase described earlier, Pol II is not responsible for capping or methylating the newly transcribed mRNA. Instead, CTD modification is directly linked to the proper processing of RNA through the recruitment of mRNA processing proteins; the most critical residues for this recruitment are serines (S) at positions 2, 5, and 7 [77-79]. Phosphorylation of S^5 (pS^5) peaks at the 5' end of genes and decreases towards the 3' end, concurrent with a rise in pS^2 [73, 80]. Recruitment of the 5' capping enzymes, mRNA-capping enzyme (RNGTT) and mRNA (guanine-N7-)-methyltransferase (RNMT), is dependent on pS^5, and formation of the 5' cap is indispensable for growth [81, 82]. Cyclin-dependent kinase 7 (CDK7, or Kin28 in yeast) is responsible for phosphorylating S^5 and is part of the cyclin-activated kinase (CAK) module of TFIIH, together with Cyclin-H and Mat1 [83, 84]. Mounting evidence exists that Mediator recruits TFIIH to the PIC and stimulates the CAK module's ability to phosphorylate S^5 [84, 85].

1.3.3. Current structural knowledge of Mediator

Early structures of Mediator were obtained using electron microscopy and were limited to low-resolution reconstructions [86-94]. These early structures delineated Mediator into discrete modules, the head (MedHead), middle (MedMiddle), and tail (MedTail) modules. The limited resolution of these structures led to errors in the assignment of subunits to these different modules, which was not rectified until more detailed subunit deletion experiments were performed [95, 96]. In the end, the density that was labeled MedHead in these early studies is MedTail, the density that was labeled MedMiddle also includes MedHead, and the density that was labeled MedTail is the hook domain of MedMiddle. These higher resolution structures allowed accurate

docking of existing x-ray crystallography structures of MedHead and portions of MedMiddle into the density, showing for the first time that MedHead interacts with Pol II [97-103]. This had been previously shown genetically and biochemically as many MedHead subunits were initially identified through a genetic screen for mutations that suppress truncations of the Pol II CTD [104, 105]. A short peptide of the RPB1 CTD was co-crystallized bound to MedHead and, based on the location of the CTD modeled into the full yeast complex, also likely serves to stabilize the interface between MedHead and MedMiddle [97, 106, 107]. In apo Mediator, this interface is disrupted, allowing separation of MedHead and MedMiddle from each other [108]. MedHead and MedMiddle form core Mediator, the minimal complex necessary to activate transcription *in vitro* [107, 109]. These modules have been further separated into structural domains. MedHead is composed of the neck, shoulder, fixed jaw, and flexible jaw domains, while MedMiddle is composed of the hook, connector, knob, plank, and beam domains.

MedTail, primarily ignored in structural studies to date, serves as a hub for binding transcription factors [110, 111]. Structures of the human Med25 activator-binding domain and Med23 have been solved, but a complete structure for MedTail has yet to be solved [112-114]. Transcription factor binding sites within the genome can be promoter-proximal or thousands of bases away from the TSS, leading to a wide range of spatial orientations between enhancers and promoters [115]. Early structures of Mediator suggested that conformational changes upon TFIIF or activator binding could stabilize the complex and the activation of transcription, respectively [94, 116]. However, higher resolution structures are necessary to verify if this is the case. Understanding the conformational landscape of MedTail will also shed light on the possible spatial orientations of activators to their target promoters, defining limits on the 3D organization of the genome.

The fourth module of Mediator, the dissociable kinase module (MedKinase), was identified early on as a negative regulator of Mediator's role in transcription activation [117-119]. Later studies would show that MedKinase sterically competes with Pol II for access to binding cMed, spanning across the hook domain of MedMiddle and MedHead [96, 120]. MedKinase is preferentially found at Med-bound enhancers and leaves before engagement with Pol II [121, 122].

Atomic models of Med-PICs are currently limited to yeast [67, 106-108]. These studies provided the first atomic-resolution structural details for the interaction between MedHead and MedMiddle. They highlighted differences in how MedHead interacts with Pol II, suggesting that the interface between Mediator and Pol II is not rigid [67, 108]. Functional and structural studies have identified a minimal core Mediator (cMed), devoid of both MedKinase and MedTail [109, 123]. Structures of Mediator have poorly defined density for MedTail, leaving open the question of where most transcription factors bind [106, 108, 124].

A second highly flexible part of Med-PIC is the cyclin-activated kinase (CAK) module of TFIIH. The location of the CAK module in both free TFIIH and the PIC is highly mobile [66, 125-127]. The first Med-PIC structures located its position for the first time, at the end of the hook of Mediator, directly above TFIIH [67, 106]. However, the orientation of the individual subunits within this density has not yet been determined. The crystal structure of the human CAK module was very recently solved, placing the C-terminal globular portion of Mat1 as a wedge between the Cyclin H-CDK7 dimer [127]. The N-terminus of Mat1 forms a long helix that spans across the TFIIH horseshoe and contains a RING domain that interacts with the RPB7 OB domain and TFIIE E-linker helices [67]. Orienting the CAK module subunits within Med-PIC will provide essential information on the mechanism of the addition of the indispensable pS^5 post-translational modification of the CTD.

CHAPTER 2:

Structure of a paramyxovirus polymerase complex

2.1. Introduction

At the advent of this project, only a single structure of an NNS virus's L protein had been determined and contained only a small fragment of its cofactor P [11]. P is highly divergent, and existing structures of the oligomerization domain show that paramyxovirus P forms a four-helix bundle that is more than twice as long as the dimers formed by rhabdovirus P [47-50]. Here, I used cryo-electron microscopy (EM) to solve the structure of the PIV5 L-P complex at 4.3 Å resolution. All five domains (RdRp, PRNTase, CD, MTase, and CTD) of L are well resolved in the density map. The structure reveals two discrete binding interfaces on the L protein surface for the binding of P-OD and P-XD, respectively. A 1.4 Å resolution crystal structure of the PIV5 P-OD confirms the presence of a four-helix bundle arranged in a parallel orientation, consistent with other paramyxoviruses except for MuV. The tetrameric P-OD protrudes away from L, forming minimal interactions with L, while a single copy of a P-XD is bound near the nucleotide entry site of L. The priming loop of L adopts a transcription elongation conformation, and an intrusion loop occupies the active site of the RdRp domain. A significant conformational rearrangement of the domains responsible for nascent mRNA 5' methylation highlights the dynamic nature of the L protein, revealing a crucial mechanism for the spatial-temporal regulation of RNA synthesis. Comparisons with the structure of the VSV and RSV L-P complexes reveal fundamental differences in both L and P across different NNS viruses, critical for our understanding of the full catalytic cycle of these essential enzymes.

2.2. Methods

2.2.1. Cells and viruses

Baby hamster kidney (BHK) -21 and BSR T7/5 (BHK cells constitutively expressing T7 RNA polymerase) cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS),

tryptose phosphate broth (TPB), and 1% penicillin-streptomycin. 500 µg/ml G418 was also added to every third passage of BSR T7/5 cells. *Spodoptera frugiperda* (Sf9) cells, maintained in SF900 II SFM (Gibco) medium containing 10% FBS and 1% penicillin-streptomycin, were used for generating recombinant baculovirus (rBV) stocks and protein expression.

2.2.2. Plasmid construction

A FLAG tag was introduced into the C-terminus of L, and a histidine tag into the N-terminus of P. These L and P genes were separately cloned into the pCAGGS vector for expression in mammalian cells. To generate rBVs co-expressing L-P, the codon-optimized L gene under the control of the polyhedrin promoter and the P gene under the control of the p10 promoter were simultaneously cloned into the pFastBac Dual vector (Invitrogen). Eight different constructs (165-278, 172-278, 183-278, 203-278, 207-278, 214-278, 1-278, 178-392) encoding the PIV5 (strain W3A) P were amplified and sub-cloned in pET28a expression vector. All the constructs were in frame to produce the N-terminal 6-His tag followed by a thrombin cleavage site.

2.2.3. Expression and Purification of PIV5 L-P and P-OD

The rBV expressing L-P was generated in Sf9 cells by following the transfection of bacmid DNAs with Cellfectin (Invitrogen). Following the determination of virus titers by plaque assay, Sf9 cells were infected with the rBVs at a multiplicity of infection (MOI) of 1. At 96-120 h post-infection, the cell lysates were harvested by centrifugation (5,000 rpm, 50 min, 4°C), and the cell pellets were suspended with buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1 mM DTT, 1 mM EDTA, 1 mM PMSF and EDTA-free complete protease inhibitor (Roche)). The lysates were briefly sonicated and centrifuged (25,000 rpm, 60 min, 4°C) to obtain the supernatant containing soluble L-P complex. L-P was first purified by affinity chromatography

using anti-FLAG M2 affinity gel (SIGMA) and eluted with elution buffer (25 mM HEPES pH 7.4, 500 mM NaCl, 10% glycerol, 0.2% Tween 20, 1 mM DTT, and 100 μg/ml FLAG peptide (SIGMA)). For further purification, the L-P complex was passed through a Superdex 200 10/300 column (GE Healthcare) in buffer A (25 mM HEPES pH 7.4, 500 mM NaCl, 5% glycerol, 0.2% Tween 20, 1 mM DTT, and 6 mM $MgSO_4$). The peak fractions were collected for analysis by EM.

2.2.4. Electron microscopy

Negative stain samples were prepared using 400 mesh copper grids with a thin layer of continuous carbon that was glow-discharged in air for 10 seconds with 25 W of power. Purified L-P was diluted to 33 nM in buffer B (25 mM HEPES pH 7.4, 500 mM NaCl, 5 mM DTT, 5% glycerol, 6 mM $MgSO_4$, and 0.2% Tween 20) and incubated for 10 minutes on a grid in a homemade humidity chamber at 4 °C. The grid was sequentially incubated on 4, 50 μL drops of 2% uranyl formate solution for 5, 10, 15, and 20 seconds, and blotted dry with #1 filter paper (Whatman). Images were collected on a Jeol 1400 equipped with a Gatan 4k × 4x CCD camera at 30,000X magnification (3.71 Å/pixel), a defocus range of -1 to -2 μm, and 20 $e^-/Å^2$ total electron dose using Leginon [128].

Cryo-EM samples were prepared using C-Flat 4/1 400 mesh copper grids (EMS) covered with a thin layer of continuous carbon that were glow-discharged in air for 10 seconds with 5 W of power. Purified L-P was diluted to 0.25 μM using buffer B. For each sample, 3.5 μL was incubated with 0.05% glutaraldehyde for 5 minutes on ice in the dark. The sample was applied to a grid suspended in a Vitrobot operating at 4 C with 100% humidity. After 90 seconds, the sample was blotted with 25 force for 4 seconds and immediately plunged into liquid ethane cooled to liquid nitrogen temperatures. Images were collected using semi-automated data collection in Leginon [128] on a JEOL 3200FS microscope at 200 kV equipped with a Gatan K2 direct detector

and omega energy filter operating in super-resolution mode at a magnification of 30,000X (0.597 Å/pixel), defocus range from -1.5 to -4.5 μm, and a dose of 2.02 e^-/A^2/frame for 40 frames (Table 2.1, Figure 2.1A). Two data sets were collected on separately prepared samples with 1077 and 2607 micrographs, respectively.

2.2.5. Image processing

For negative-stained samples, particles were picked using DogPicker, extracted, and 2D classified using iterative MSA/MRA topological alignment within the Appion data processing software [129-132]. A particle stack of 96,043 particles with a box size of 80 x 80 pixels was subjected to iterative, multi-reference projection-matching 3D refinement using libraries from the EMAN2 software package, starting with a circular mask of 163 Å and increasing to 193 Å, 237 Å, and finally 282 Å [133].

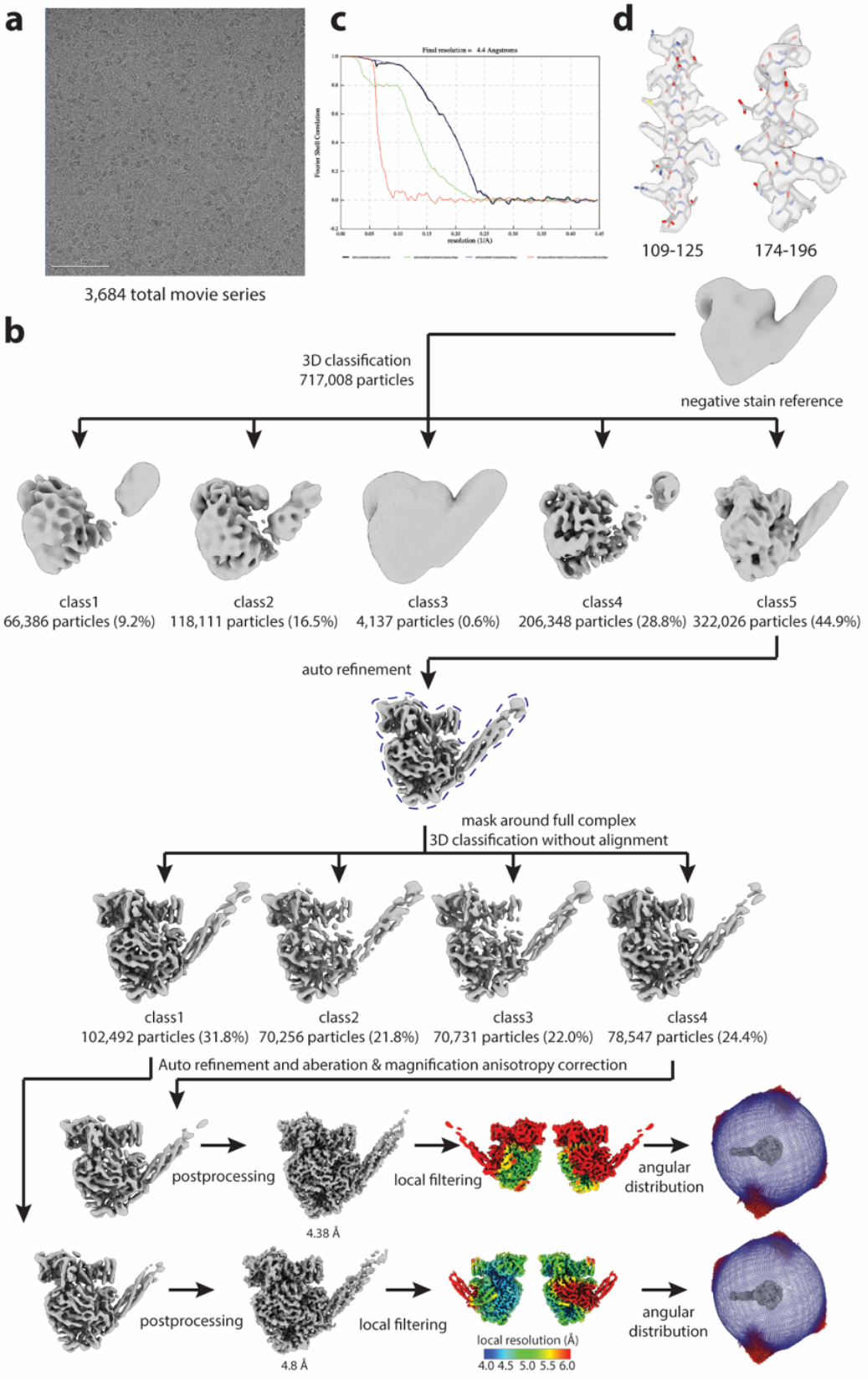
Cryo-EM micrographs were binned by two and motion-corrected using Motioncor2 with dose-weighting [134]. Particle coordinates were selected using gAutomatch (developed by Kai Zhang, MRC Laboratory of Molecular Biology, Cambridge, UK). gCTF version 0.50 was used for per-particle CTF estimation [135]. Particles were extracted using Relion 1.4, and all subsequent processing was done in Relion 2 or Relion 3 [136-138]. The initial set of 717,008 particles was extracted with a further binning by 2, resulting in a pixel size of 2.24 Å. 142 rounds of 3D classification using the negative stain volume as an initial reference, a circular radius of 300 Å, and a soft mask around the complex were performed (Figure 2.1B). Class 5 displayed signs of alpha-helices, so particles corresponding to that class were unbinned by a factor of 2 (pixel size 1.12 Å) and subjected to an automatic 3D refinement, which resulted in a reconstruction of 6.0 Å. 3D classification for 100 iterations without alignment was performed to sort out compositional or conformational heterogeneity. Classes 1 and 4 improved in their resolution after classification,

and each class was then refined separately and post-processed to 4.8 and 5.2 Å, respectively. Higher-order aberration and magnification anisotropy calculations were performed in Relion 3.1-beta, and re-refinement and post-processing of both classes resulted in final resolutions of 4.38 and 4.8 Å, respectively. All resolutions reported use the gold-standard Fourier shell correlation (FSC) at the 0.143 criterion [139].

Table 2.1. L-P cryo-EM data collection, refinement, and validation statistics

	PIV5 L-P class 1 (EMDB-21095) (PDB 6V85)	PIV5 L-P class 4 (EMDB-21096) (PDB 6V86)
Data collection and processing		
Magnification	30,000	30,000
Voltage (kV)	200	200
Electron exposure (e^{-}/Å^2)	76.5	76.5
Defocus range (μm)	-1.5 - -4.5	-1.5 - -4.5
Pixel size (Å)	1.12	1.12
Symmetry imposed	C1	C1
Initial particle images (no.)	717,008	717,008
Final particle images (no.)	102,493	78,547
Map resolution (Å)	4.38	4.63
FSC threshold	0.143	0.143
Map resolution range (Å)		
Refinement		
Initial model used (PDB code)		
Model resolution (Å)		
FSC threshold		
Model resolution range (Å)		
Map sharpening *B* factor (Å^2)	-197	-265
Model composition		
Non-hydrogen atoms	17895	17798
Protein residues	2240	2229
Ligands	ZN:2	ZN:2
B factors (Å^2)		
Protein	77.11	127.37
Ligand	280.69	280.69
R.m.s. deviations		
Bond lengths (Å)	0.006	0.004
Bond angles (°)	0.878	0.825
Validation		
MolProbity score	2.47	2.49
Clashscore	21.82	23.19
Poor rotamers (%)	0.40	0.70
Ramachandran plot		
Favored (%)	86.05	86.63
Allowed (%)	13.54	13.14
Disallowed (%)	0.41	0.23

Figure 2.1. **L-P cryo-EM data collection and image processing.** A) Representative electron micrograph of L-P. Individual particles are monodisperse and easily visible by eye. B) All 717,008 initial particles were subjected to 3D classification in Relion-2 using a negative stain volume of L-P as an initial reference. Particles were binned by 2 for an initial pixel size of 2.24 Å to improve processing time. Following many rounds of classification, particles corresponding to class 5 were selected, un-binned to a pixel size of 1.12 Å, and auto-refined. A second round of classification, this time without alignment, was performed and the best two classes, classes 1 and 4, were individually auto-refined, corrected for higher-order aberrations and magnification anisotropy, and post-processed to generate final maps. Class 1 was used to build the atomic model of L-P. The core of both classes is at the highest resolution, with P-OD being at a significantly lower resolution than the rest of the complex. The final reconstructions are free of preferential views, as indicated by the angular distribution plots. C) FSC curves for the final L-P refinement of class 1. D) Two representative helices showing the final model docked into the electron density.


a
3,684 total movie series
c
Final resolution = 4.4 Angstroms
Fourier Shell Correlation
resolution (1/A)
d
109-125
174-196
b
3D classification
717,008 particles
negative stain reference
class1
66,386 particles (9.2%)
class2
118,111 particles (16.5%)
class3
4,137 particles (0.6%)
class4
206,348 particles (28.8%)
class5
322,026 particles (44.9%)
auto refinement
mask around full complex
3D classification without alignment
class1
102,492 particles (31.8%)
class2
70,256 particles (21.8%)
class3
70,731 particles (22.0%)
class4
78,547 particles (24.4%)
Auto refinement and aberation & magnification anisotropy correction
postprocessing
4.38 Å
local filtering
angular
distribution
postprocessing
4.8 Å
local filtering
local resolution (Å)
4.0 4.5 5.0 5.5 6.0
angular
distribution

2.2.6. Model building

A homology model for the PIV5 P-XD was generated using the MuV P-XD crystal structure [44]. Rigid body docking into the EM density was unambiguous.

A homology model of the PIV5 L protein was generated using the VSV L structure [11]. This model was fit as a rigid body into the class 1 EM density using the RdRp and PRNTase domains only. The CD was fit as a rigid body into its position in the density. This required a small rotation away from the RdRp-PRNTase module. A homology model of the MTase-CTD module using the hMPV crystal structure was generated and, along with a homology model of the VSV MTase-CTD module, was fit as a rigid body into its novel position in the PIV5 density. The model of L was built manually in Coot [140], utilizing the homology model as a guide. Flexible loops and portions where the density was not well resolved were left unmodeled if the C-alpha path was ambiguous.

In many cases, there are differences in the secondary structure between PIV5 and VSV. In these cases, bulky side chains visible in the density and secondary structure prediction was utilized to determine the correct register of the polypeptide. The zinc-binding sites identified in the VSV structure were conserved and displayed density where the zinc ions are located, suggesting the presence of metal ions in our density. These sites were built according to the corresponding sites in VSV. Ramachandran outliers and other errors introduced during manual model building were fixed before multiple rounds of Phenix real-space refinement [141]. The resulting complete model, including P, was refined using Phenix real space refine.

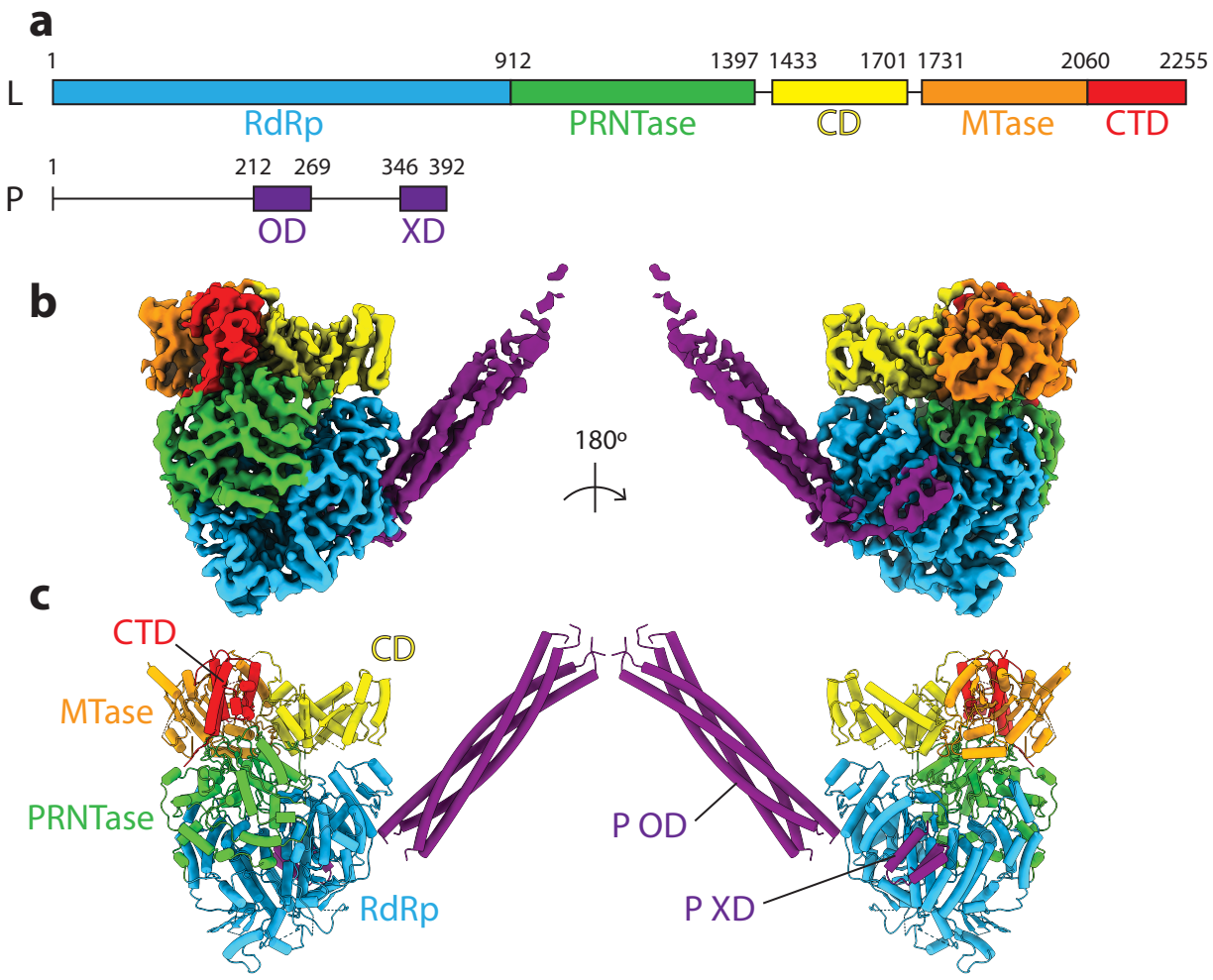
For class 4, the CD-MTD-CTD module was fit as a rigid body into the density and re-refined using Phenix real space refine [141]. All visualization, figures, and movies were generated in UCSF Chimera X [142].

2.3. Results

2.3.1. Cryo-EM structure of the PIV5 L-P complex

Single-particle cryo-EM of the purified complex gave rise to a 3D reconstruction at 4.3 Å resolution (Figures 2.1B, 2.2, Table 2.1). Local resolution estimation for the entire complex ranged from 4.0 Å near the core of the RdRp and PRNTase domain to >6.0 Å for the P-OD (Figure 2.1B). All domains of L could be unambiguously assigned to the density. Four copies of P form a long four-helix bundle that extends from the RdRp domain with a single P-XD bound near the nucleotide entry tunnel of L (Figure 2.2). I generated a homology model for PIV5 L based on the VSV L structure and used it as a guide to manually build a nearly complete model from residues Arg5 to His2225. For simplicity, all comparisons of the PIV5 structure to *Pneumoviridae* L-P structures will use the RSV structure because of the high similarity between the RSV and hMPV L-P structures [25, 29, 32]. Any reference to hMPV structures refers to the crystal structure of the hMPV MTase-CTD dimer [24]. Similarly, all comparisons of the PIV5 structure to *Rhabdoviridae* L-P structures will use the VSV structure because of the high similarity between the VSV and RaV L-P structures [11, 27, 28].

Figure 2.2. **The architecture of the PIV5 L-P complex.** A) Domain diagrams of PIV5 L and P proteins. RdRp, RNA-dependent RNA polymerase (cyan); PRNTase, poly-ribonucleotidyltransferase (green); CD, connecting domain (yellow); MTase, methyltransferase (orange); CTD, C-terminal domain (red); P-OD, oligomerization domain (purple); P-XD, X domain (purple). B & C) Electron density (B) and atomic model (C) of the PIV5 L-P complex with domains colored as depicted in (A).



2.3.2. Structural architecture of the PIV5 L RdRp-PRNTase module

The RdRp domain adopts a similar architecture, composed of fingers, palm, and thumb sub-domains, as observed in other viral polymerases (Figure 2.3A). The PRNTase domain is located above the RdRp, in the same position that it occupies in the VSV and RSV structures (Figure 2.4). The PRNTase domain also contains CR IV and adopts a very similar fold in the VSV and RSV structures except for differences in some key areas (Figure 2.3B). The priming loop is highly flexible in our PIV5 structure, but I could trace its path, which adopts a non-initiation conformation similar to the RSV priming loop (Figure 2.5A). This positions the GxxT (1218-1220) motif much closer to the HR motif, in a conformation that likely could not accommodate the guanosine diphosphate. The HR motif is located within a flexible intrusion loop, which does not impinge on the central cavity in the RSV and VSV structures (Figure 2.5). Instead, it projects out into the central cavity in our structure and would clash with the position of the priming loop in the initiation conformation. Displacement of the intrusion loop would be required to accommodate RNA in the active site [33], suggesting a possible tug-of-war between the priming loop and intrusion loop that could regulate transcription initiation.

Figure 2.3. **Domain comparisons between PIV5, VSV, RSV, and hMPV.** A) PIV5, VSV, and RSV all form the canonical right-hand fold found in viral polymerases. The palm is stabilized by a β-sheet, the fingers are a largely alpha-helical segment, and the thumb is formed by a few long helices at the very C-terminus of the RdRp domain. B) Comparison of the PRNTase domains between PIV5, VSV, and RSV. The overall folds are quite similar, with some critical differences around the conserved HR and GxxT motifs. C) The connection domains are the least conserved among NNS viruses, but conserved secondary structure elements common to PIV5 and VSV are present. D) The center of the MTase domain is formed by a highly conserved β-sheet, which contains some of the catalytic K-D-K-E triad residues found in these viruses. E) The CTD of PIV5, VSV, and hMPV are not very well conserved except for the α6 helix, which contains the KxxxKxxG motif found in PIV5 and hMPV. In VSV, the first lysine is not conserved, and the middle lysine is replaced by arginine.

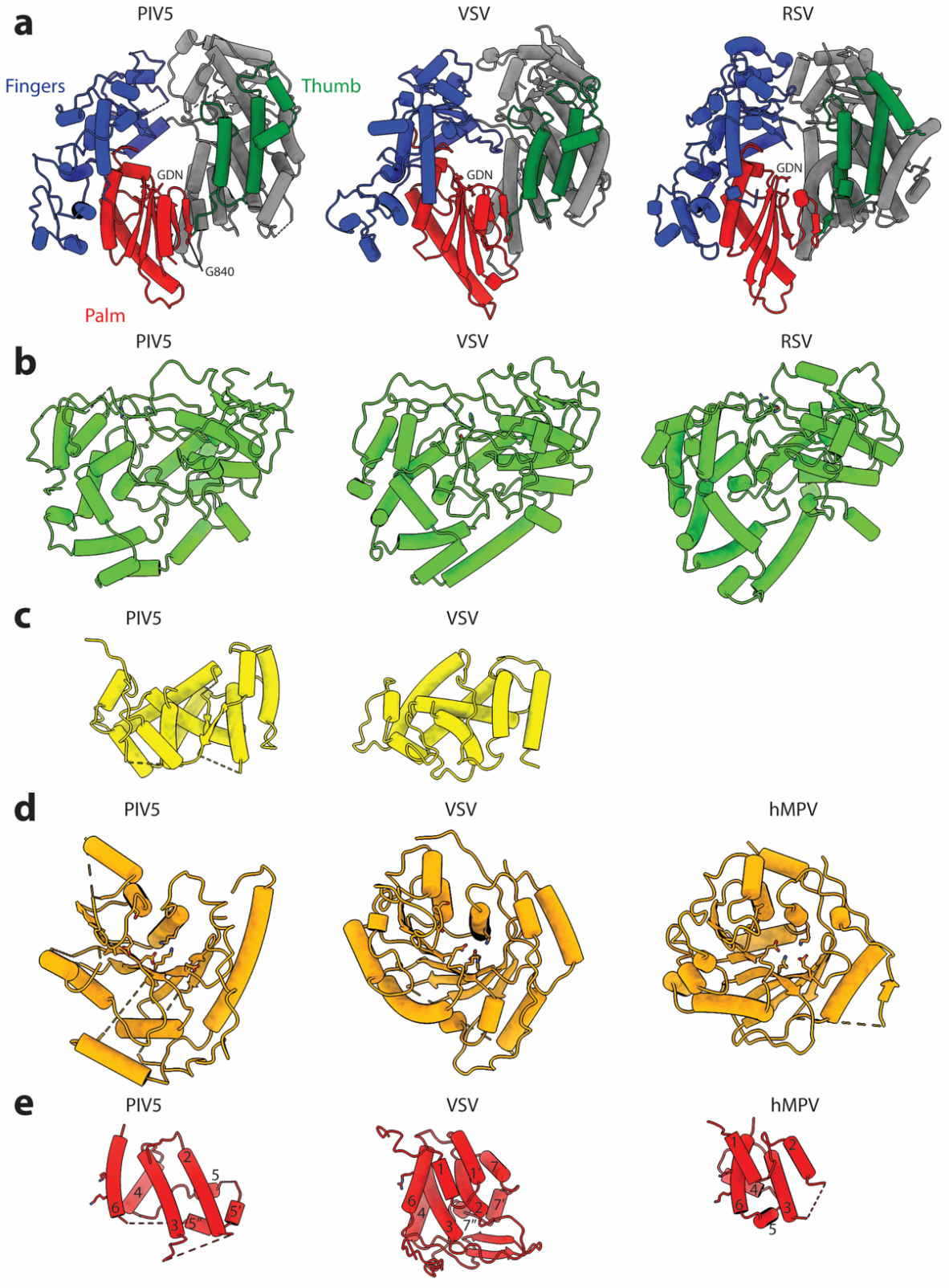

a
PIV5
VSV
RSV
Fingers
Thumb
GDN
GDN
GDN
G840
Palm
b
PIV5
VSV
RSV
c
PIV5
VSV
d
PIV5
VSV
hMPV
e
PIV5
VSV
hMPV

Figure 2.4. **Comparison of PIV5, VSV, and RSV L-P complexes.** A-C) PIV5 (A), RSV (B), and VSV (C) L-P complexes were aligned based on the RdRp domain (cyan). The PRNTase domains (green) are in similar positions relative to the RdRp in all three structures. The P-OD of PIV5 is significantly longer than the P-OD of RSV and protrudes further away from the RdRp. The PIV5 P-XD is in roughly the same position as the single C-terminal helix in the RSV structure. There are no visible P linker helices in the PIV5 structure as there are in the RSV structure. The large rearrangement of the methyltransferase (MTase, orange) and C-terminal domains (CTD, red) is visible between the PIV5 and VSV structures. The relative orientations of the domains are shown in the insert with the arrowhead representing the direction of the protein backbone. CD, connecting domain (yellow); P protein (purple).

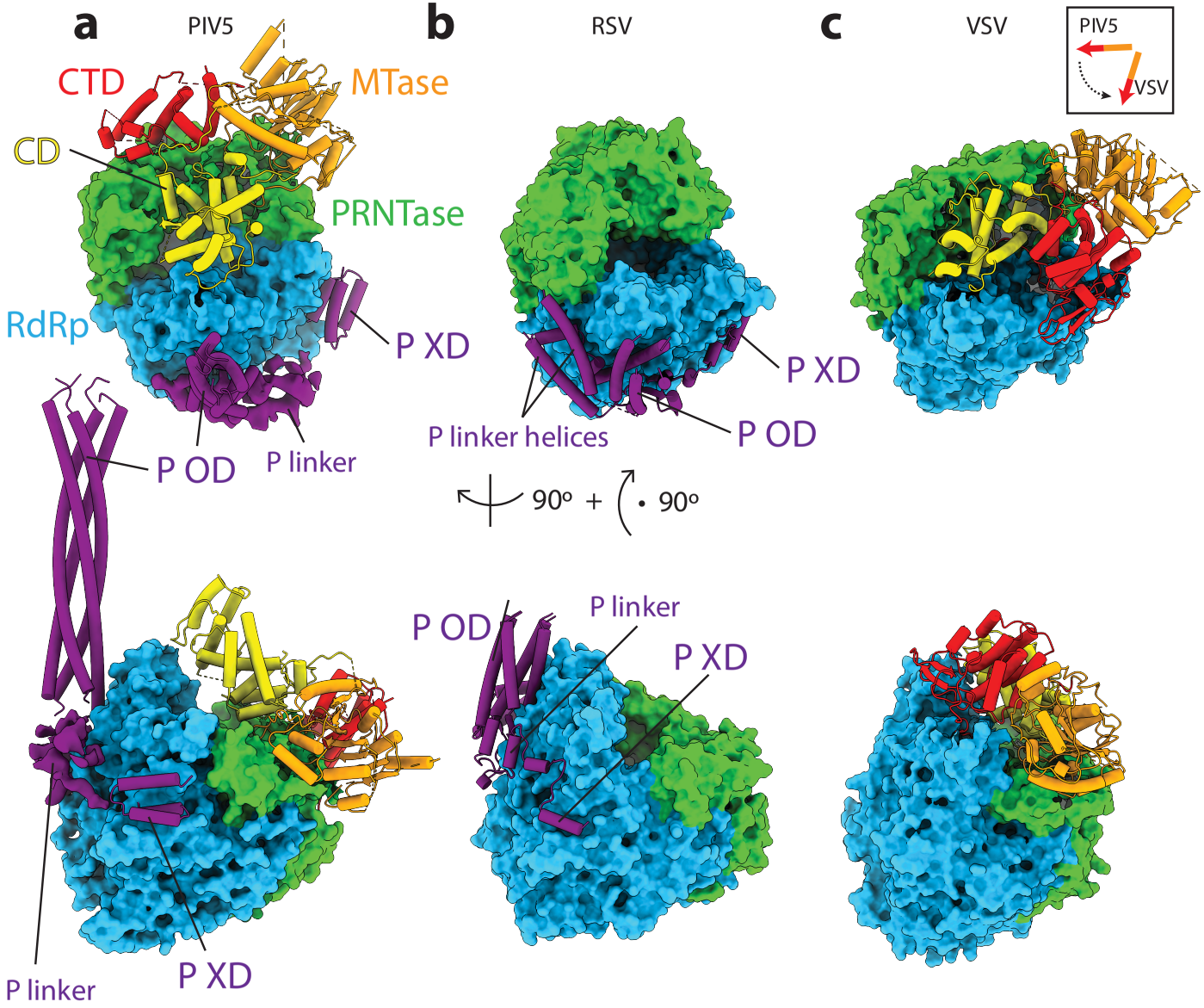
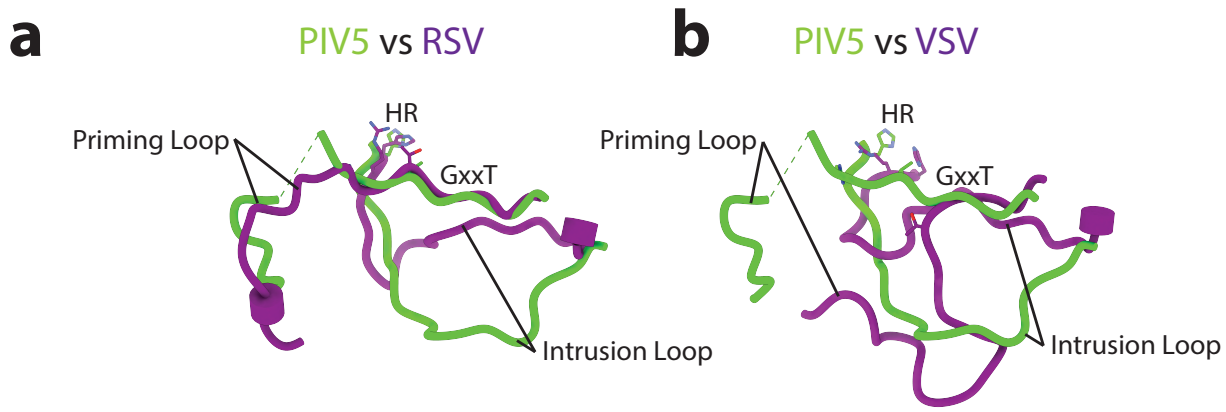

a
PIV5
CTD
MTase
CD
PRNTase
RdRp
P XD
P OD
P linker
P linker
P XD
b
RSV
P linker helices
P XD
P OD
90° + 90°
P OD
P linker
P XD
c
VSV
PIV5
VSV

Figure 2.5. **Comparison of the PIV5, RSV, and VSV priming loop and intrusion loop.** A) The PIV5 priming loop adopts the same elongation conformation as in the RSV structure. The intrusion loop projects out into the central cavity between the RdRp and PRNTase domains. B) In the VSV structure, the priming loop in the initiation conformation would sterically clash with the position of the PIV5 intrusion loop. The VSV and RSV intrusion loops are in the same position with minor differences.

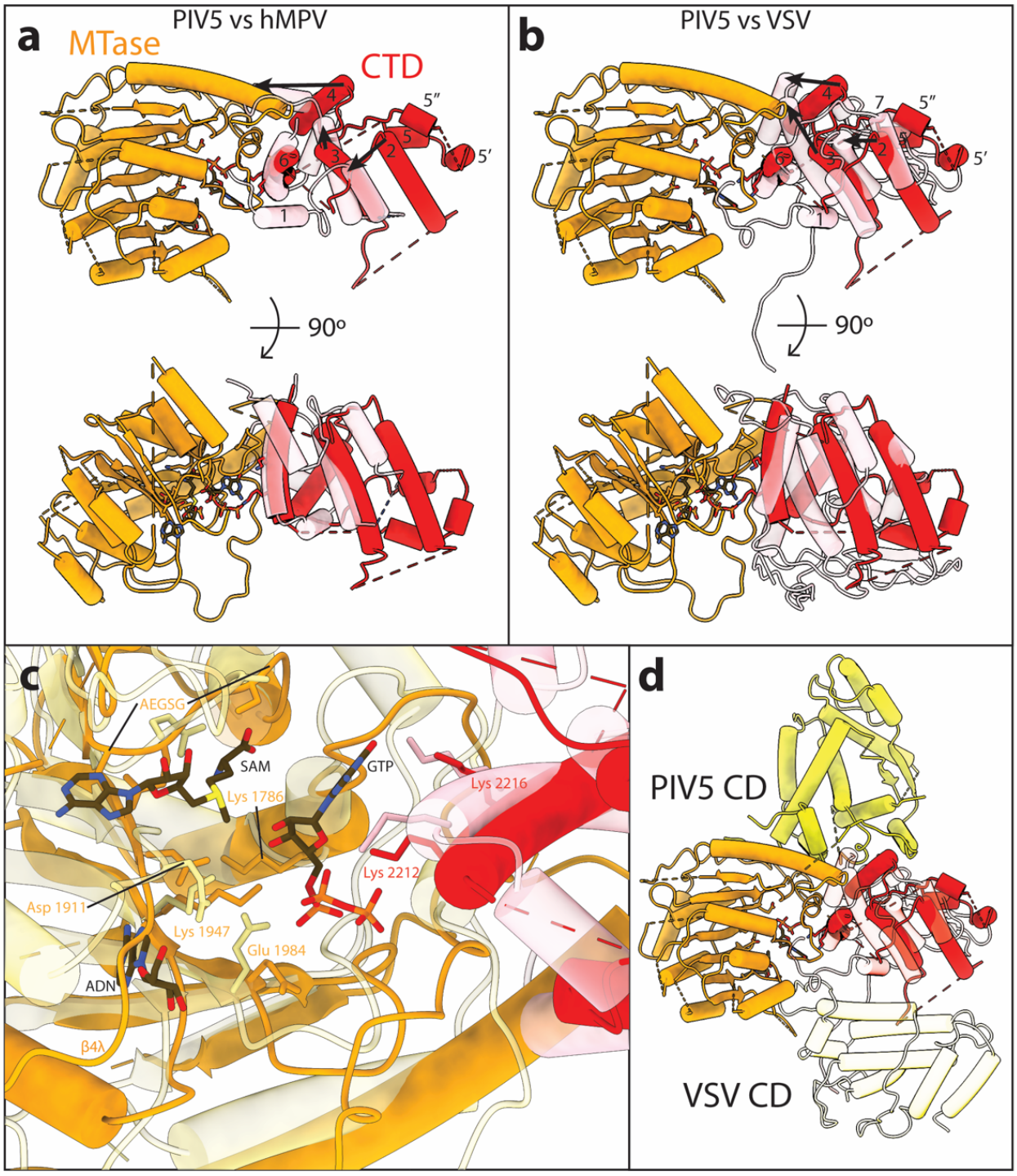


2.3.3. The MTase and CTD domains adopt a unique conformation

The MTase and CTD domains of NNS viruses are implicated in capping and methyltransferase activities [22, 24, 37]. The MTase domain adopts a very similar fold across all three structures, consistent with its dual function in methylation and capping (Figure 2.3D). The CTD of PIV5 differs from VSV and hMPV in that it does not contain the α1 helix (Figure 2.3E, 2.6A-B). Instead of an α5 helix, PIV5 contains multiple small helices named α5, α5', and α5'' that occupy a similar location as a β-sheet in the VSV CTD. The concerted shifts of helices α2 and α3 in the PIV5 structure relative to hMPV are likely driven by interactions with the CD that is absent in hMPV (Figure 2.6A and D). The flipping of helix α4 in the hMPV structure is likely prevented in PIV5 due to stabilization by the CD. The presence of the CD on the other face of the MTase-CTD dimer in the VSV structure shifts helices α2, α3, and α4 even further than the hMPV structure from their PIV5 location (Figure 2.6B and D). Together these results highlight the role of the CD in influencing the packing of the CTD.

Superposition of the hMPV MTase domain positions all conserved residues in the same location in our PIV5 structure, suggesting a conserved binding mode for substrates SAM and GTP (Figure 2.6C). Despite the significant changes in the packing of the CTD, the α6 helix, which contains the conserved KxxKxxG motif, maintains its location relative to the MTase active site, supporting the role of this motif in stabilizing substrates within the active site.

Figure 2.6. **Detailed comparison of the MTase and CTD folds and active site.** A) Comparison of the fold of the CTD (red) between PIV5 (opaque) and hMPV (transparent). The MTase-CTD module is aligned using only the MTase domain, and the MTase of hMPV is hidden for simplicity. Only a small shift of α2, a flip of α4, and an expansion of α5 are present between the two structures. B) Comparison of the fold of the CTD between PIV5 (opaque) and VSV (transparent). The MTase-CTD module is aligned using only the MTase domain, and the MTase of VSV is hidden for simplicity. More extensive shifts of α2, α3, and α4 are present. Helices α5, α5', and α5'' of PIV5 are replaced by a β-sheet in VSV. C) Close up of the active site of the MTase of PIV5 (opaque) and hMPV (transparent). The conserved motifs K-D-K-E, A/GxGxG, and KxxxKxxG are shown as side chains. The position of substrates SAM, GTP, and ADN from the hMPV structure are superimposed. D) Comparison of the location of the CD relative to the MTase-CTD dimer between PIV5 (opaque) and VSV (transparent). This rearrangement is likely responsible for the shifting of secondary structure elements highlighted in A and B.


a
PIV5 vs hMPV
MTase
CTD
5″
5′
90°
b
PIV5 vs VSV
7
5″
5′
90°
c
AEGSG
SAM
GTP
Lys 1786
Lys 2216
Lys 2212
Asp 1911
Lys 1947
Glu 1984
ADN
β4λ
d
PIV5 CD
VSV CD

The position of the MTase-CTD dimer relative to the RdRp-PRNTase module in my PIV5 structure is significantly different from the VSV structure (Figure 2.7). In PIV5, the MTase active site sits directly on top of the PRNTase domain, resulting in a distance between the conserved HR motif and the MTase active site of 25.7 Å, rather than 52.5 Å in the VSV structure. This rearrangement of the MTase-CTD module results in entirely different interfaces between it and the RdRp-PRNTase-CD module. It is likely facilitated by the flexible linker region between the CD and MTase domains.

In the recently published pneumovirus L-P complex structures, the CD-MTase-CTD module is not visible in the density, even though full-length proteins were used, and MTase activity was shown to be present (Figure 2.4A-B) [25, 26, 29]. This suggests that these three domains can separate from the RdRp-PRNTase module and rearrange before re-associating. I was able to further separate the particles into two classes that slightly differed in the position of the CD-MTase-CTD module relative to the RdRp-PRNTase module, showing that there is a tendency of these modules to separate from each other (Figure 2.8). The observed movement of the CD-MTase-CTD module indicates that this module mainly behaves as a rigid body, leading to the conclusion that the CD does not stay stably associated with the RdRp-PRNTase module while the MTase-CTD undergoes the conformational rearrangement. This is consistent with the disappearance of the CD in the pneumovirus structures and enabled by the long flexible linker between the RdRp and CD domains.

Figure 2.7. **Comparison of the relative position between the conserved HR motif and the MTase active site between the PIV5 and VSV structures.** The MTase and CTD are positioned directly above the PRNTase domain in the PIV5 structure (opaque). The distance between the HR motif and the GTP modeled into the MTase active site is 25.7 Å. In the VSV structure, the distance increases to 52.5 Å due to the active site projecting away from the RdRp domain.

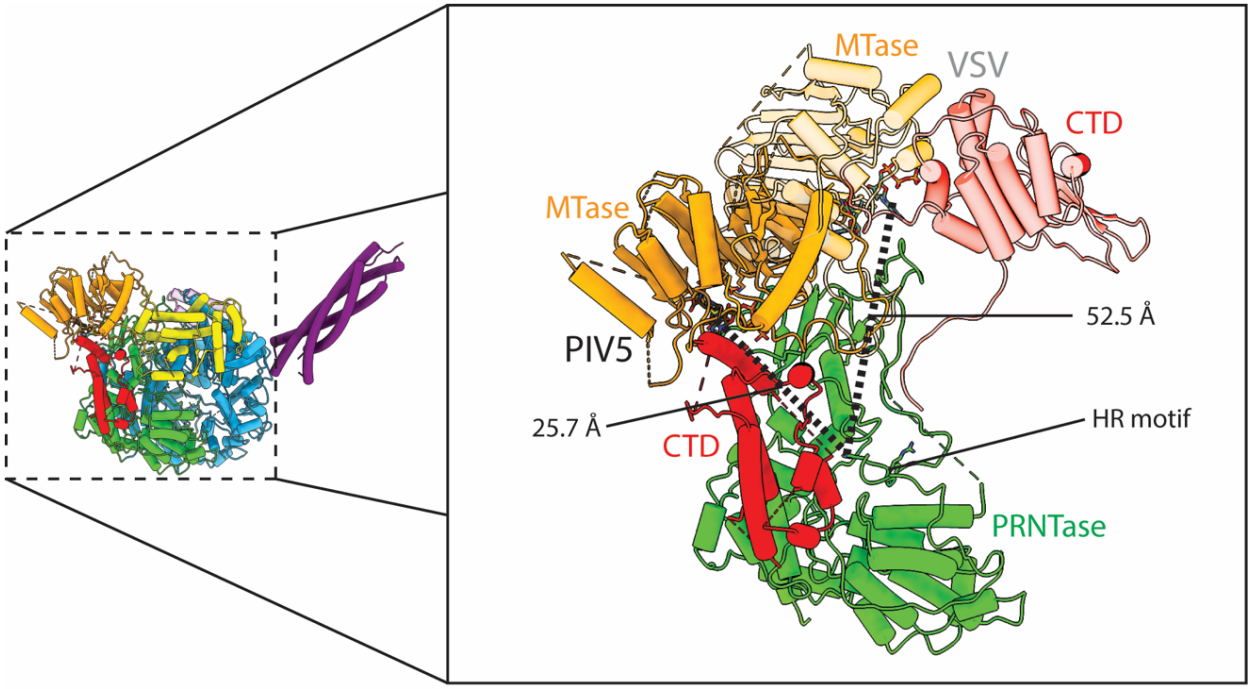
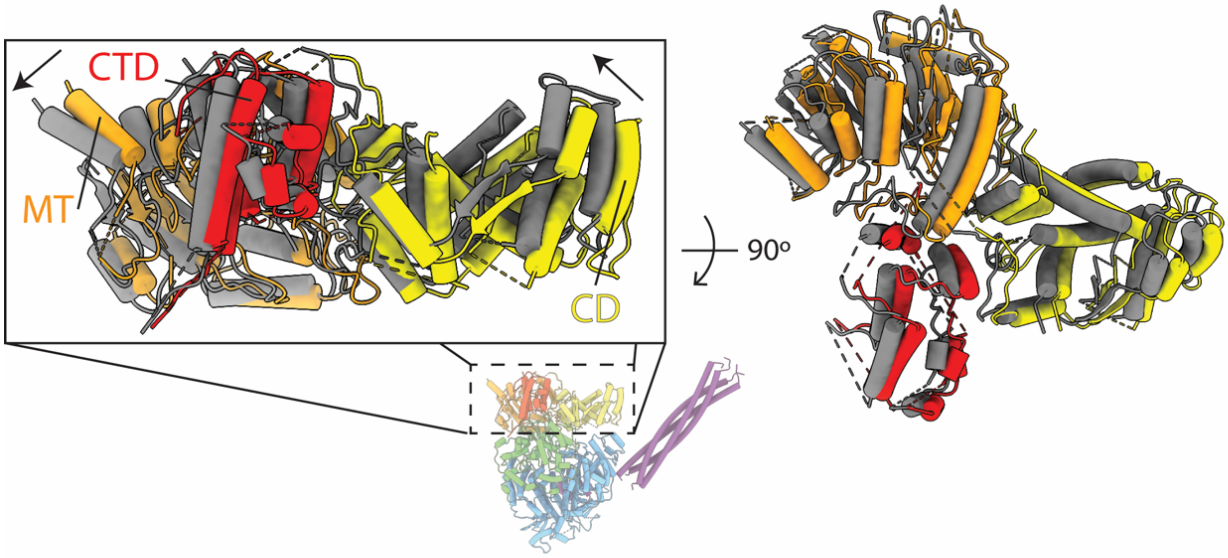


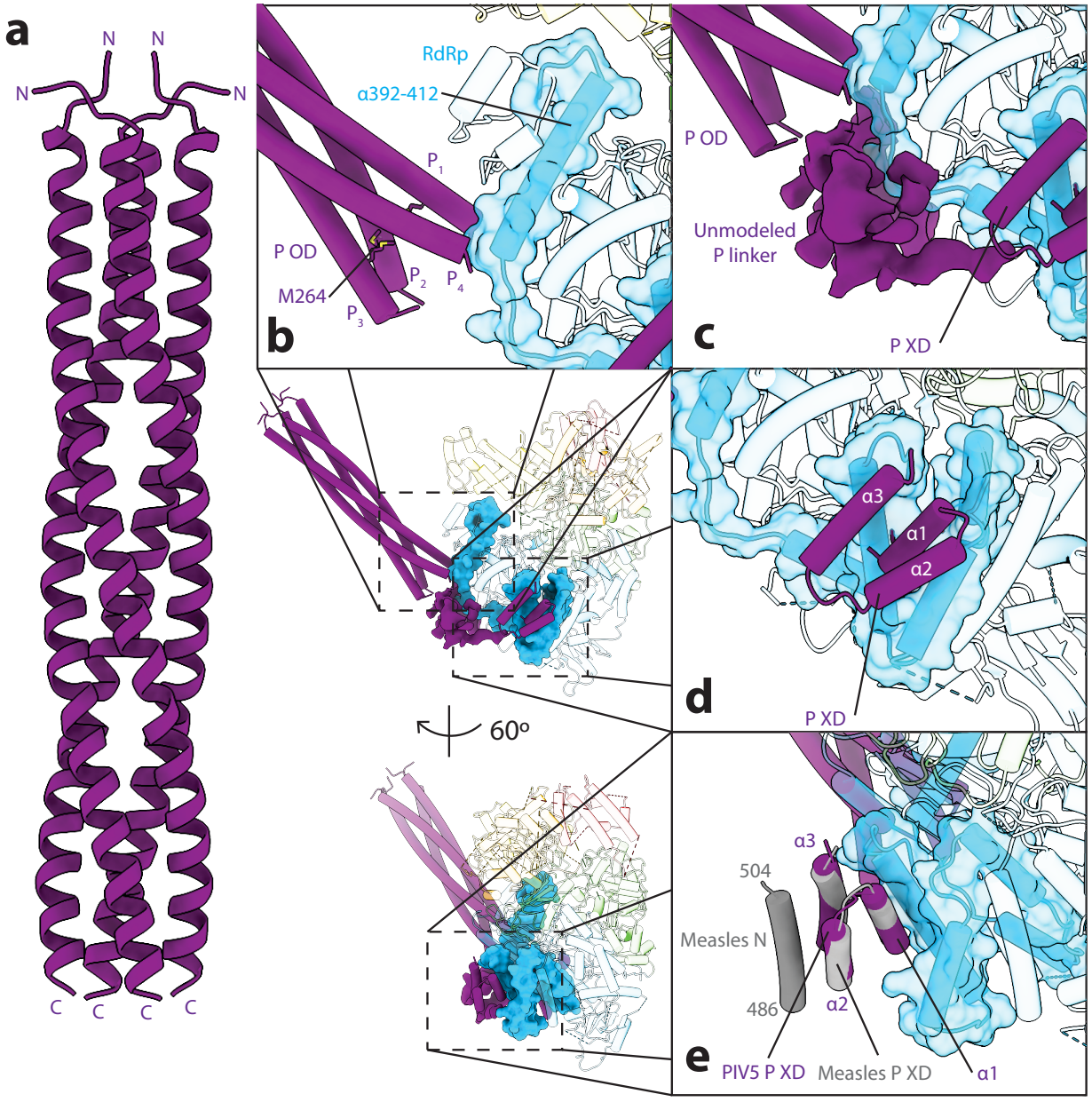
Figure 2.8. **The flexibility of the CD-MTase-CTD module.** Comparison of the position of the CD-MTase-CTD module between the two best classes obtained from 3D classification. The movement behaves mostly as a rigid body, up and away from P-OD, and allows for considerable changes across the interface between this module and the RdRp-PRNTase module.



2.3.4. Structure of tetrameric PIV5 P-OD

The PIV5 P-OD is located in the middle of the P protein and is responsible for oligomerization (Fig. 2.2A). Disrupting the stability of the coiled-coil region or the kink near residues 339-341 in MeV P-OD results in the disruption of MeV gene expression [143]. A tetrameric helical bundle is visible in my reconstruction, directly associating with the RdRp domain of L and protruding away from it (Figure 2.2B-C). I could not unambiguously determine the orientation of each chain of P within this density due to the limited resolution in this region. A postdoc in Robert Lamb's lab, Megha Aggarwal, crystallized the P-OD from residue 172 to 278 and determined the structure at 1.4 Å resolution (Figure 2.9A, 2.10A). This structure confirmed that PIV5 P forms an all-parallel four-helix bundle with two long parallel helices in the asymmetric unit. The all-parallel orientation of P-OD is consistent with the vast majority of paramyxovirus P-OD crystal structures that have been determined to date. Residues 198-271 were visible in the crystallographic electron density map, and the coiled-coil region comprises residues 203-270. The residues immediately N-terminal to the P-OD from NiV and SeV form a helical cap. However, I could not identify a similar feature in either of our crystallographic or cryo-EM density maps. Sequence alignments suggest that those N-terminal helices in NiV and SeV P are insertions and absent in the PIV5 P protein.

Figure 2.9. **Interaction interfaces of PIV5 L and P proteins.** A) Crystal structure of the oligomerization domain (OD) of the PIV5 P protein. The OD forms an all-parallel four-helix bundle with one helix from each of four chains of P. B) Interfaces between P-OD (purple) and L (cyan). The L fragment that is necessary and sufficient to interact with P is shown as an opaque surface; the rest of L is shown as a transparent cartoon. Helix 392-412 is the only portion of L that interacts with P-OD. C) Interaction between the unmodelled P density and L. This density is not as well resolved as the OD or XD and does not form extensive contacts with L except at the base of P-OD. D&E) Interaction of P-XD (residues 346-392) and L. Helices α1 and α3 of P-XD form the interface of the XD with L. The portion of L that interacts with P-XD spans residues 303-350. Superposition of Measles virus P-XD bound to a C-terminal fragment of N (486-504) with PIV5 P-XD (E).


a
N
N
N
N
C
C
C
C
b
RdRp
α392-412
P1
P OD
M264
P2
P4
P3
c
P OD
Unmodeled
P linker
P XD
60°
d
α3
α1
α2
P XD
e
α3
504
Measles N
486
α2
PIV5 P XD
Measles P XD
α1

2.3.5. A Bipartite interface on L for binding P-OD and P-XD

To fit the crystal structure of P-OD into my cryo-EM density map, I needed first to determine the correct orientation of P-OD relative to the EM density. I was able to take advantage of the identification of a single P-XD associated with L, which will be described in detail later. Even though the density between P-OD and P-XD is poorly resolved, P-OD must anchor to L at its C-terminus due to the P-XD location, resulting in the N-terminus of P-OD projecting away from L (Figure 2.9B). The distance between the first residue of P-XD (L342) and the end of P-OD that anchors to L is 43 Å, while the distance to the end of P-OD that is far away from L is 145 Å. With 77 residues between P-OD and P-XD, the more plausible orientation of P-OD is with the C-terminus binding to L. This is consistent with the orientation of the P-OD of RSV and hMPV [25, 26].

2.3.5.1. P-OD forms minimal interactions with L

With the correct orientation of P-OD determined, the crystal structure could be docked into my EM density as a rigid body without significant changes. The C-terminus of chains P_1 and P_4 of P-OD interacts with a single helix of L formed by residues 392-412 (Figure 2.9B). The RSV P-OD, although significantly shorter than the PIV5 P-OD, also engages with the homologous helix of L (residues 455-476, Figure 2.4A-B). This helix maps perfectly to the end of a 408-residue fragment of the closely related MuV L, which was shown to be both necessary and sufficient for interacting with P [53]. The interaction between SeV L and P required a larger fragment of L, residues 1-1146 [23]. This is significantly larger than MuV, but smaller fragments of L may make helix 392-412 unstable, preventing stable association. Replacement of the MeV P-OD up to residue 360 (264 in PIV5) with the GCN4 tetrameric domain did not affect the binding of L and P,

consistent with my structure where L only binds to P-OD C-terminal of residue 264 (Figure 2.9B) [58].

2.3.5.2. The flexible linker between P-OD and P-XD is unstructured

There is electron density for the flexible linker C-terminal of P-OD that I could not assign to any individual chain of P with confidence or build a model for (Figure 2.9C). This flexible linker region associates with loop (384-391) of L, immediately C-terminal of helix 392-412. Replacement of the homologous region of the MeV P with the GCN4 tetrameric domain disrupted the interaction between L and P, highlighting the importance of this interface [58]. The flexible linker appears to correspond to P_4 based on homology to RSV, although my density is not resolved enough to be confident in this assignment (Figure 2.9B-C). The path of the linker is also different than RSV, where it interacts with helix 829-850 of L (Figure 2.10B-C). In PIV5, the density loops away from L, and the lack of a stabilizing interaction with L is likely the cause of its low resolution. I see no evidence for additional interactions between the P linker helices of P_1, P_2, and P_3, like those present in RSV (Figure 2.4A-B).

Figure 2.10. **Atomic details of P.** A) Electron density and atomic model of P determined by x-ray crystallography. Only a representative portion of the full OD is shown for simplicity. B&C) Comparison of the path of the linker regions in between P_4 (purple) and the XD of PIV5 and RSV. In PIV5, this region loops away from L, forming minimal interactions (B). In RSV, this region forms extensive interactions with L, stabilizing this region (C). D) Close up of critical residues involved in the interaction between PIV5 L and P-XD. A hydrogen bond network between P-T349, L-H315, P-Q352, and L-Q349 forms the structural core of the interface. E) In the RSV structure, the C-terminal helix of P shifts down one turn of the predominant interacting helix of L (350-378), and the interaction is stabilized through hydrophobic interactions between P-S220, P-L223, L-L361, L-T365, and L-I398.


a
b
c
d
e
H315
Q349
Q352
T349
350-378
L361
S220
L223
I398
T365

2.3.5.3. Interaction between P-XD and L

As mentioned earlier, a single P-XD is bound to L near the nucleotide entry site (Figure 2.9D). A homology model of PIV5 P-XD based on the MuV P-XD structure fits perfectly into my EM density, with the N-terminus of the P-XD naturally connecting to the unmodeled P linker density. Helices α1 and α3 of P-XD form the interface that interacts with L, consistent with mutational studies of MeV P [58]. The part of L that interacts with P-XD spans residues 303-350. Interestingly, the N-terminal 298 residues of MeV L were not sufficient to bind P, defining a fragment of L from 303-412 as the region that is sufficient to bind MeV P [53]. This region is a subset of CR I, which spans residues 227-419. A separate study on MeV showed that P-XD residues V463, S466, and H498 are essential for interacting with L [58]. The structurally equivalent residues in PIV5 P-XD, T349, Q351, and K384, are all in direct proximity to the surface of L (Figure 2.10D). T349 and Q351 form a hydrogen bond network with residues H315 and Q349 of L. These interactions are conserved in the closely related hPIV2 but not in other paramyxoviruses. For example, a T349M mutation in simian virus 41 (SV41) P has a corresponding H315Y mutation in L, creating a pi-aromatic interaction that would also stabilize this interface. In MeV, a hydrophobic core is created by P-V463, L-L305, P-S466, and L-I339. These findings are consistent with the biochemical studies on MeV and suggest that this interface may have co-evolved and is maintained beyond just the rubulaviruses [58]. RSV lacks the same P-XD fold as PIV5 and other paramyxoviruses; instead, a single helix of RSV P is found in roughly the same position as α1 of the PIV5 P-XD (Figure 2.10E). There are only an additional 13 residues C-terminal to this helix of RSV P, not enough to form an additional two helices. The interactions between this helix of RSV P and L shifts down one turn of helix 350-378 (+3 residues) of L and forms a hydrophobic core containing residues P-S220, P-L223, L-L361, and L-I398 that stabilizes the interface (Figure 2.10E). Because the interaction between L and P does not involve a large

interface, it could allow transient binding of P to L, which would be essential for an association-disassociation mechanism of L-P interaction.

2.3.6. Dual binding surfaces on P-XD

Helices α1 and α3 of P-XD directly interact with L, but helix α2 is solvent-exposed (Figure 2.9D). Interestingly, MeV P-XD binds the MoRE motif of N through an interface formed by α2 and α3 [58]. Superposition of the MeV P-XD-N-MoRE structure with PIV5 P-XD reveals that this association does not interfere with its interaction with L, and there are no significant structural changes in P-XD as well (Figure 2.9E). Thus, there are two completely independent binding sites on P-XD for L and N. This could provide a stable contact between L and N to allow recycling of N monomers while they are displaced from the genome.

CHAPTER 3:

Structure of a human transcription initiation complex

3.1. Introduction

Despite many efforts to structurally characterize Mediator-bound PICs [67, 106, 108], a mechanism for the phosphorylation of the Pol II CTD is missing due to the lack of structural information for the CAK module of TFIIH within the full complex. Atomic resolution models are missing for most of MedTail, an essential module for the recruitment of Mediator to specific genomic locations by transcription factors, which will inform how enhancers and promoters come together within the nucleus. Also, understanding the conformational landscape of this large complex is central to understanding the role of flexibility in the mechanism of initiation.

In this study, I present the first structure of a human Med-PIC assembled on a closed promoter DNA construct, with the TBP subunit replacing the full TFIID complex. Human Mediator is held together by a central scaffold subunit, Med14, which forms two contact sites with MedTail. The precise orientation of the CAK module within Med-PIC is revealed, with clear density for the Pol II CTD in the active site. A second CTD binding site between MedHead and MedMiddle shows how Mediator positions the rest of the CTD for phosphorylation by CDK7. Many regions of Mediator that interact with transcription factors are flexibly tethered, facilitating its assembly. The structure also provides key insights into the conformational landscape of Mediator relative to the PIC.

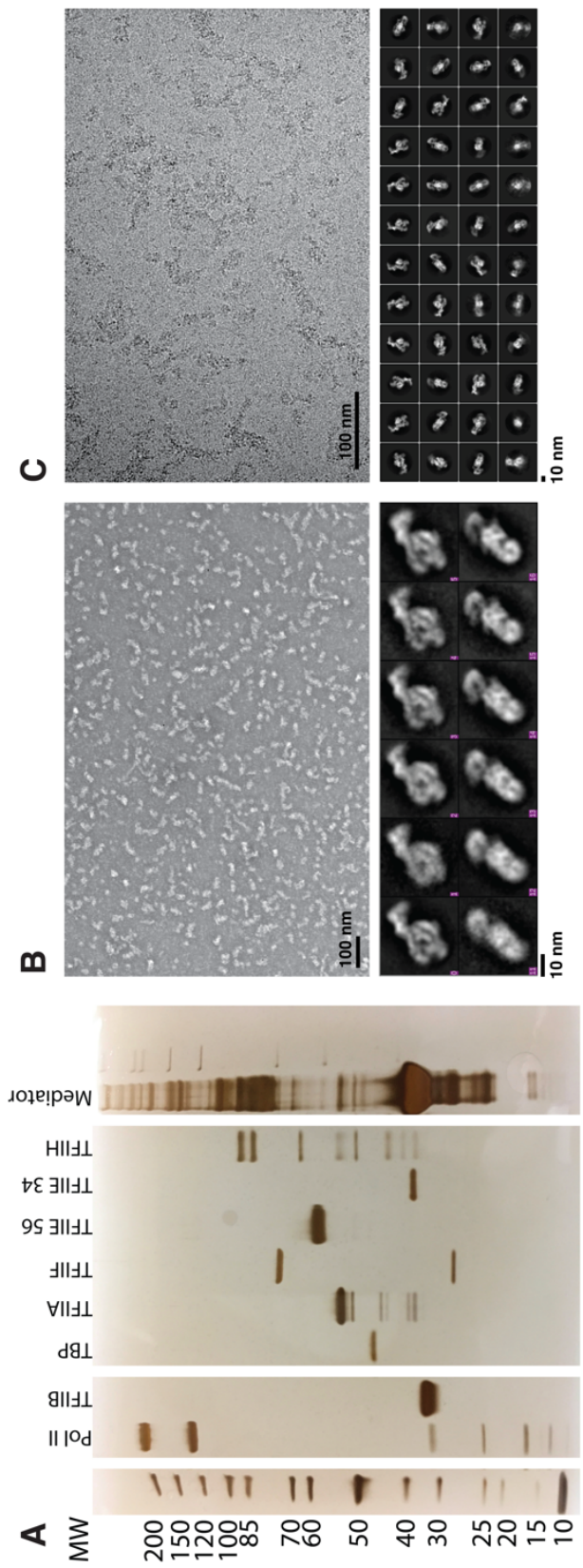
3.2. Methods

3.2.1. Purification of PIC components

Pol II and TFIIH were purified endogenously from HeLa cells, and TFIIB, TFIIA, TBP, TFIIE, and TFIIF were purified recombinantly, as previously described (Figure 3.1A) [66, 92, 126]. Mediator was purified endogenously from HeLa cells, as previously described [92]. In short, HeLa cell nuclear extract was run over a phosphocellulose column using 0.1M KCl HEMG (20 mM

HEPES, 10 mM EDTA, 2 mM $MgCl_2$, and 10% glycerol). Stepwise elution of protein complexes was performed at 0.1 M, 0.3 M, 0.5, M and 1.0 M HCl HEMG. The 0.5 M and 1.0 M elutions were dialyzed against 0.1 M HEMG before being subjected to further affinity purification using a GST-VP16 fusion protein bound to glutathione Sepharose resin (GE). Following a 3-hour incubation with the affinity resin, the resin was washed 5 times with 50 column volumes of 0.5 M KCl HEGN (20 mM HEPES, 10 mM EDTA, 10% glycerol, 0.1% NP-40), followed by one wash with 50 column volumes of 0.15 M KCl HEGN (0.02% NP-40). Mediator was eluted using 30 mM glutathione in 0.15 M TEGN (20 mM Tris pH 7.9, 0.1 mM EDTA, 10% glycerol, and 0.02% NP-40). Mediator is found in both the P0.5M and P1M fractions, and in our hands, no significant difference in Mediator composition is seen between the two fractions.

Figure 3.1. **Assembly of Med-PIC.** A) SDS PAGE gel of purified Med-PIC factors. Lanes were rearranged for clarity. B-C) Representative negative stained (B) and cryogenic (C) electron micrograph and class averages show intact Med-PIC complexes with multiple views.


A
MW
200
150
120
100
85
70
60
50
40
30
25
20
15
10
Pol II
TFIIB
TBP
TFIIA
TFIIF
TFIIE 56
TFIIE 34
TFIIH
Mediator
B
100 nm
10 nm
C
100 nm
10 nm

3.2.2. Assembly of hMed-PIC

Human Med-PIC complexes were assembled as previously described for the PIC with the following changes to accommodate the incorporation of Mediator into the complex [66, 126]. For negative staining, three subcomplexes were assembled in parallel. First, 0.25 pmol of a super core promoter DNA template (sense: 5'-GAAGGGCGCCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT-3', anti-sense: 5'-/5BiotinTEG/ACTGGGGAATTCCATGGTCCGTAGGCACGTCTGCTCGGCTCGAGTGTTCGATCGCGACTGAGGACGAACGCGCCCCCACCCCCTTTTATAGGCGCCCTTC-3') was mixed with 1.8 pmol TFIIB, 2 pmol TBP, 1 pmol TFIIA. 0.1 pmol Pol II was mixed with 0.7 pmol TFIIF in a second tube. In a third tube, 1.5 pmol Mediator was mixed with 2.5 pmol TFIIE56, 7.6 pmol TFIIE34, and 1 pmol TFIIH. The salt concentration of each solution was adjusted to 100 mM KCl with the addition of buffers A (12 mM HEPES pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, and 0.05% NP-40) and B (12 mM HEPES pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 1 mM DTT, and 0.05% NP-40) After 30 minutes at room temperature (RT), all components were combined and incubated for an additional 30 minutes at RT before binding to T1 streptavidin beads (Fisher Scientific) at RT for 15 minutes. Assembled complexes were washed with buffer C (10 mM HEPES pH 7.9, 10 mM Tris pH 7.9, 5% glycerol, 5 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, and 0.05% NP-40) and eluted with buffer D (10 mM HEPES pH 7.6, 5% glycerol, 10 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, 0.05% NP-40, and 30 units EcoRI-HF (New England Biolabs)).

Complex assembly for cryo-EM was identical to negative staining samples. Assembled complexes were always used fresh for microscopy and never flash-frozen to maintain the structural integrity of the complex.

3.2.3. Electron Microscopy

Negative stain samples were prepared using 400 mesh copper grids (Electron Microscopy Sciences) with a thin layer of continuous carbon on a nitrocellulose support film that was glow-discharged in air for 10 seconds with 25 W of power using the PELCO easiGlow (TED PELLA). Purified Med-PIC complexes in buffer D were cross-linked with 0.05% glutaraldehyde for 10 minutes on ice and incubated for 10 minutes on a grid in a homemade humidity chamber at 4 °C. The grid was sequentially incubated on 4, 40 μL drops of 2% uranyl formate solution for 5, 10, 15, and 20 seconds and blotted dry with #1 filter paper (Whatman). Images were collected on a Jeol 1400 equipped with a Gatan 4k × 4x CCD camera at 30,000X magnification (3.71 Å/pixel), a defocus range of -1.5 to -3 μm, and 20 $e^-/Å^2$ total electron dose using Leginon (Figure 3.1B) [144].

Cryo-EM samples were prepared using Quantifoil 2/1 300 mesh copper grids (EMS). Grids were glow discharged in air for 10 seconds with 5 W of power using the PELCO easiGlow, and then a thin layer of graphene oxide was applied as described previously [145]. Eluted Med-PIC samples (~3.5 μL) were incubated with 0.05% glutaraldehyde for 10 minutes on ice in the dark. The sample was applied to a grid suspended in a Vitrobot operating at 4 °C with 100% humidity. After 5 minutes, the sample was blotted with 10 force for 4 seconds and immediately plunged into liquid ethane cooled to liquid nitrogen temperatures. A data set of 19,881 images was collected at the Pacific Northwestern Center for Cryo-EM (PNCC) (Figure 3.1C). Images were collected using semi-automated data collection in Serial EM [146] on a 300 kV Titan Krios-3 microscope (Thermo Fisher) equipped with a Gatan K3 direct detector operating in super-resolution mode at a magnification of 30,000X (0.5295 Å/pixel). Images were collected using a defocus range of -2 to -4 μm with a 45-frame exposure taken over a total of 2.1 seconds using a dose rate of 15 e^-/pixel/second for a total dose of 31.5 $e^-/Å^2$. 19,881 images total were collected (Table 3.1).

3.2.4. Image processing

For negative-stained samples, particles were picked using DogPicker, extracted, and 2D classified using iterative MSA/MRA topological alignment within the Appion data processing software [147-150]. A particle stack of at least 50,000 particles with a box size of 144 x 144 pixels was subjected to iterative, multi-reference projection-matching 3D refinement using libraries from the EMAN2 software package to generate an initial reference for cryo-EM data processing [151].

RELION 3.1 was used for all pre-processing, 3D classification, model refinement, post-processing, and local-resolution estimation jobs [136]. To pre-process the cryo-EM data, movie frames were aligned using RELION's implementation while binning by a factor of 2 (1.059 Å/pixel). After motion correction, micrographs were manually inspected, resulting in the exclusion of 3,903 micrographs from further processing. Particles were automatically picked using Gautomatch (developed by K. Zhang, MRC Laboratory of Molecular Biology, Cambridge, UK), and the local CTF of each micrograph was determined using Gctf or CTFFIND-4 [150, 152].

An initial particle stack of 885,514 particles was binned by a factor of 4 (4.236 Å/pixel), extracted, and subjected to an initial round of 3D classification using the negative stain reconstruction (low-pass filtered to 30 Å) as an initial reference (Figure 3.2). Class 5 (156,383 particles) showed sharp and clear structural features of Med-PIC, so it was selected for further processing. The selected particles were 3D auto-refined, re-centered, and re-extracted without binning (1.059 Å/pixel, box size = 450 pixels). Another round of 3D auto-refinement was performed with a soft mask applied around the whole complex, resulting in a 5.79 Å resolution reconstruction. All reported resolutions correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criterion [153]. Per-particle CTF refinement was performed by first estimating magnification anisotropy, then per-particle defocus and per-micrograph astigmatism, and finally beam tilt, followed by Bayesian particle polishing. 3D auto-refinement using the

polished particles yielded a 4.8 Å resolution map. The map was post-processed using DeepEMhancer [154] (Figure 3.3). This map is the full Med-PIC map used for deposition.

To improve the map quality of distal portions of the complex that showed significant averaging out in the Med-PIC map, we segmented the Med-PIC map into 7 bodies: MedTail, Med14Med24, MedHead, MedMiddle-CAK, cPIC, cTFIIH, and Med1. For each body, a partial soft mask was applied to the corresponding volume, and particles were signal subtracted, re-centered, and re-extracted with a suitable box size (384, 360, 384, 320, 320, 288, and 288 pixels, respectively). Next, the particles are binned by a factor of 2 (2.118 Å/pixel) and 3D auto-refined locally with an initial angular sampling interval of 3.7°. Subsequent three-class 3D classification without alignment was performed, and the class with the best features and resolution (See Figure 3.2 for particle numbers of each selected class) was selected, un-binned (1.059 Å/pixel), auto-refined, and post-processed. Local resolution of the maps was estimated with RELION 3.1.

3D variability analysis was performed on the Med-PIC, MedΔTail-PIC, and Med-CAK maps using CryoSPARC [155]. For Med-PIC, a soft mask was applied, signal subtraction was performed, and the subtracted stack was binned by 2 (2.118 Å/pixel), re-centered, and re-boxed (280 pixels) in Relion. This stack was transferred to CryoSPARC for masked non-uniform refinement, which resulted in a 4.3 Å resolution map. 3D variability analysis was performed on the aligned stack after filtering to 5 Å resolution, and the first three principal components were selected for analysis. A similar strategy was used for the remaining two maps with box sizes of 180 and 270 pixels, respectively. Both maps gave 4.3 Å resolution maps after non-uniform refinement.

UCSF Chimera and UCSF Chimera X were used for all volume segmentation, figure and movie generation, and rigid-body docking [142, 156]. In parallel with post-processing done in

RELION3.1, DeepEMhancer was applied on the refined maps to better correct local B-factors and yielded cleaner maps for model building and docking [154].

Table 3.1. **Med-PIC cryo-EM data collection, refinement, and validation statistics**

	#1 Med-PIC (EMDB: 23255, PDB: 7LBM)	#2 cPIC (EMDB: 23256)	#3 cTFIIH (EMDB: 23257)	#4 Med Head (EMDB: 23258)	#5 Med Middle-CAK (EMDB: 23259)	#6 Med14C (EMDB: 23260)	#7 Med Tail (EMDB: 23261)	#8 Med1 (EMDB: 23262)
Data collection and processing								
Microscope	Titan Krios-3							
Voltage (kV)	300							
Camera	Gatan K3							
Magnification	30k							
Pixel size at detector (Å/pixel)	1.059							
Total electron exposure (e–/Å^2)	~31							
Exposure rate (e$^-$/pixel/sec)	15							
Number of frames collected during exposure	45							
Defocus range (μm)	-2.0 to -4.0							
Automation software	SerialEM							
Energy filter slit width	N/A							
Micrographs collected (no.)	19,881							
Micrographs used (no.)	15,978							
Total extracted particles (no.)	885,514							
Reconstruction	Med-PIC	cPIC	cTFIIH	Med Head	Med Middle-CAK	Med14C	Med Tail	Med1
Refined particles (no.)	156,383	54,801	44,471	47,138	43,779	35,447	79,952	108,383
Final particles (no.)	156,383	54,801	44,471	47,138	43,779	35,447	79,952	108,383

Point-group or helical symmetry parameters	C1	C1	C1	C1	C1	C1	C1	C1
Resolution (global, Å)								
FSC 0.5 (unmasked/masked)	9.82/7.4	7.1/3.82	9.22/8.25	7.92/4.45	9.28/7.82	7.8/4.5	7.7/4.12	8.28/7.25
FSC 0.143 (unmasked/masked)	7.68/4.8	4.15/3.4	7.9/7.1	6.33/4.0	7.9/6.5	5.25/4.0	4.6/3.6	6.98/5.8
Resolution range (local, Å)	3.4 to >10	3 to 7	6 to >10	3.6 to 8	6 to >10	3.6 to 8	3.3 to 7.5	5 to 9
Map sharpening B factor (Å^2)	-126	-76	-248	-92	-183	-92	-81	-221
Map sharpening methods	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN
Model composition								
Protein	15,877							
Ligands	19							
RNA/DNA	128							
Model Refinement								
Refinement package	Phenix							
-real or reciprocal space	Real							
Model-Map scores								
-CC	0.46							
-Average FSC								
B factors (Å^2)								
Protein residues	80.23							
Ligands	85.15							
RNA/DNA	71.22							
R.m.s. deviations from ideal values								
Bond lengths (Å)	0.009							
Bond angles (°)	1.236							

Validation	
MolProbity score	2.57
CaBLAM outliers	5.7
Clashscore	30.04
Poor rotamers (%)	0.71
C-beta deviations	0.03
Ramachandran plot	
Favored (%)	87.88
Outliers (%)	0.65

Figure 3.2. **Med-PIC cryo-electron microscopy processing pipeline.** An initial 3D classification of all particles resulted in a single class that could be refined to an overall resolution of 4.8 Å. Focused local refinements on subcomplexes were performed for cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, Med1, and MedTail by binning by a factor of 2, re-centering, and signal subtracting away the rest of the complex. 3D classification without alignment was performed to select a subset of particles that were unbinned and refined to 3.4, 7.1, 4.0, 6.5, 4.0, 5.8, and 3.6 Å, respectively.

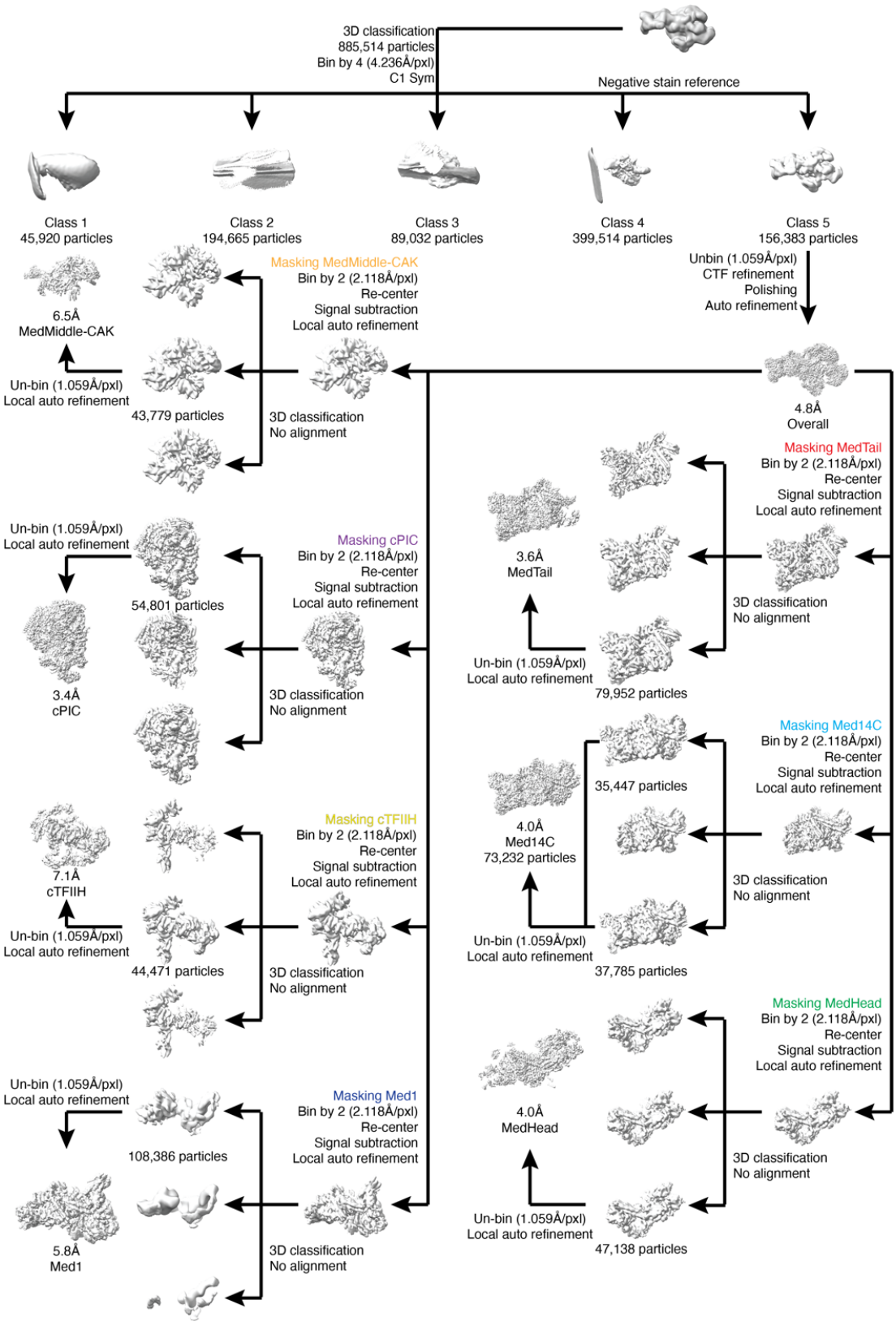
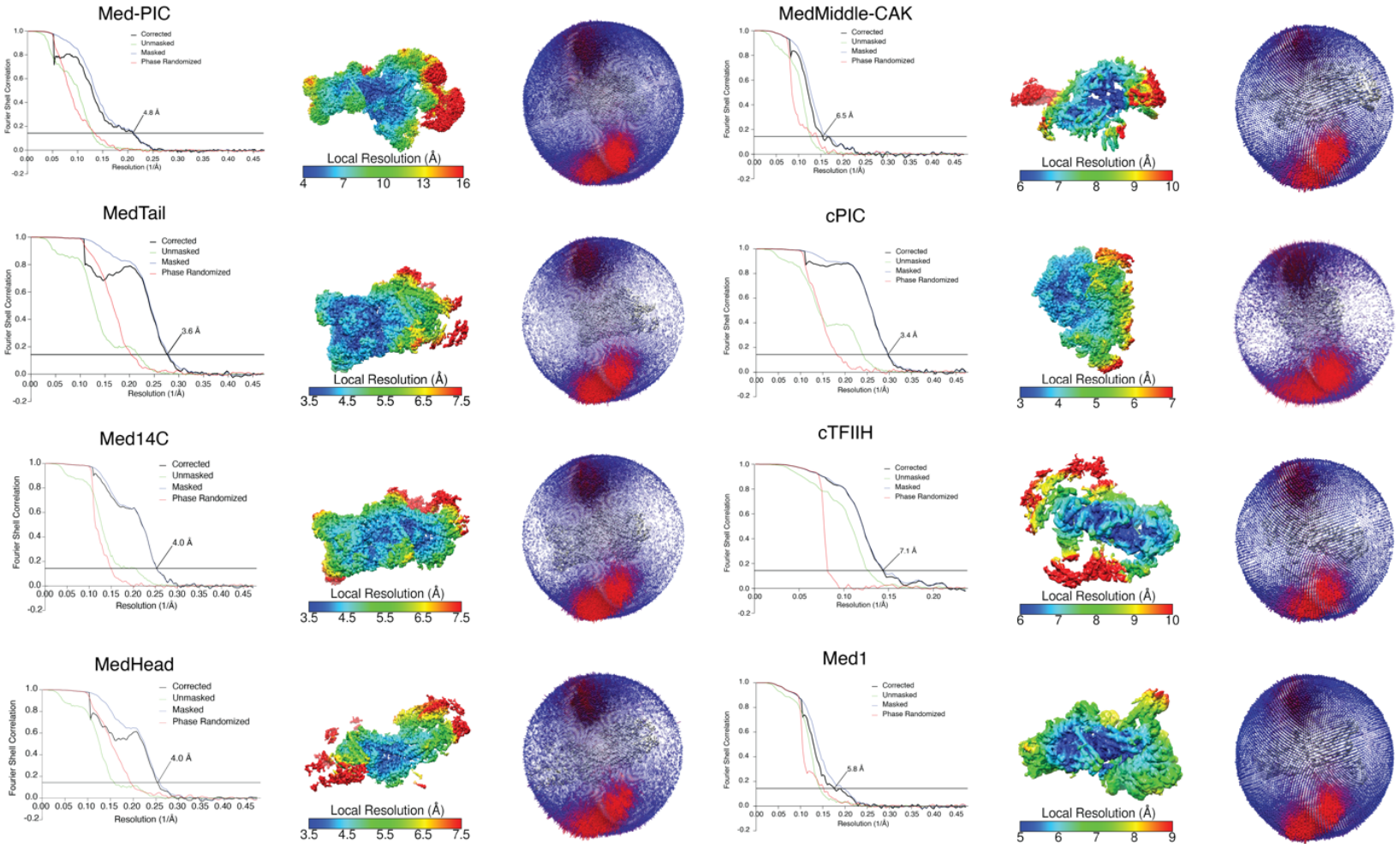

3D classification
885,514 particles
Bin by 4 (4.236Å/pxl)
C1 Sym
Negative stain reference
Class 1
45,920 particles
Class 2
194,665 particles
Class 3
89,032 particles
Class 4
399,514 particles
Class 5
156,383 particles
Unbin (1.059Å/pxl)
CTF refinement
Polishing
Auto refinement
4.8Å
Overall
Masking MedMiddle-CAK
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
43,779 particles
Un-bin (1.059Å/pxl)
Local auto refinement
6.5Å
MedMiddle-CAK
Masking MedTail
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
79,952 particles
Un-bin (1.059Å/pxl)
Local auto refinement
3.6Å
MedTail
Masking cPIC
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
54,801 particles
Un-bin (1.059Å/pxl)
Local auto refinement
3.4Å
cPIC
Masking Med14C
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
35,447 particles
37,785 particles
Un-bin (1.059Å/pxl)
Local auto refinement
4.0Å
Med14C
73,232 particles
Masking cTFIIH
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
44,471 particles
Un-bin (1.059Å/pxl)
Local auto refinement
7.1Å
cTFIIH
Masking MedHead
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
47,138 particles
Un-bin (1.059Å/pxl)
Local auto refinement
4.0Å
MedHead
Masking Med1
Bin by 2 (2.118Å/pxl)
Re-center
Signal subtraction
Local auto refinement
3D classification
No alignment
108,386 particles
Un-bin (1.059Å/pxl)
Local auto refinement
5.8Å
Med1

Figure 3.3. **Cryo-electron microscopy map quality.** Map resolution, local resolution maps, and angular distributions for the full Med-PIC map and cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1 focused refinements. The overall reconstruction shows significant variation in local resolution that improves in each subcomplex following focused refinement.


Med-PIC
Corrected
Unmasked
Masked
Phase Randomized
Fourier Shell Correlation
Resolution (1/Å)
4.8 Å
Local Resolution (Å)
4 7 10 13 16
MedMiddle-CAK
6.5 Å
6 7 8 9 10
MedTail
3.6 Å
3.5 4.5 5.5 6.5 7.5
cPIC
3.4 Å
3 4 5 6 7
Med14C
4.0 Å
cTFIIH
7.1 Å
MedHead
Med1
5.8 Å
5 6 7 8 9

3.2.5. Model building

cPIC:

The human cPIC bound to a closed DNA template (PDB:5IYA) [66] was fit as a rigid body into the cPIC density map as an initial model using UCSF Chimera (Table 3.2) [156]. Manual adjustments were made in Coot thanks to the high resolution of the cPIC map [140]. Modifications made to TFIIE were guided by the human TFIIE crystal structure (PDB: 5GPY) [157]. The cPIC model was real space refined in Phenix to the cPIC map [141].

Table 3.2. **Med-PIC model building starting models and model confidence**

Domain	Chain ID	Map	Prior knowledge	Level of confidence
cPIC	A-T	cPIC	human cPIC PDB:5IYA human TFIIE PDB: 5GPY	Atomic level
$_{MED}$CTD	A	MedHead	PDB:4GWQ	Backbone trace
$_{CDK}$CTD	A	MedMiddle-CAK	PDB: 1QMZ	Backbone trace
TFIIH-CAK	d-f	MedMiddle-CAK	PDB: 6XBZ	Backbone trace
cTFIIH	W-c	MedMiddle-CAK	PDB: 6NMI	Backbone trace
Med1		Med1	-	-
Med4	s	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med6	g	MedHead	PDB:5U0S	Atomic level
Med7	t	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med8	h	MedHead	PDB:5U0S	Atomic level
Med9	u	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med10	v	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med11	i	MedHead	PDB:5U0S	Atomic level
Med14	r	MedMiddle-CAK	PDB:5OQM	Backbone trace
		MedHead	PDB:5U0S	Atomic level
		Med14C	-	Atomic level
Med15	z	MedTail	PDB:2EBK	Atomic level
Med16	0	MedTail	PDB:2MZH	Atomic level
Med17	j	MedHead	PDB:5U0S	Atomic level
Med18	k	MedHead	PDB:2HZM	Atomic level
Med19	w	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med20	l	MedHead	PDB:2HZM	Atomic level
Med21	x	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med22	m	MedHead	PDB:5U0S	Atomic level
Med23	1	MedTail	PDB:6H02	Atomic level
Med24	2	MedTail	-	Atomic level
Med25	3	MedTail	PDB: 2KY6	Atomic level
Med26		-	-	-
Med27	n	Med14C	-	Atomic level
Med28	o	Med14C	-	Atomic level
Med29	p	Med14C	-	Atomic level
Med30	q	Med14C	-	Atomic level
Med31	y	MedMiddle-CAK	PDB:5OQM	Backbone trace

MedHead:

S. pombe subunit structures (PDB:5U0S) [108] of Med6, 8, 17, and 22 or the *S. cerevisiae* (PDB:5OQM) [67] structure of Med11 were used as initial models for building the human structures using the MedHead map in Coot [140] (Figure 3.4, Table 3.2). Map quality was sufficient to see side chains of bulky residues what was crucial for determining the register of the sequences. Final models were built by threading the human sequences onto the yeast structures and making any necessary adjustments guided by sequence alignment and secondary structure prediction from the MPI Bioinformatics Toolkit [158] and Jpred4 [159], respectively. The Med14C map was used for building portions of Med17C, Med27, Med28, Med29, and Med30, which were better resolved in that map than MedHead.

Homology models of Med18 and Med20 were built using the MPI Bioinformatics Toolkit and Modeller [158, 160]. These homology models were aligned to the structure of the Med18-Med20 dimer structure (PDB:2HZM) [161], flexible fitted into the post-processed map of MedHead using ISOLDE [162] in UCSF Chimera X [142] and manually adjusted in Coot. Med28 and Med30 were built using secondary structure prediction and their known closer association with MedHead than Med27 and Med29 (Figure 3.5). They could be correctly assigned in the density by initially noting that Med30 has a much longer flexible linker between helices than Med28. Med27 was built by identifying the location of the C-terminus through homology to the *S. pombe* structure. The N-terminus and Med29 were built by identifying the remaining helical density that closely matched secondary structure prediction and identifying the register based on clear bulky side chain density. Med27 was also validated due to the proximity of its N- and C-terminal ends. MedHead subunits were combined, and real space refined in Phenix to the MedHead map.

Figure 3.4. **Comparison of conserved MedHead subunits.** Each MedHead subunit is shown with the corresponding homology model from either *S. pombe* (PDB:5U0S) or *S. cerevisiae* (PDB:5OQM) and the sequence alignment used to build the human model. Colored bars above the sequence alignment show portions for which models were built, excluding any small missing loops. Sequence alignments only include those portions of each subunit for which sequence alignment was successful. Secondary structure prediction is shown for the C-terminal extension of Med17 not found in yeast using PsiPred. Models colored as in Figure 3.8.

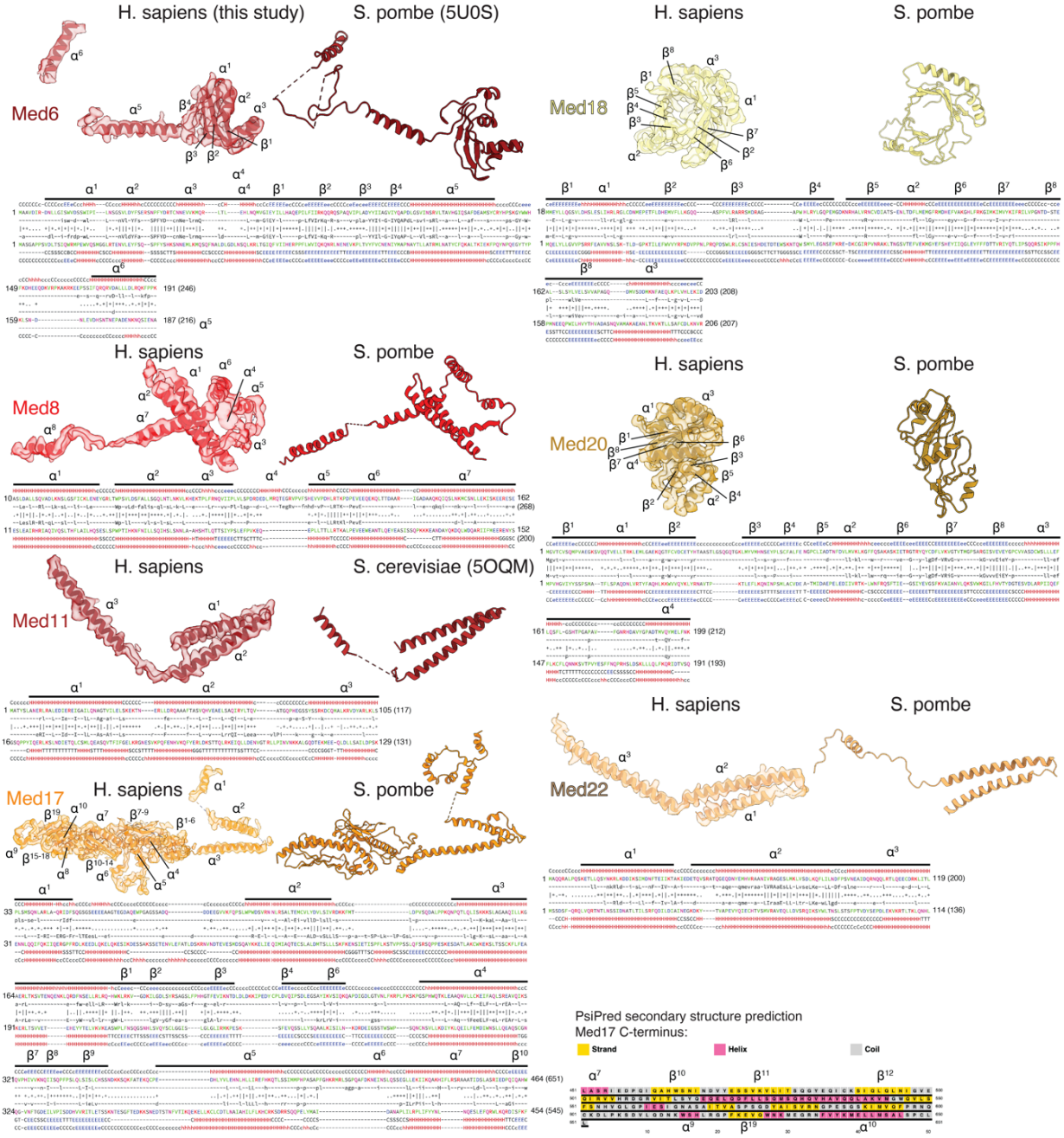
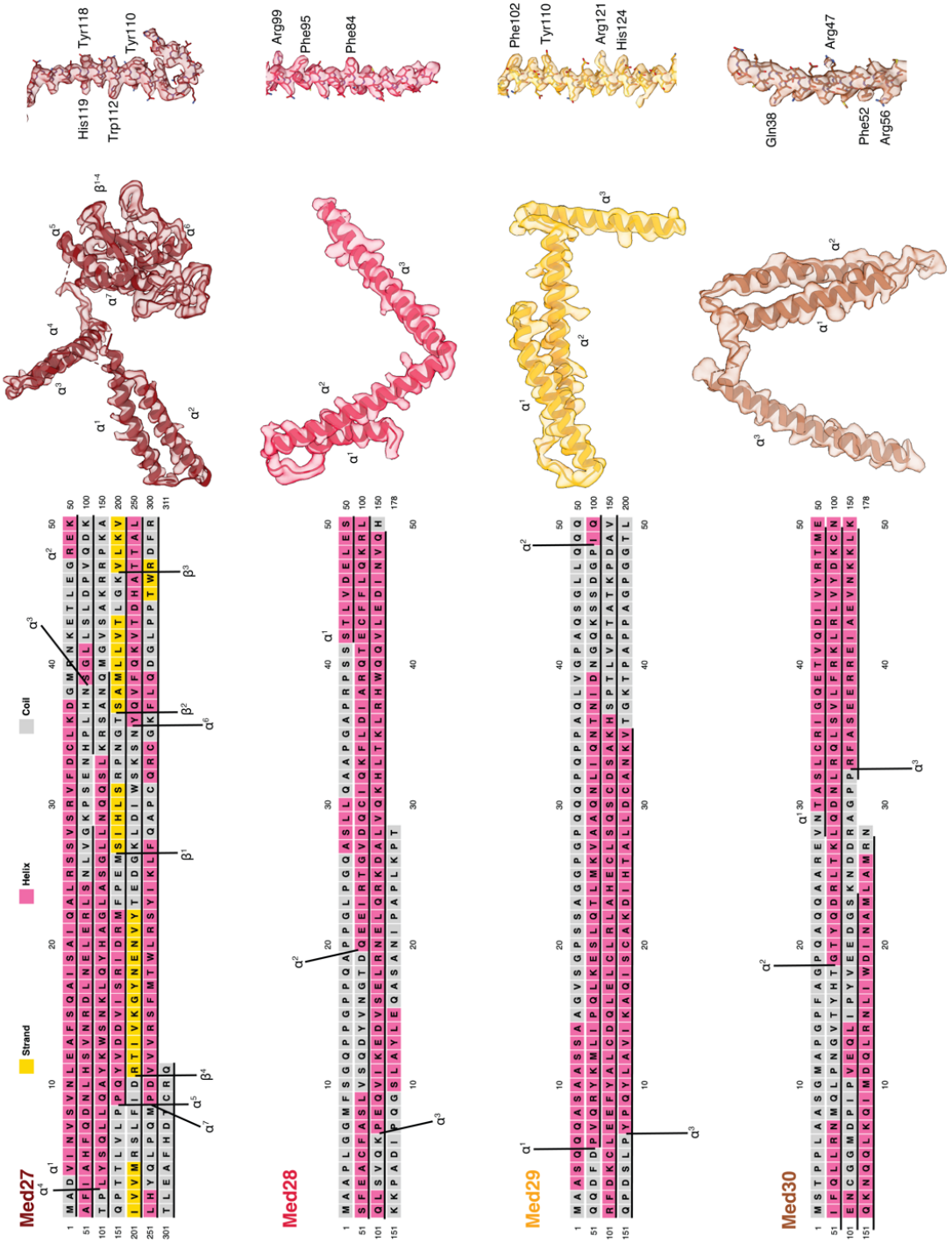

H. sapiens (this study)
S. pombe (5U0S)
Med6
H. sapiens
S. pombe
Med18
H. sapiens
S. pombe
Med8
H. sapiens
S. pombe
Med20
H. sapiens
S. cerevisiae (5OQM)
Med11
H. sapiens
S. pombe
Med17
H. sapiens
S. pombe
Med22
PsiPred secondary structure prediction
Med17 C-terminus:
Strand
Helix
Coil

Figure 3.5. **Models of MedHead subunits Med27, Med28, Med29, and Med30.** Secondary structure prediction using PsiPred closely agrees with secondary structure visible in the density maps and allowed the building of atomic models for each subunit. Secondary structure elements are labeled on both the protein sequence and the models. Representative model-to-map fits (far right) show clear density for bulky side chains that enabled atomic model building. Models colored as in Figure 3.8.


Med27
Med28
Med29
Med30
Strand
Helix
Coil
His119
Tyr118
Trp112
Tyr110
Arg99
Phe95
Phe84
Phe102
Tyr110
Arg121
His124
Gln38
Arg47
Phe52
Arg56

Med14C:

Yeast Med14 (PDB:5OQM) [67] was fitted into the Med14C density as rigid body and used to guide building the RM1 and RM2 domains guided by sequence alignment and secondary structure prediction. The C-terminus was built guided by secondary structure prediction and the high quality of the density in this area. Residues 968 to 1167, which are not predicted to form common secondary structure elements, were missing in the density, but we were able to build much of the final RM domain, which displays the typical one helix-four strand-two helix-four strand fold. The quality of the map at the very C-terminus was not sufficient to build loops between secondary structure elements or identify the correct register of the final beta-sheets. Med14C was real space refined in Phenix against the Med14C map.

MedTail:

Human Med23 (PDB:6H02) [163] was fit as a rigid body into the map of MedTail, and manual adjustments were made in Coot (Figure 3.6). Med16 was built by first locating the seven-stranded WD-40 domain in the map of MedTail. A homology model for this domain was built using PDB:5MZH [164]. Manual alignment of this model into the density was performed in UCSF Chimera by noting the connectivity of the domain to the C-terminus of the protein. The model was then manually improved in Coot. The C-terminus was built by following the density from the C-terminus of the WD-40 domain. Clear helices were visible for the rest of the density and showed clear side-chain density for bulky residues, allowing manual building for the rest of the protein. Med24 is predicted to be almost entirely helical and was localized above Med23 in the density. This was the only remaining largely helical density where a subunit of this size could be located in MedTail. The register was established by identifying the longest predicted helices using secondary structure prediction, locating possible densities, and identifying bulky side chains.

The von Willebrand factor type A (vWA) domain of Med25 was built by first building a homology model (PDB:3V4V_B) [165] using Modeller in the MPI Bioinformatics Toolkit and rigid body docking it into any unmodeled density remaining in MedTail. Manual adjustments were made to the final model in Coot.

The RWD of Med15 (residues 677-786) was built by first building a homology model (PDB:2EBK) using Modeller in the MPI Bioinformatics Toolkit and then rigid body fitting it into any unmodeled density remaining in MedTail using UCSF Chimera. Manual adjustments were then made using Coot. Residues 617-652 were built by using secondary structure prediction and looking in the direction of the N-terminus of the RWD domain. The two helices, one with a large kink in it, showed clear side-chain density that matched the predicted sequence of Med15. MedTail subunits were combined and real space refined in Phenix to the MedTail map.

Figure 3.6. **Models of MedTail.** Secondary structure prediction and model-to-map fit for each subunit of MedTail. Underlined sequences indicate portions of each subunit for which models were built. Models colored as in Figure 3.8.

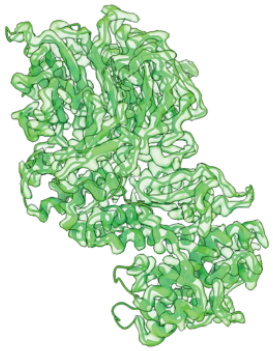
Med15

Strand Helix Coil

```
1   MDVSGQETDWRSTAFRQKLVSQIEDAMRKAGVAHSKSSKDMESHVFLKAK  50
51  TRDEYLSLVARLIIHFRDIHNKKSQASVSDPMNALQSLTGGPAAGAAGIG  100
101 MPPRGPGQSLGGMGSLGAMGQPMSLSGQPPPGTSGMAPHSMAVVSTATPQ  150
151 TQLQLQQVALQQQQQQQQFQQQQQQAALQQQQQQQQQQQQFQAQQSAMQQQF  200
201 QAVVQQQQQLQQQQQQQQHLIKLHHQNQQQIQQQQQQLQRIAQLQLQQQQ  250
251 QQQQQQQQQQQQQALQAQPPIQQPPMQQPQPPPPSQALPQQLQQMHHTQHHQ  300
301 PPPQPQQPPVAQNQPSQLPPQSQTQPLVSQAQALPGQMLYTQPPLKFVRA  350
351 PMVVQQPPVQPQVQQQQTAVQTAQAAQMVAPGVQMITEALAQGGMHIRAR  400
401 FPPTTAVSAIPSSSIPLGRQPMAQVSQSSLPMLSSPSPGQQVQTPQSMPP  450
451 PPQPSPQPGQPSSQPNSNVSSGPAPSPSSFLPSPSPQPSQSPVTARTPQN  500
501 FSVPSPGPLNTPVNPSSVMSPAGSSQAEEQQYLDKLKQLSKYIEPLRRMI  550
551 NKIDKNEDRKKDLSKMKSLLDILTDPSKRCPLKTLQKCEIALEKLKNDMA  600
601 VPTPPPPPVPPTKQQYLCQPLLDAVLANIRSPVFNHSLYRTFVPAMTAIH  650
651 GPPITAPVVCTRKRRLEDDERQSIPSVLQGEVARLDPKFLVNLDPSHCSN  700
701 NGTVHLICKLDDKDLPSVPPLELSVPADYPAQSPLWIDRQWQYDANPFLQ  750
751 SVHRCMTSRLLQLPDKHSVTALLNTWAQSVHQACLSAA              788
```

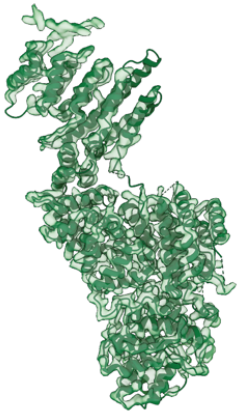
Med16

```
1   MCDLRRPAAGGMMDLAYVCEWEKWSKSTHCPSVPLACAWSCRNLIAFTMD  50
51  LRSDDQDLTRMIHILDTEHPWDLHSIPSEHHEAITCLEWDQSGSRLLSAD  100
101 ADGQIKCWSMADHLANSWESSVGSLVEGDPIVALSWLHNGVKLALHVEKS  150
151 GASSFGEKFSRVKFSPSLTLFGGKPMEGWIAVTVSGLVTVSLLKPSGQVL  200
201 TSTESLCRLRGRVALADIAFTGGGNIVVATADGSSASPVQFYKVCVSVVS  250
251 EKCRIDTEILPSLFMRCTTDLNRKDKFPAITHLKFLARDMSEQVLLCASS  300
301 QTSSIVECWSLRKEGLPVNNIFQQISPVVGDKQPTILKWRILSATNDLDR  350
351 VSAVALPKLPISLTNTDLKVASDTQFYPGLGLALAFHDGSVHIVHRLSLQ  400
401 TMAVFYSSAAPRPVDEPAMKRPRTAGPAVHLKAMQLSWTSLALVGIDSHG  450
451 KLSVLRLSPSMGHPLEVGLALRHLLFLLEYCMVTGYDWWDILLHVQPSMV  500
501 QSLVEKLHEEYTRQTAALQQVLSTRILAMKASLCKLSPCTVTRVCDYHTK  550
551 LFLIAISSTLKSLLRPHFLNTPDKSPGDRLTEICTKITDVDIDKVMINLK  600
601 TEEFVLDMNTLQALQQLLQWVGDFVLYLLASLPNQGSLLRPGHSFLRDGT  650
651 SLGMLRELMVVIRIWGLLKPSCLPVYTATSDTQDSMSLLFRLLTKLWICC  700
701 RDEGPASEPDEALVDECCLLPSQLLIPSLDWLPASDGLVSRLQPKQPLRL  750
751 QFGRAPTLPGSAATLQLDGLARAPGQPKIDHLRRLHLGACPTEECKACTR  800
801 CGCVTMLKSPNRTTAVKQWEQRWIKNCLAVEGRGPDACVTSRASEEAPAF  850
851 VQLGPQSTHHSPRTPRSLDHLHPEDRP                          877
```



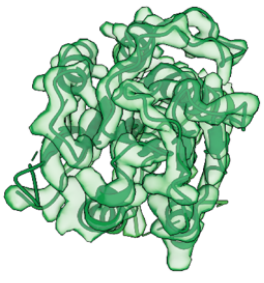
Med24

```
1   MKVVNLKQAILQAWKERWSDYQWAINMKKFFPKGATWDILNLADALLEQA  50
51  MIGPSPNPLILSYLKYAISSQMVSYSSVLTAISKFDDFSRDLCVQALLDI  100
101 MDMFCDRLSCHGKAEECIGLCRALLSALHWLLRCTAASAERLREGLEAGT  150
151 PAAGEKQLAMCLQRLEKTLSSTKNRALLHIAKLEEASSWTAIEHSLLKLG  200
201 EILANLSNPQLRSQAEQCGTLIRSIPTMLSVHAEQMHKTGFPTVHAVILL  250
251 EGTMNLTGETQSLVEQLTMVKRMQHIPTPLFVLEIWKACFVGLIESPEGT  300
301 EELKWTAFTFLKIPQVLVKLKKYSHGDKDFTEDVNCAFEFLLKLTPLLDK  350
351 ADQRCNCDCTNFLLQECGKQGLLSEASVNNLMAKRKADREHAPQQKSGEN  400
401 ANIQPNIQLILRAEPTVTNILKTMDADHSKSPEGLLGVLGHMLSGKSLDL  450
451 LLAAAAATGKLKSFARKFINLNEFTTYGSEESTKPASVRALLFDISFLML  500
501 CHVAQTYGSEVILSESRTGAEVPFFETWMQTCMPEEGKILNPDHPCFRPD  550
551 STKVESLVALLNNSSEMKLVQMKWHEACLSISAAILEILNAWENGVLAFE  600
601 SIQKITDNIKGKVCSLAVCAVAWLVAHVRMLGLDEREKSLQMIRQLAGPL  650
651 FSENTLQFYNERVVIMNSILERMCADVLQQTATQIKFPSTGVDTMPYWNL  700
701 LPPKRPIKEVLTDIFAKVLEKGWVDSRSIHIFDTLLHMGGVYWFCNNLIK  750
751 ELLKETRKEHTLRAVELLYSIFCLDMQQVTLVLLGHILPGLLTDSSKWHS  800
801 LMDPPGTALAKLAVWCALSSYSSHKGQASTRQKKRHREDIEDYISLFPLD  850
851 DVQPSKLMRLLSSNEDDANILSSPTDRSMSSSLSASQLHTVNMRDPLNRV  900
901 LANLFLLISSILGSRTAGPHTQFVQWFMEECVDCLEQGGRGSVLQFMPFT  950
951 TVSELVKVSAMSSPKVVLAITDLSLPLGRQVAAKAIAAL              989
```



Med25

```
1   MVPGSEGPARAGSVVADVVFVIEGTANLGPYFEGLRKHYLLPAIEYFNGG  50
51  PPAETDFGGDYGGTQYSLVVFNTVDCAPESYVQCHAPTSSAYEFVTWLDG  100
101 IKFMGGGGESCSLIAEGLSTALQLFDDFKKMREQIGQTHRVCLLICNSPP  150
151 YLLPAVESTTYSGCTENLVQQIGERGIHFSIVSPRKLPALRLLFEKAAP  200
201 PALLEPLQPPTDVSQDPRHMVLVRGLVLPVGGGSAPGPLQSKQPVPLPPA  250
251 APSGATLSAAPQQPLPPVPPQYQVPGNLSAAQVAAQNAVEAAKNQKAGLG  300
301 PRFSPITPLQQAAPGVGPPFSQAPAPQLPPGPPGAPKPPPASQPSLVSTV  350
351 APGSGLAPTAQPGAPSMAGTVAPGGVSGPSPAQLGAPALGGQQSVSNKLL  400
401 AWSGVLEWQEKPKPASVDANTKLTRSLPCQVYVNHGENLKTEQWPQKLIM  450
451 QLIPQQLLTTLGPLFRNSRMVQFHFTNKDLESLKGLYRIMGNGFAGCVHF  500
501 PHTAPCEVRVLMLLYSSKKKIFMGLIPYDQSGFVNGIRQVITNHKQVQQQ  550
551 KLEQQQRGMGGQQAPPGLGPILEDQARPSQNLLQLRPPQPQPQGTVGASG  600
601 ATGQPQPQGTAQPPPGAPQGPPGAASGPPPPGPILRPQNPGANPQLRSLL  650
651 LNPPPPQTGVPPPQASLHHLQPPGAPALLPPPHQGLGQPQLGPPLLHPPP  700
701 AQSWPAQLPPRAPLPGQMLLSGGPRGPVPQPGLQPSVMEDDILMDLI     747
```

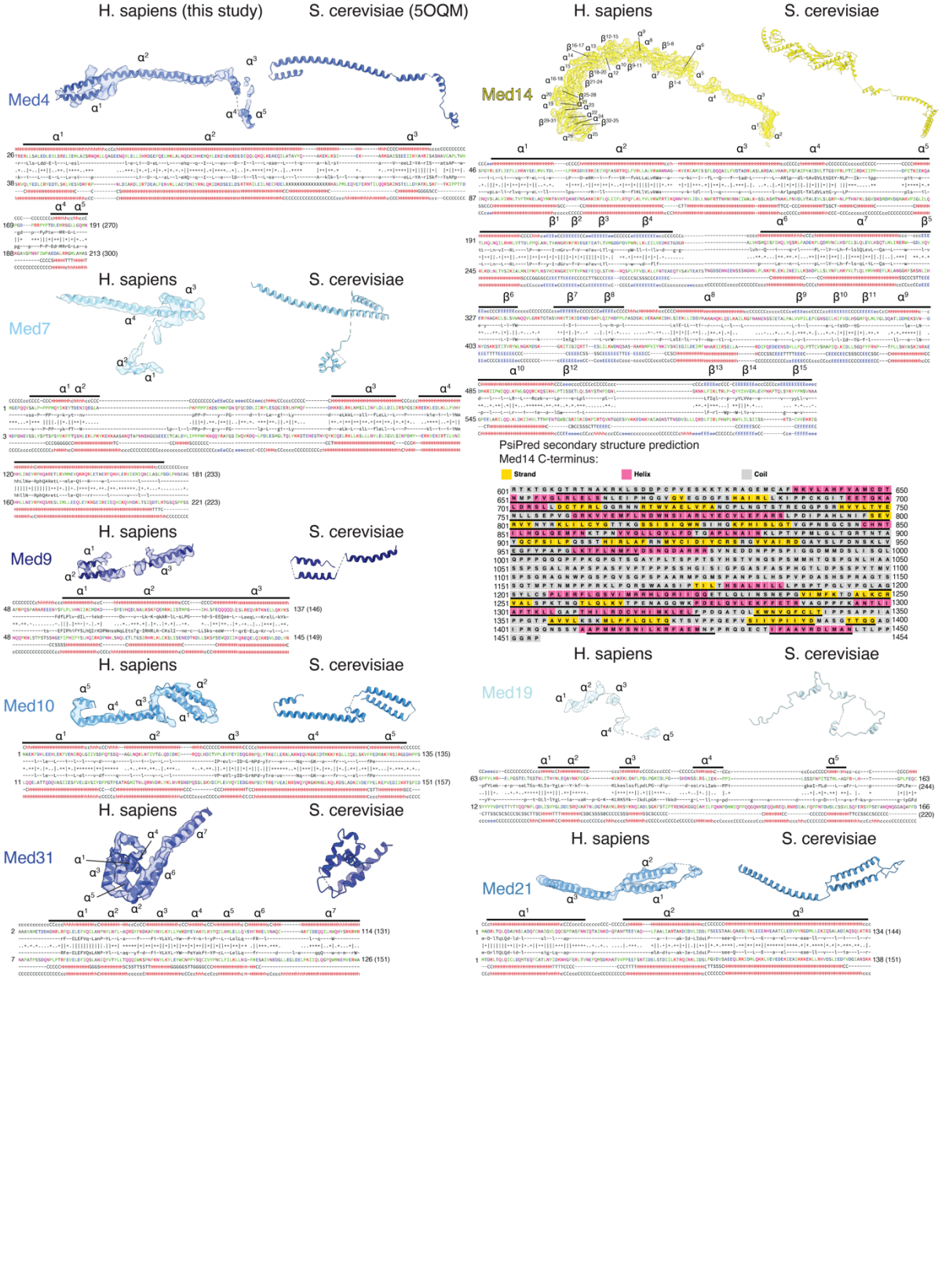


MedMiddle:

Homology models for Med4, 7, 9, 10, 14 (1-195), 19, 21, and 31 were created using sequence alignment and secondary structure prediction to their *S. cerevisiae* counterparts (PDB:5OQM) [67] in Coot (Figure 3.7, Table 3.2). These homology models were flexible fitted into the MedMiddle-CAK density using Namdinator [166]. Manual inspection of the results, including building an additional C-terminal helix in Med31, N-terminal helix of Med17, and C-terminal helix of Med6, was done in Coot.

The $_{MED}$CTD structure was created by first aligning the yeast MedHead-CTD structure (PDB:4GWQ) [97] to our human structure using Med8. The peptide was used as an initial model to rigid body fit into the MedHead density. Clear density for the sidechains of two Y^1 residues was visible in the MedHead density, and the remaining model was built using Coot.

Figure 3.7. **Comparison of MedMiddle subunit models.** Each MedMiddle subunit is shown with the corresponding homology model from *S. cerevisiae* (PDB:5OQM) and the sequence alignment used to build the human model. Colored bars above the sequence alignment show portions for which models were built, excluding any small missing loops. Sequence alignments only include those portions of each subunit for which sequence alignment was successful. Secondary structure prediction is shown for the C-terminal extension of Med14 not found in yeast using PsiPred. Models colored as in Figure 3.8.


H. sapiens (this study)
S. cerevisiae (5OQM)
H. sapiens
S. cerevisiae
Med4
Med14
Med7
Med9
Med10
Med19
Med31
Med21
PsiPred secondary structure prediction
Med14 C-terminus:
Strand
Helix
Coil

TFIIH-CAK:

The human CAK module structure (PDB: 6XBZ) [167] was fit as a rigid body into the MedMiddle-CAK density. The CDK2-cyclin A-peptide substrate structure (PDB: 1QMZ) [168] was aligned using CDK2 to align to CDK7. The substrate peptide structure didn't need any adjustment to fit into the MedMiddle-CAK density. The sequence was mutated to the consensus sequence of the Pol II CTD, maintaining the SP motif in the substrate with S^5P^6 in the CTD and truncated to match the density visible in the structure. The model was combined with the MedMiddle structure and real space refined using Phenix. The CAK and MedMiddle subunits were combined, and real space refined in Phenix to the MedMiddle-CAK map.

cTFIIH:

The human apo-TFIIH structure (PDB: 6NMI) [169] was used as an initial model for building into the TFIIH density. Because of differences in the shape of the horseshoe, individual subunits were docked into the density as rigid body. Portions of p62 and p44 for which there was no density in our structure were removed. XPB undergoes a conformational change between its position in the cTFIIH structure and its structure in the PIC. To model this, we separately rigid body docked residues 34 to 164, 165 to 296, 297 to 502, and 503 to 730 into the density and refined the connections between those domains in Coot. The aligned subunits were combined, and real space refined using Phenix to the cTFIIH map.

Med-PIC:

The cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, and MedTail maps were segmented to remove overlapping segments and fit into the Med-PIC map. The models corresponding to each map were aligned with the maps, combined, and validated using Phenix.

3.3. Results

3.3.1. Structural characterization of the human Med-PIC

The Med-PIC complex was assembled by extending our previous protocol for assembling the closed complex PIC from purified factors to accommodate the addition of Mediator (Figure 3.1A) [66]. In contrast to previous protocols where factors were added in a stepwise manner, three subcomplexes, DNA-TBP-TFIIB-TFIIA, Pol II-TFIIF, and TFIIE-TFIIH-Mediator, were first assembled and were next incubated together. Negative stain electron microscopy (EM) of assembled complexes indicated that a subset of particles contained all components of Med-PIC and that significant conformational heterogeneity exists (Figure 3.1B).

A cryo-EM data set was collected, and 2D classification in Relion-3 showed many classes representing the full complex (Figure 3.1C, Table 3.1) [170]. A subset of 156,383 particles refined to a resolution of 4.8 Å, but due to the high intrinsic flexibility of Med-PIC, distal regions including MedMiddle, MedTail, and TFIIH, are significantly averaged out in the post-processed map. Focused refinements on subcomplexes were used to improve the resolution of all portions of the density compared to the full complex (Figure 3.2, 3.3). These regions were chosen because either the subcomplex behaves like a rigid body within the full complex, as is the case for the core PIC (cPIC), cTFIIH, MedHead, MedTail, and MedMiddle-CAK, or to center a region within the box to improve its resolution, as in the case of Med1 and Med14C. These refinements improved the resolution of the vast majority of MedTail, Med14, MedHead, and Pol II to 3.5 Å or better (Figure 3.3) and that of flexible regions, including Med1, MedMiddle-CAK, and cTFIIH, to 5.8, 6.5, and 7.1 Å, respectively. These improvements allowed the building, refining, or flexible fitting of atomic models for nearly the entire complex (Figure 3.8A-B, Table 3.2). Overall, the structure of Med-PIC is highly similar to previous human PIC complexes and yeast Med-PIC complexes (Figure 3.9) [66, 67, 106]. The presence of Mediator does not cause significant changes in the structures

of Pol II or the GTFs, including TFIIB, TBP, TFIIA, TFIIE, and TFIIF. Med-PIC is compatible with the incorporation of TFIID as no clashes are observed upon superimposing the structure of TFIID-TFIIA-DNA (Figure 3.9) [171].

Figure 3.8. **Structure of the human Mediator-bound pre-initiation complex.** A) Composite density map for Med-PIC built from the focused refinement maps for cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1. The colors of the subunits will be repeated throughout the manuscript. B) Model of the human Mediator-bound pre-initiation complex. Gray, Pol II; Dark Gray, general transcription factors; Pink, TFIIH core; Salmon, CDK7; Violet, cyclin H; Medium Violet Red, Mat1; Cyan, DNA, Reds, MedHead; Blues, MedMiddle; Yellow, Med14; Greens, MedTail.

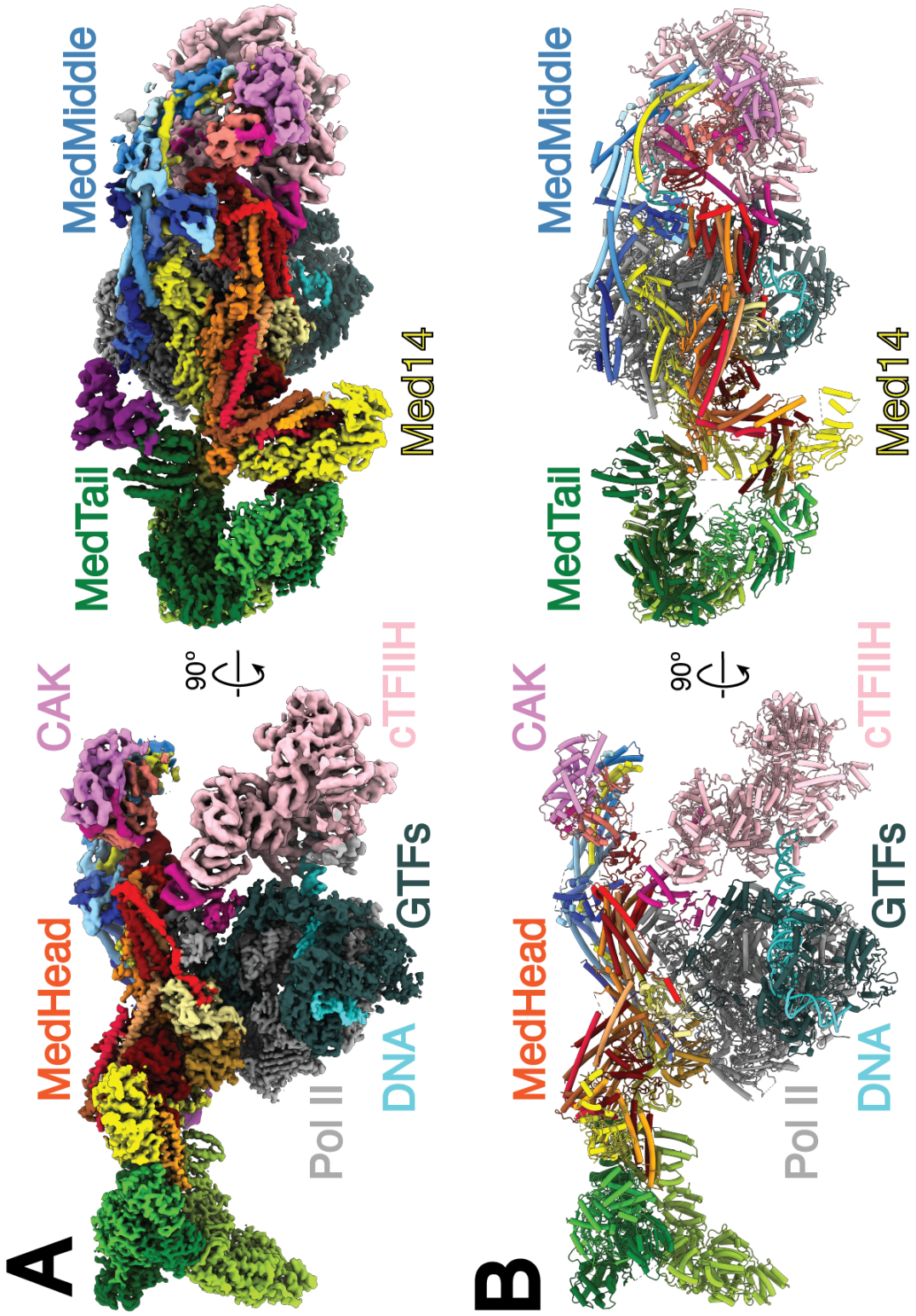
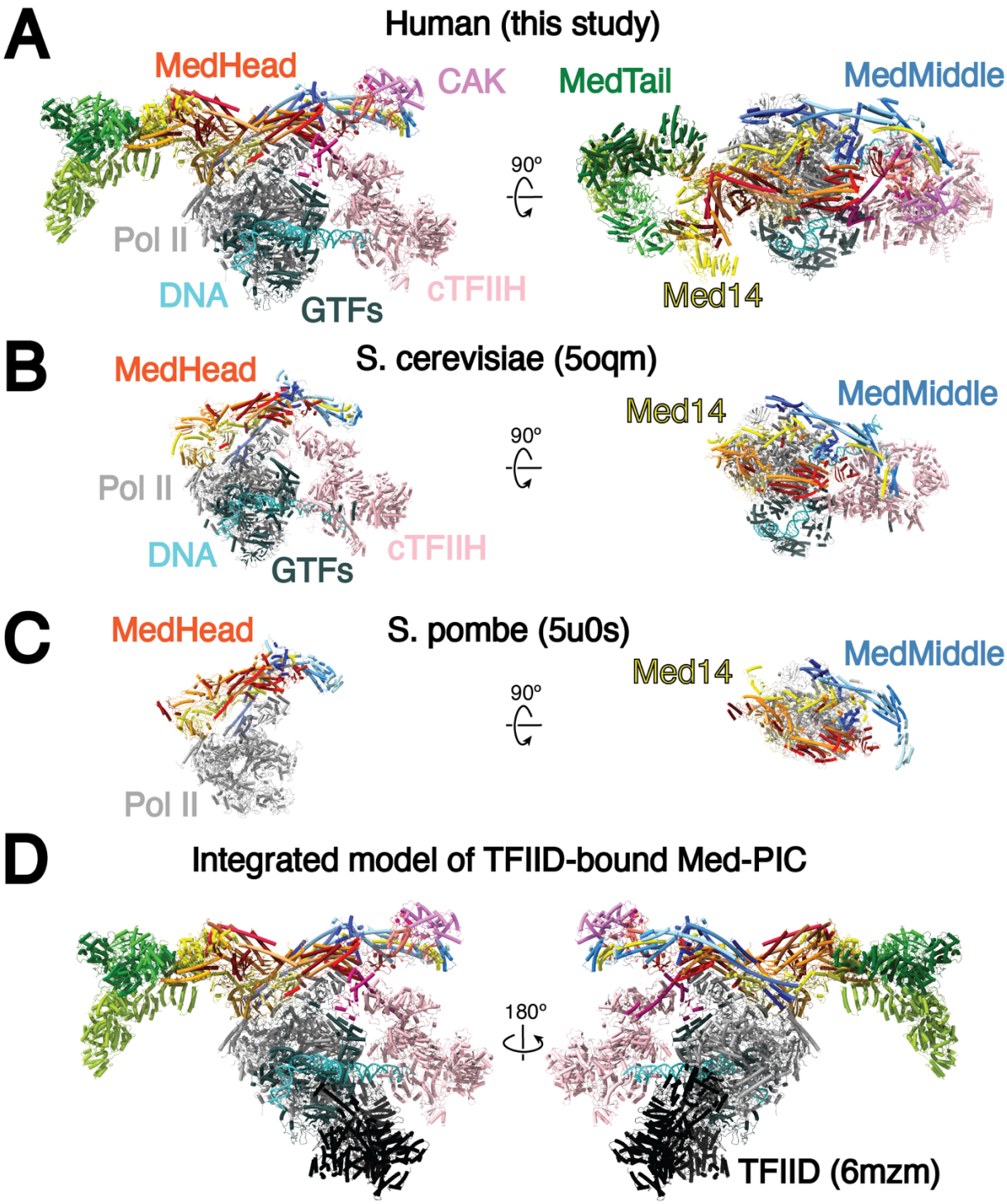

A
MedHead
CAK
MedTail
MedMiddle
90°
Pol II
DNA
GTFs
cTFIIH
Med14
B
MedHead
CAK
MedTail
MedMiddle
90°
Pol II
DNA
GTFs
cTFIIH
Med14

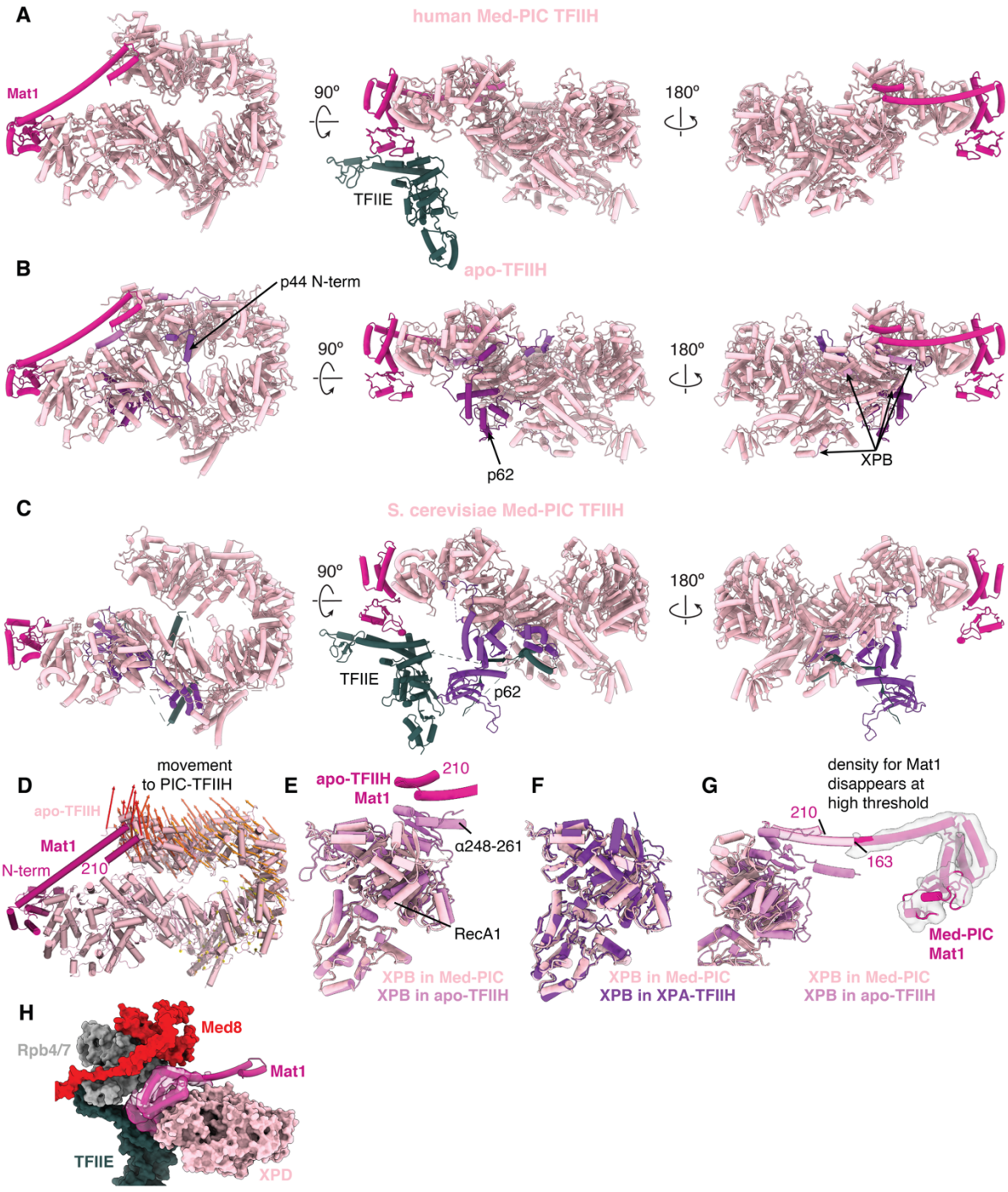
Figure 3.9. **Comparison of Med-PICs between yeast and humans and integrated model of a TFIID-bound Med-PIC.** A) Structure of the human Med-PIC as shown in Figure 3.8. B-C) Structure of *S. cerevisiae* and *S. pombe* Med-PIC complexes. The *S. cerevisiae* complex was reconstituted without MedTail, which the *S. pombe* complex is missing the GTFs. The overall architecture of the complexes does not differ dramatically between species. D) Integrated model of a TFIID-bound Med-PIC complex created by aligning the DNA from the TFIID-TFIIA-DNA complex (PDB: 6MZM) with the Med-PIC complex. No clashes are observed in this complex, suggesting no changes in Med-PIC architecture would be necessary to accommodate TFIID binding. TFIIA and the DNA from the TFIID-TFIIA-DNA complex are hidden for simplicity. Models colored as in Figure 3.8.


A
Human (this study)
MedHead
CAK
MedTail
MedMiddle
90°
Pol II
DNA
GTFs
cTFIIH
Med14
B
S. cerevisiae (5oqm)
MedHead
MedMiddle
Med14
90°
Pol II
DNA
GTFs
cTFIIH
C
MedHead
S. pombe (5u0s)
MedMiddle
Med14
90°
Pol II
D
Integrated model of TFIID-bound Med-PIC
180°
TFIID (6mzm)

3.3.2. Architecture of TFIIH

cTFIIH undergoes a conformational change from the apo structure (PDB:6NMI) [172] to its structure in Med-PIC that involves an opening of the horseshoe (Figure 3.10A-D). Structural changes within the ATP-dependent DNA helicase XPB subunit that contacts downstream DNA result in a large rotation of the damage recognition domain (DRD, residues 195-296) away from Mat1 and towards the DNA, breaking the interaction between Mat1 and helix 248-261 of XPB (Figure 3.10E). A slight shift of the RecA1 domain towards the DNA occurs upon DNA engagement. This conformation of XPB is nearly identical to that of XPB in the TFIIH-XPA-DNA structure (PDB:6RO4) formed during the initial steps of nucleotide excision repair (NER) (Figure 3.10F). We did not observe any density for helix 248-261 of XPB even though the rest of the DRD was well structured. The loss of this contact site between XPB and Mat1 results in the C-terminal half of the long Mat1 helix (residues 163-210) being visible only at a much lower map threshold (Figure 3.10G). XPD and Med8 sandwich the visible portion of the Mat1 helix, and RPB4/7 and TFIIE also contribute to the stabilization of the rest of the Mat1 N-terminus (Figure 3.10H). Thus, assembly of the CAK module into Med-PIC does not require significant structural changes in Mat1. Even if the C-terminal half of the long helix is significantly more flexible in its elongated state, it could still connect to the Mat1 C-terminus bound to the CAK module. The opening of TFIIH also results in loss of density for the N-terminus of p44 (residues 1-50), which bridges across the horseshoe (Figure 3.10A-B). This is also seen in the scMed-PIC (Figure 3.10C). No density is observed for significant portions of p62 (residues 1-106, 148-371) that are present in both the apo-TFIIH structure and the *S. cerevisiae* Med-PIC structure. In the scMed-PIC, these portions of p62 interact with TFIIE, but this interaction is seemingly not essential for complex assembly (Figure 3.10C) [67].

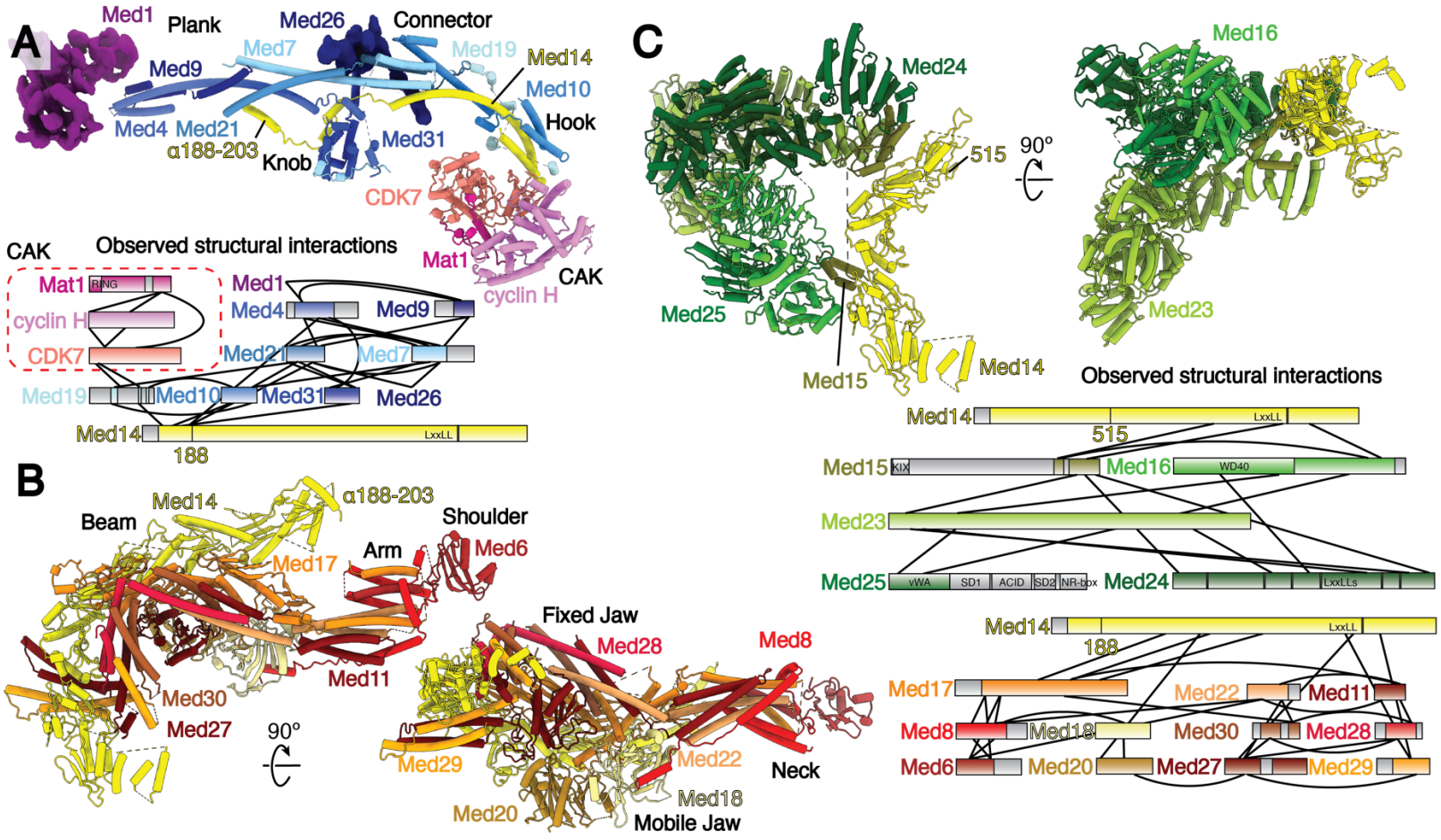
Figure 3.10. **Comparison of cTFIIH structure between human Med-PIC (A), apo-TFIIH (B), and scMed-PIC (C).** Structure of TFIIH within Med-PIC exhibits a much more open structure than apo-TFIIH. Models built in apo-TFIIH that are absent in the Med-PIC map are shown in light violet (XPB), medium violet (p44), and dark violet (p62). Mat1 and cTFIIH are colored as in Figure 3.8. TFIIE is shown in dark slate gray and interacts with the N-terminus of Med1. The structure of TFIIH within *S. cerevisiae* Med-PIC also adopts the more open shape seen in the human Med-PIC but has more extensive interactions between TFIIE and p62 that form a second stabilizing interface that is absent in the human structure. D) Structure of apo-TFIIH showing movements within the complex that accompany incorporation into Med-PIC. Length of movement is colored from yellow to red. E) Comparison of the structure of XPB in Med-PIC versus apo-TFIIH (PDB:6NMI) shows a rotation of the DRD of XPB towards the DNA, breaking contacts with Mat1. Helix 248-261, which forms contacts with Mat1 in apo-TFIIH, is not visible in the Med-PIC structure. F) Comparison of the structure of XPB in Med-PIC versus the XPA-TFIIH-DNA structure. No notable difference is seen between the two structures. G) Density for the Mat1 long helix disappears at a high threshold in the cTFIIH map due to the loss of Mat1-XPB contacts. H) The N-terminus of Mat1 is stabilized through interactions with XPD, RPB4/7, Med8, and TFIIE. When shown at a realistic threshold, density for Med1 (transparent surface) is only visible for portions stabilized by these subunits. Models are colored as in Figure 3.8.


A
human Med-PIC TFIIH
Mat1
90°
TFIIE
180°
B
apo-TFIIH
p44 N-term
90°
p62
180°
XPB
C
S. cerevisiae Med-PIC TFIIH
90°
TFIIE
p62
180°
D
movement
to PIC-TFIIH
apo-TFIIH
Mat1
N-term
210
E
apo-TFIIH
Mat1
210
α248-261
RecA1
XPB in Med-PIC
XPB in apo-TFIIH
F
XPB in Med-PIC
XPB in XPA-TFIIH
G
density for Mat1
disappears at
high threshold
210
163
Med-PIC
Mat1
XPB in Med-PIC
XPB in apo-TFIIH
H
Med8
Rpb4/7
Mat1
TFIIE
XPD

3.3.3. Structure of human Mediator

The human Mediator complex within Med-PIC is divided into three modules, held together by the central Med14 scaffold subunit (Figure 3.11). MedMiddle closely resembles the structure of its yeast counterpart [67, 106]. Homology models for the human MedMiddle subunits Med4, 7, 9, 10, 19, 21, and 31, based on the *S. cerevisiae* ortholog structures, were built using the MedMiddle-CAK map (Figure 3.7, 3.11). The N-terminal 200 residues of Med14 were modeled similarly. Additional density near the connector domain of MedMiddle could be assigned to Med26, a metazoan-specific subunit, that has been shown to localize in this part of Mediator and interact with Med4, 7, and 19 (Figure 3.11A) [96]. The C-terminus of Med26 is sufficient to interact with Mediator, strongly suggesting the C-terminus of Med26 is what can be seen, leaving the N-terminus flexibly attached. The N-terminus has been shown to interact with the super elongation complex (SEC), which is responsible for the release of paused Pol II through phosphorylation of the Pol II CTD and SPT5 by CDK9 [173].

Figure 3.11. **Models and observed structural interactions for human Mediator.** A-C) Model and observed structural interaction diagram for MedMiddle and the CAK module of TFIIH (A), MedHead (B), and MedTail (C). The N-terminus of the scaffold subunit Med14 extends the length of MedMiddle. Putative density for Med1 and Med26 are shown and colored purple and dark blue, respectively. The C-terminus of Med14 forms extensive interactions with MedHead. MedTail also interacts with the C-terminus of Med14, but on the opposite face. Portions for which models were built are shown in color; unmodeled sections are shown in gray. Known domains are shown with a light-to-dark (top-to-bottom) gradient. Everything else is shown with a dark-to-light gradient. Models colored as in Figure 3.8.


A
Med1
Plank
Med26
Connector
Med7
Med19
Med14
Med9
Med10
Hook
Med4
Med21
α188-203
Knob
Med31
CDK7
Mat1
CAK
cyclin H
CAK
Observed structural interactions
Mat1
RING
cyclin H
CDK7
Med1
Med4
Med9
Med21
Med7
Med19
Med10
Med31
Med26
Med14
LxxLL
188
B
Med14
α188-203
Beam
Shoulder
Arm
Med6
Med17
Fixed Jaw
Med28
Med8
Med11
Med30
Med27
90°
Med29
Med22
Neck
Med18
Med20
Mobile Jaw
C
Med24
515
90°
Med16
Med25
Med15
Med14
Med23
Observed structural interactions
Med14
LxxLL
515
Med15
KIX
Med16
WD40
Med23
Med25
vWA
SD1
ACID
SD2
NR-box
Med24
LxxLLs
Med14
LxxLL
188
Med17
Med22
Med11
Med8
Med18
Med30
Med28
Med6
Med20
Med27
Med29

Additional unmodeled density attributed to the N-terminus of Med1 is located between the plank domain (Med4 and Med9) and MedTail subunit Med24 (Figure 3.12A). This is consistent with the location of Med1 shown in both yeast and humans previously [96]. Density for the plank, Med1, and the N-terminus of Med24 is significantly worse than surrounding areas, indicating that this portion of Mediator moves independently of MedMiddle and MedTail. Previous structures of yeast Med-PICs show interactions between Med9 and the foot domain of Pol II (Figure 3.13A-B) [67, 108]. In *S. pombe,* Med4 and Med9 also interact with Med1, but there is no change in the overall structure compared to *S. cerevisiae,* where Med1 was not included during complex assembly. The contact between Med9 and the foot domain of Pol II is broken in the human Med-PIC. Instead, Med9 is very close to RPB8, and the interactions between Med4 and Med9 with Med1 are retained (Figure 3.13C). These differences are likely driven by the presence of the larger MedTail in the human Med-PIC, which positions Med1 further away from the plank through interactions with Med24.

Figure 3.12. **Key interfaces in Med-PIC.** A) Putative density for Med1 is located between the N-terminus of Med24 and the end of the plank domain formed by Med4 and Med9. B) Two helices of Med15, residues 617 to 649 are sandwiched between Med27 and Med29 and together form one of the two main interfaces between MedHead and MedTail. Models for Med14 and Med16 are shown as surface representations. C) Med17 stabilizes the fixed jaw on one face and interacts with the RM1 and RM2 domains of Med14 on the other face. Models are shown as either ribbon (Med17) or surface (all other subunits). D) A C-terminal extension of Med17 interacts with the RWD domain of Med15 (surface), which is located in a pocket formed by the MedTail subunits Med23 and Med24. E) The vWA domain of Med25 is located in a pocket formed by Med16 and Med23 (surface representations).

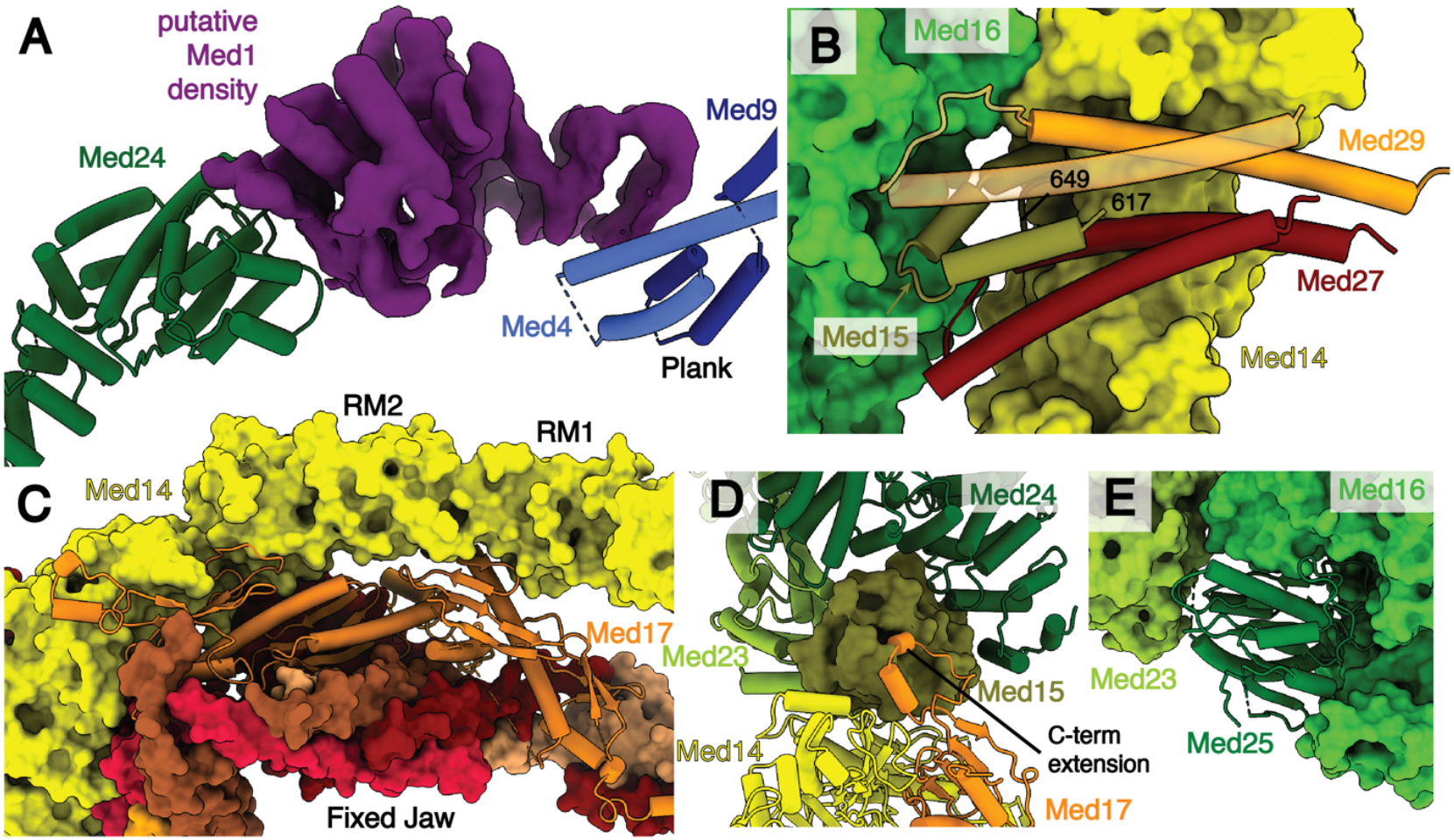
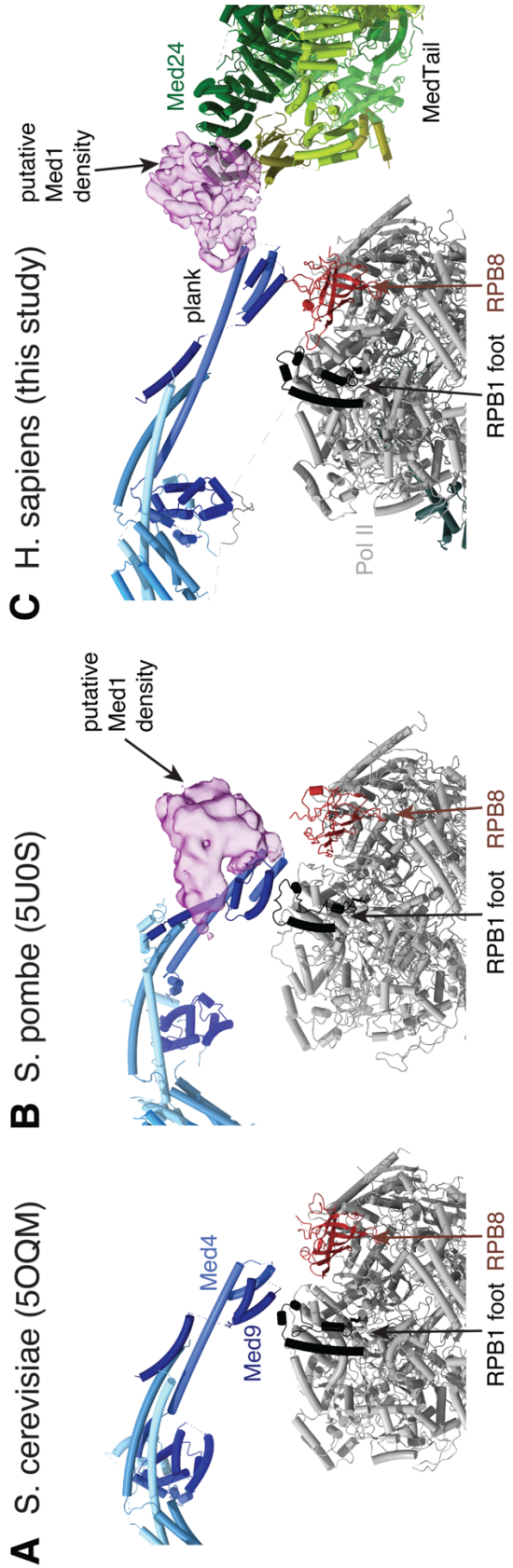

A
putative
Med1
density
Med24
Med9
Med4
Plank
B
Med16
Med29
649
617
Med27
Med15
Med14
C
Med14
RM2
RM1
Med17
Fixed Jaw
D
Med24
Med23
Med15
C-term
extension
Med14
Med17
E
Med16
Med23
Med25

Figure 3.13. **Comparison of Mediator plank domain interactions with Pol II.** A) The *S. cerevisiae* plank domain interacts with the RPB1 foot (black). The Med1 subunit is not present in the *S. cerevisiae* structure. B) The *S. pombe* plank domain interacts with both the RPB1 foot (black) and putative Med1 density, suggesting that the presence of Med1 is not sufficient to break plank-foot interactions. C) The human Mediator plank domain does not interact with the RPB1 foot (black). Instead, Med1 is stabilized by interactions with Med24 of MedTail, which pulls the plank along with it. Models are colored as in Figure 3.8.


A S. cerevisiae (5OQM)
Med4
Med9
RPB1 foot
RPB8
B S. pombe (5U0S)
putative Med1 density
RPB1 foot
RPB8
C H. sapiens (this study)
putative Med1 density
Med24
MedTail
plank
RPB1 foot
RPB8
Pol II

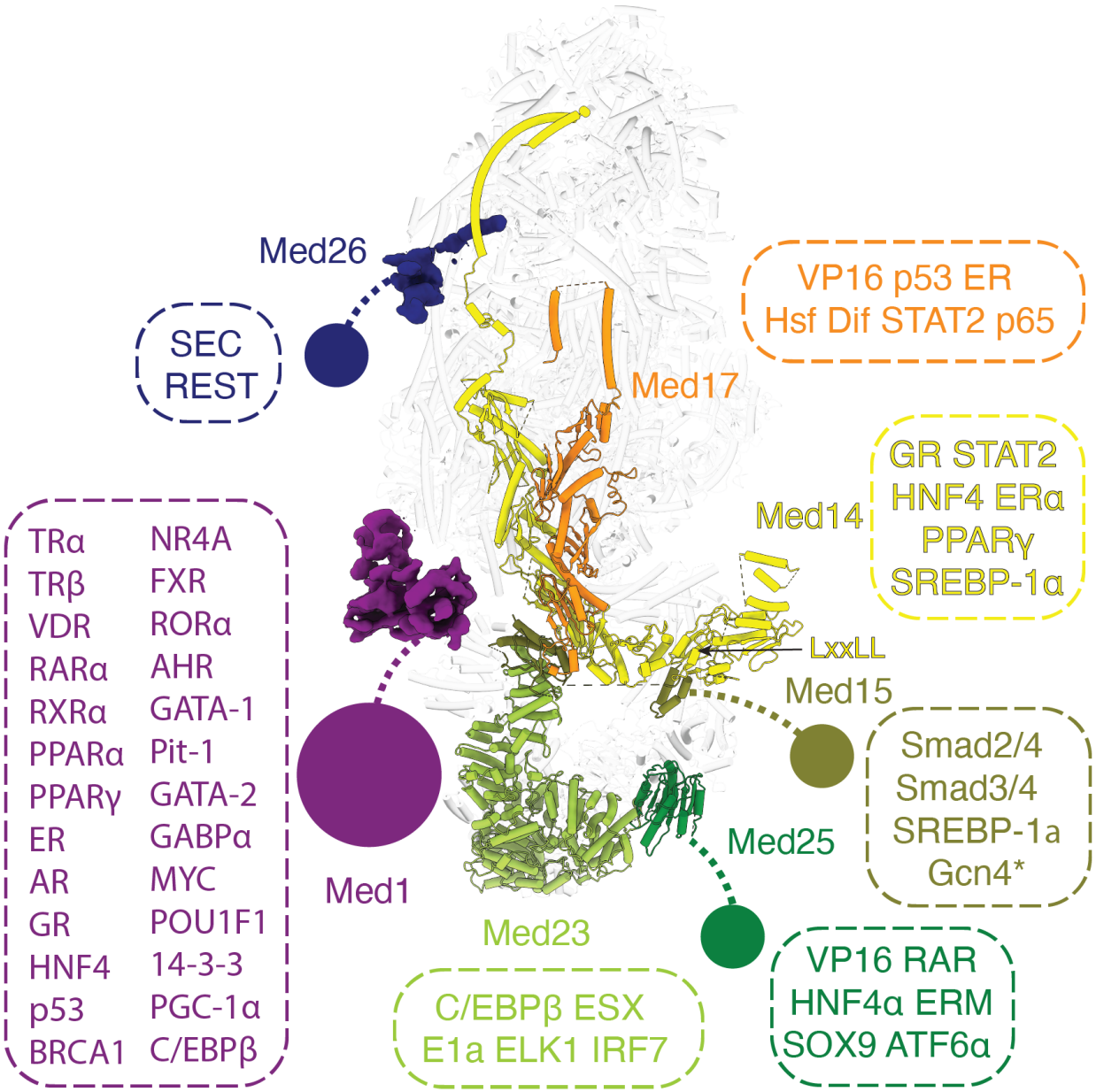
MedHead adopts a very similar structure to the yeast model except for the presence of the additional subunits Med27, 28, 29, and 30 (Figure 3.4, 3.11B). These subunits, which were assigned previously to either MedHead or MedTail [62, 96], exhibit extensive interactions with the fixed jaw of MedHead and were therefore assigned to MedHead. The C-terminus of the scaffold subunit Med14 extends the RM1 and RM2 repeats visible in yeast structures and wraps around MedHead, serving as a clear divider between MedHead and MedTail. Med17, a scaffold subunit within MedHead, stabilizes the fixed jaw on one face and interacts with the RM1 and RM2 repeats of Med14 on the other (Figure 3.11B, 3.12B).

MedTail connects to the rest of Mediator through two relatively small interfaces with MedHead and Med14. Two C-terminal domains of Med15 are crucial for forming both interfaces. The first contact site is located near the C-terminus of Med14. Two helices each from Med27 and Med29 project underneath Med14, with two helices of Med15 (residues 617-649) wedged between them (Figure 3.12C). A concave surface on Med16 makes contact with both this site and Med14. The second site is formed by a C-terminal extension between β^{18} and β^{19} (residues 596-620) of Med17 that interacts with the Ring-WD40-DEAD domain (RWD) of Med15 (residues 674-692) (Figure 3.4, 3.12D). The RWD domain of Med15 is wedged in a large cavity between Med23 and Med24.

The rest of MedTail is formed by subunits Med16, Med23, Med24, and Med25. Med16 is divided into N-terminal and C-terminal domains, with the N-terminus forming a 7-blade WD-40 domain and the C-terminus forming a mostly helical domain that constitutes much of the first interface with MedHead described above (Figure 3.6, 3.12B). The N-terminus of Med24 interacts with Med1 and is much more flexible than the rest of MedTail. We could only identify a single domain of Med25, the von Willebrand factor type A (vWA) domain, wedged in a pocket formed by Med16 and Med23 (Figure 3.12E).

Almost all domains that are bound by transcription factors in Mediator, including the N-terminus of Med15, the N-terminus of Med25, and the C-terminus of Med1, are flexibly attached to the main body and not visible in the density map (Figure 3.14). The first visible portion of Med15 is located underneath MedTail, near the upstream DNA, allowing its N-terminus to easily engage with DNA-bound transcription factors. The C-terminus of Med1 contains the NR-boxes important for nuclear receptor (NR) binding [174]. Many NRs also bind to a C-terminal fragment (1147-1454) of Med14 [175-177]. These two binding interfaces for NRs are quite far from each other (Figure 3.14). The NR AF-1 and AF-2 domains that mediate these interactions are at opposite ends of NR sequences, suggesting that NRs might have to stably associate with the full complex to bridge these two interfaces.

Figure 3.14. **Location of Mediator domains and subunits that interact with transcriptional activators or elongation factors.** Flexible tethered domains are indicated by solid circles connected by dashed lines. All interactions shown are between human factors except Gcn4 which is from yeast and indicated by an asterisk.


Med26
SEC
REST
VP16 p53 ER
Hsf Dif STAT2 p65
Med17
GR STAT2
HNF4 ERα
PPARγ
SREBP-1α
Med14
LxxLL
Med15
Smad2/4
Smad3/4
SREBP-1a
Gcn4*
Med25
VP16 RAR
HNF4α ERM
SOX9 ATF6α
Med23
C/EBPβ ESX
E1a ELK1 IRF7
Med1
TRα
TRβ
VDR
RARα
RXRα
PPARα
PPARγ
ER
AR
GR
HNF4
p53
BRCA1
NR4A
FXR
RORα
AHR
GATA-1
Pit-1
GATA-2
GABPα
MYC
POU1F1
14-3-3
PGC-1α
C/EBPβ

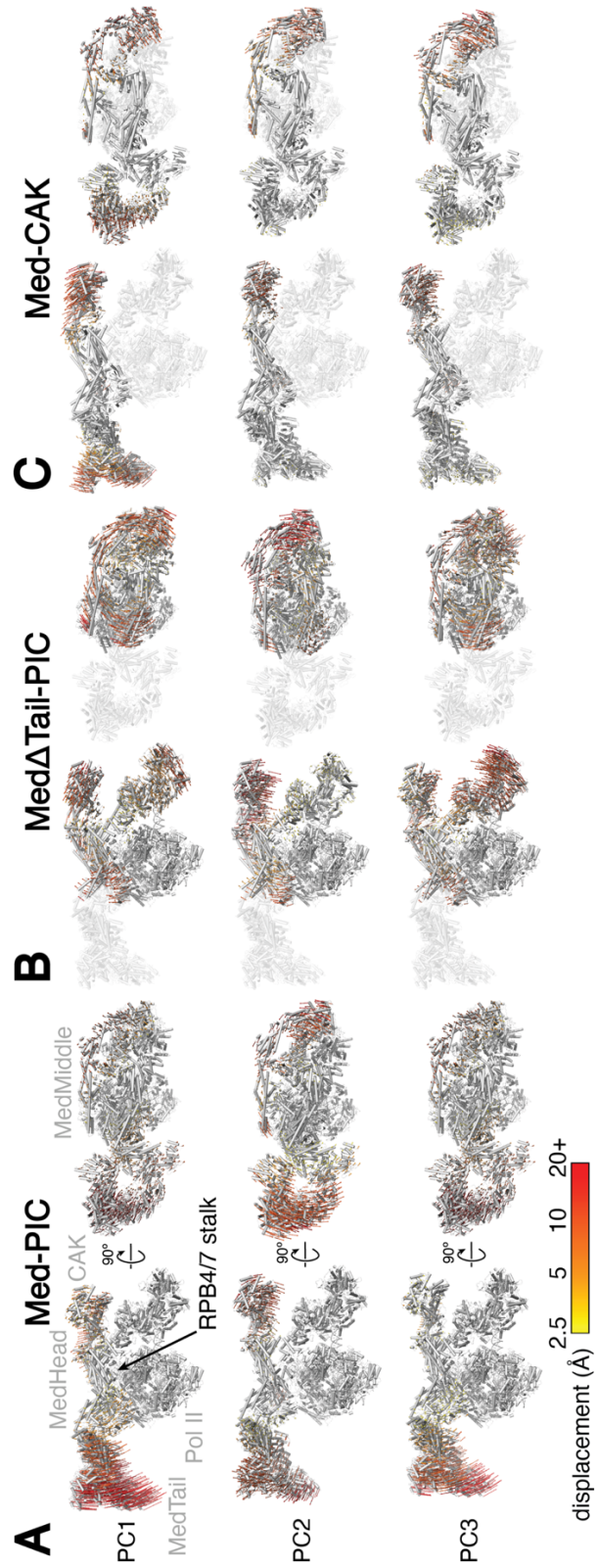
The VP16 activation domain (AD) used to purify Mediator for this study binds to the ACID domain located at the N-terminus of Med25 [112, 113]. The VP16 AD appears to stay bound to Mediator during complex assembly. Due to the absence of density of the ACID domain bound to the VP16 AD in this structure, we can conclude that the ACID domain remains flexibly tethered upon activator binding. It has been hypothesized that conformational changes following activator binding to Mediator could lead to the activation of Med-PIC [90, 116, 178]. Given that so many of the activator-binding domains within Mediator are flexibly tethered to the main body, it is unlikely that this is a universal mechanism for activating Med-PIC for transcription.

While we were in the process of publishing these results, a preprint manuscript describing the structure of the mouse Mediator complex became available [179]. While we do not have access to the models or maps to make a detailed comparison, the overall architecture of Mediator appears highly conserved. The putative locations of Med1 and Med26 described earlier are also in agreement with the mouse structure. A second manuscript, describing the structure of *Chaetomium thermophilum* Mediator was published during this time [180]. While many subunits present in other species, including Med27, Med28, Med29, Med30, Med23, and Med24, are missing in *Chaetomium thermophilum*, the overall architecture appears conserved. The largest potential functional difference is that MedTail appears significantly more flexible in the *Chaetomium thermophilum* structure.

3.3.4. Flexibility of Med-PIC

Because of the size of Med-PIC and the number of rigid bodies required, multi-body refinement in Relion-3 was computationally prohibitive. Instead, we performed non-uniform refinement and 3D variability analysis in CryoSparc [155] which shows a broad distribution of movement of Mediator relative to the PIC (Figure 3.15). This observation explains the low resolution or missing density far from the center of the post-processed map. We performed this analysis on three portions of Med-PIC: Med-PIC, MedΔTail-PIC, and Med-CAK (Figure 3.15). Analysis of the first three principal components for each complex shows a high degree of similarity of movement with the interface between MedHead and the stalk of Pol II, serving as a pivot point for the rotation of Mediator relative to Pol II. This movement can either be up-and-down as in the case of Med-PIC PCs 1 and 3, MedΔTail-PIC PC 2, and Med-CAK PC 1, side-to-side as in the case of Med-PIC PC 2, MedΔTail-PIC PC 1, and Med-CAK PC 2, or a combination of the two as in MedΔTail-PIC PC 3 and Med-CAK PC 3.

Figure 3.15. **3D variability analysis of Med-PIC.** A) Top three principal components (PCs) of movement within Med-PIC. MedTail and the CAK module undergo the largest displacements in Med-PIC. B) Top three PCs of movement within MedΔTail-PIC. When isolating movement from MedTail, the rotation of MedHead-MedMiddle-CAK and TFIIH relative to the cPIC is more readily visible. C) Top three PCs of movement within Med-CAK. The movement of MedTail and MedMiddle-CAK is largely independent of each other. PC1 shows that the interface between MedHead and MedTail can act as a hinge, which is reasonable given how small the interface is between the two modules. Movements are colored from yellow (small) to red (large).


A
Med-PIC
B
MedΔTail-PIC
C
Med-CAK
MedHead
CAK
MedMiddle
MedTail
Pol II
RPB4/7 stalk
PC1
PC2
PC3
90°
displacement (Å)
2.5
5
10
20+

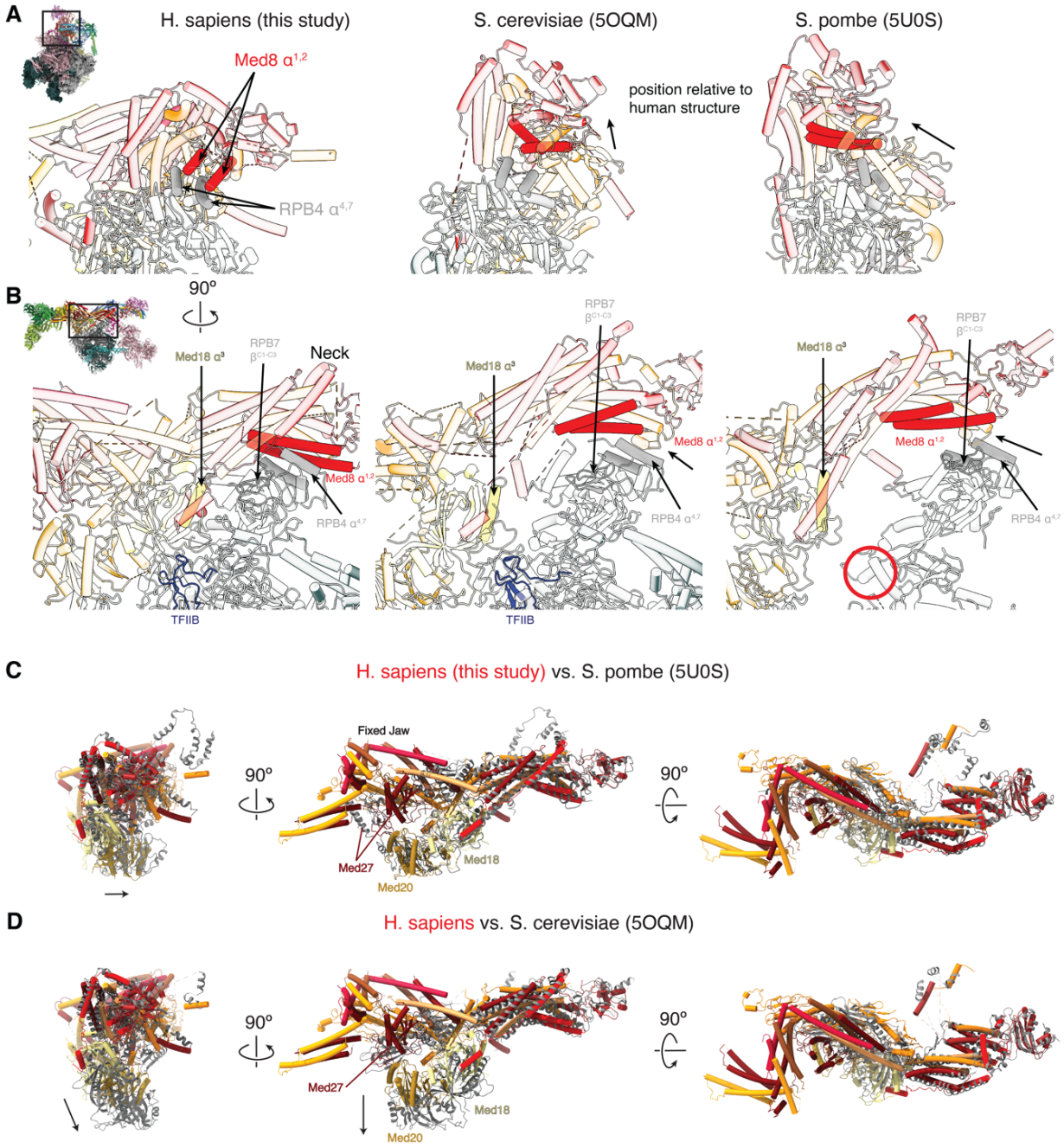
This analysis led me to compare the interface between MedHead and Pol II in my human structure with that in the existing yeast Med-PIC structures. MedHead forms a closer association with the stalk of Pol II (RPB4/7) in the human structure than seen in any previous structure to date (Figure 3.16A-B). Helices α^1 and α^2 of Med8 stack on top of helices α^4 and α^7 of RPB4 in all three Med-PIC structures available. However, the extent of that interaction differs significantly between species. In the human structure, these pairs of helices run parallel to each other, forming an extensive interface between MedHead and the RPB4/7 stalk, highlighted by close interactions between Med18 α^3 and RPB7 $\beta^{C1\text{-}C3}$ (Figure 3.16A-B). In the scMed-PIC [67], scMedHead slides towards scMedTail, resulting in a sinking of the Med18-Med20 flexible jaw away from the stalk and the RNA exit tunnel of Pol II, and a lifting of the shoulder of Mediator (Figure 3.16A-B). Due to the stabilization of the CAK module by the shoulder domain, this change would likely result in a lifting of the CAK module or shifting of the interface. In the *S. pombe* Med-PIC (spMed-PIC) structure [108], this movement is even more exaggerated with minimal overlap between the Med8 and RPB4 helices, resulting in an even larger gap between the stalk and flexible jaw and a slight rotation of spMedHead away from spMedMiddle. We were unable to identify a prominent principal component in our data set that captured the positions of yeast MedHead relative to Pol II.

Superimposing human MedHead with the scMedHead and spMedHead structures shows that they align very well with just subtle movements of the flexible jaw between species (Figure 3.16C-D). The only difference is the position of the mobile jaw, Med18, and Med20. In humans and *S. pombe*, the presence of the Med27 subunit stabilizes the mobile jaw, but its absence in *S. cerevisiae* causes the sinking of the mobile jaw away from the fixed jaw.

Comparison of the *S. pombe* apo-Mediator and Med-cPIC structures shows that Med14 contains a hinge between the RM1 and RM2 domains, which ultimately leads to the raising and

lowering of spMedMiddle, relative to spMedHead [108]. This is very similar to Med-CAK PC 1, suggesting this flexibility remains after engagement with the PIC (Figure 3.16C).

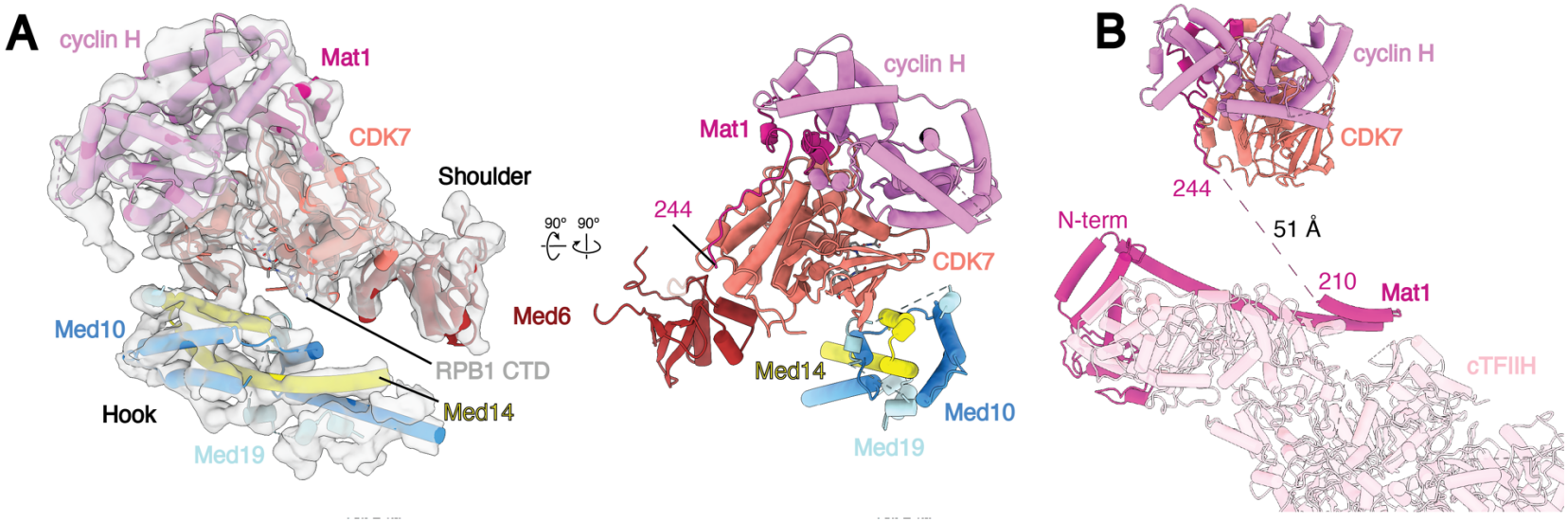
Figure 3.16. **Comparison of MedHead positions relative to Pol II.** A) Human MedHead makes extensive contacts with the stalk of Pol II with helices α^1 and α^2 from Med8 stacking on top of helices α^4 and α^7 of RPB4 and forming an extensive interface between MedHead and the RPB4/7 stalk. In *S. cerevisiae* and *S. pombe,* MedHead slides towards MedTail. B) This movement of MedHead from human to *S. cerevisiae* to *S. pombe* results in an uncovering of the RNA exit channel, occupied by TFIIB (blue ribbon), in the first two structures. The distance between Med18 α^3 and RPB7 $\beta^{C1\text{-}C3}$, which define this gap, is highlighted as opaque ribbon. A red circle denotes where the exit channel is in the *S. pombe* structure. The view of A and B relative to the full complex is shown on the far left. C-D) Superimposing human MedHead (colored/tube) with the spMedHead (C, gray/ribbon, PDB:5U0S) and scMedHead (D, gray/ribbon, PDB:5OQM) structures show that they align very well with just subtle movements of the mobile jaw (Med18 and Med20) between species. In humans and *S. pombe*, the presence of the Med27 subunit stabilizes the mobile jaw, but its absence in *S. cerevisiae* likely allows the sinking of the mobile jaw away from the fixed jaw. Models are colored as in Figure 3.8.


A
H. sapiens (this study)
S. cerevisiae (5OQM)
S. pombe (5U0S)
Med8 α1,2
RPB4 α4,7
position relative to human structure
B
90°
Med18 α3
RPB7 βC1-C3
Neck
Med8 α1,2
RPB4 α4,7
TFIIB
C
H. sapiens (this study) vs. S. pombe (5U0S)
90°
Fixed Jaw
Med27
Med18
Med20
D
H. sapiens vs. S. cerevisiae (5OQM)
90°
Med27
Med18
Med20

3.3.5. Mediator stabilizes the CAK module of TFIIH

While previous structural studies of Med-PICs established that the CAK module of TFIIH occupies a position between the shoulder and hook domains of Mediator, the position and orientation of each CAK module subunit could not be determined [67, 106-108]. Rigid body docking of the human CAK module structure into our density led to an unambiguous orientation of the CAK module with the active site of CDK7 facing the hook domain of MedMiddle (Figure 3.16A) [167]. Mediator stabilizes the CAK module through interactions involving Med6, the N-terminus of Med14, and a small fragment of Med19 (~133-148) with CDK7 (Figure 3.17A). This orientation of the CAK module positions the C-terminus of Mat1 ~50 Å from the N-terminus bound to cTFIIH, a distance easily spanned by the small fragment of Mat1 (211-243) missing in the structure (Figure 3.17B).

Figure 3.17. **Structure of TFIIH within Med-PIC** A) Docking of the CAK module (CDK7, cyclin-H, and Mat1) within the MedMiddle-CAK density. The CAK module of TFIIH is stabilized in the Med-PIC by interactions between CDK7 and Med6, the N-terminus of Med14, and a small fragment of Med19. B) The model of the complete human TFIIH complex places the two modeled segments of Mat1 (1-210, 244-308) close to each other. The missing 34 residues can easily span the 51 Å distance between the termini. Models are colored as in Figure 3.8.


A
cyclin H
Mat1
CDK7
Shoulder
90°
90°
Med10
Med6
RPB1 CTD
Hook
Med14
Med19
244
cyclin H
Mat1
CDK7
Med14
Med10
Med19
B
cyclin H
CDK7
244
51 Å
N-term
210
Mat1
cTFIIH

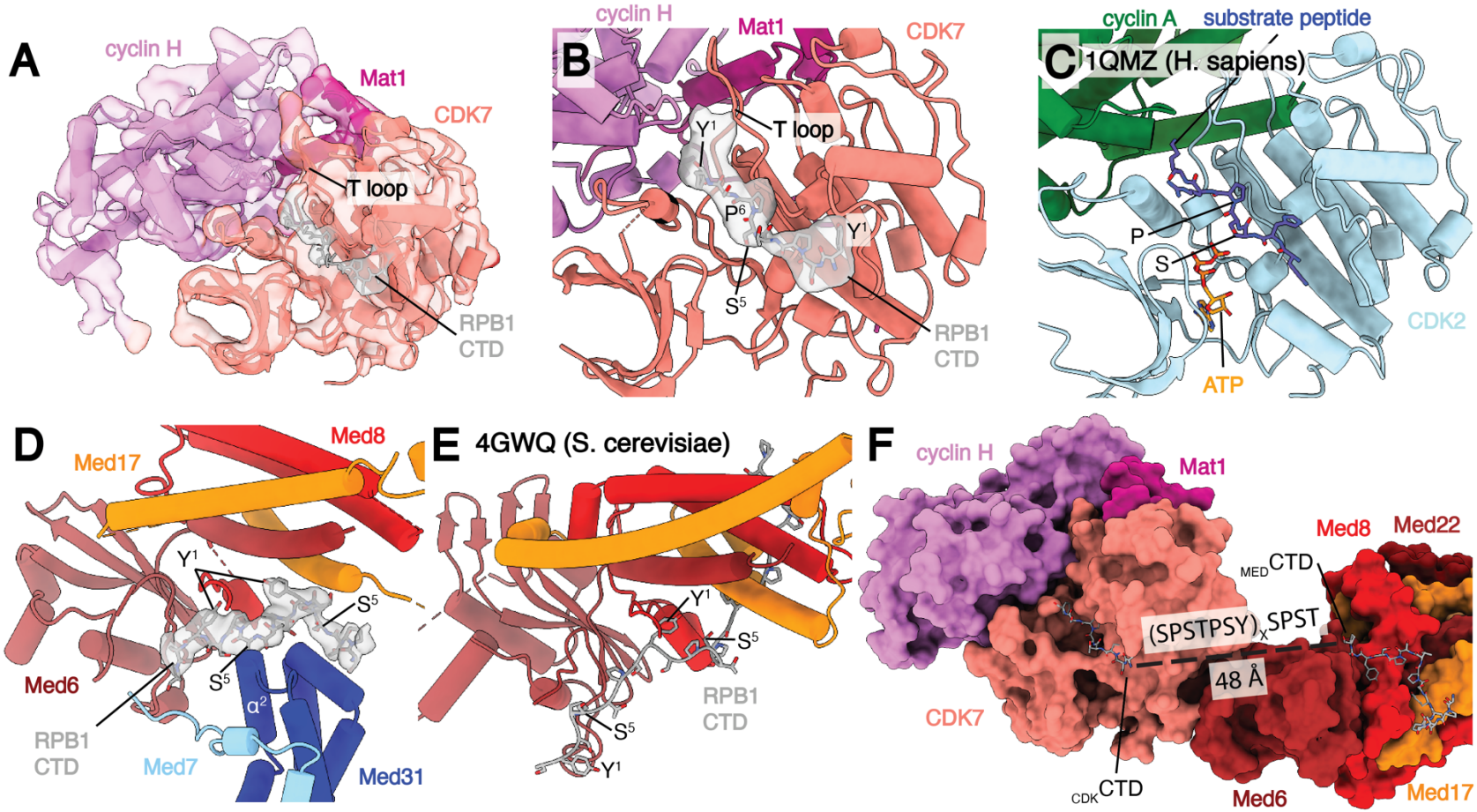
CDK7 adopts the active conformation seen in the human CAK module structure, with the T-loop projecting towards Mat1 and away from the active site (Figure 3.18A-B) [167]. Clear electron density in the CDK7 active site closely matches the location of the substrate peptide in the CDK2-cyclin-A-substrate peptide complex [168] (Figure 3.18A-C). This peptide shares the identical serine-proline sequence that is found in the RPB1 CTD targeted by the kinase. Therefore, we built a model for the RPB1 CTD in the active site that we designate as $_{CDK}$CTD.

S. cerevisiae MedHead (scMedHead) was co-crystallized with a short peptide of the RPB1 CTD, which shows slightly more than three full repeats engaged with scMedHead at the shoulder and neck domains [97]. We observe additional electron density in this same location and used the *S. cerevisiae* structure to build a model for this portion of the CTD that we will refer to as $_{MED}$CTD (Figure 3.17D-E). $_{MED}$CTD is 16 residues long, slightly more than two full repeats, and adopts a somewhat different path than the yeast structure, likely due to the presence of Med31, which interacts with the other side of $_{MED}$CTD. In scMedHead, the elongated structure of the N-terminal portion of $_{MED}$CTD forms extensive interactions with Med17. In contrast, we see clear density for $_{MED}$CTD starting to wrap around Med31. The C-terminal end of $_{MED}$CTD also does not form as extensive of an interface with Mediator as in scMedHead, due to a clash with the Med7 N-terminus. Experiments in *S. pombe* show that the CTD is necessary for interaction between MedHead and Pol II *in vitro,* suggesting that $_{MED}$CTD is critical for this interaction [106]. $_{MED}$CTD binding to Mediator would likely be disrupted following phosphorylation of Ser^5 due to close interactions between Ser^5 and the end of Med31 helix α^2 (Figure 3.18D).

The directionality of $_{MED}$CTD and $_{CDK}$CTD is the same, with the N-terminal end of $_{MED}$CTD pointing towards Pol II and the C-terminal end of $_{CDK}$CTD leading away from Med-PIC (Figure 3.18F). This observation strongly suggests that $_{MED}$CTD is N-terminal to $_{CDK}$CTD within the full CTD sequence. The distance between the termini of those two CTD fragments is 48 Å. In an

elongated state, one repeat of the CTD can span approximately 25 Å [181], so while two repeats of the CTD may be sufficient to bridge that gap, we would likely see better-defined density for the CTD in that case. Therefore, we suspect that three or more repeats are likely looped out between $_{MED}$CTD and $_{CDK}$CTD.

Figure 3.18. **Location of RPB1 CTD binding in Med-PIC.** A) Structure of the TFIIH CAK module. Segmented map of MedMiddle-CAK shows clear density representing an active conformation of the T-loop of CDK7 and density for Pol II CTD in the active site of CDK7. B) Model of the CAK module with density observed for the $_{CDK}$CTD in the active site. A consensus sequence of the Pol II CTD is modeled due to limited resolution. The T-loop is in the extended, active conformation. C) Model of the CDK2-cyclin A-substrate peptide structure shows high similarity to the CAK module structure with the conserved SP motif that is common to substrates of both enzymes. D) Model and density of $_{MED}$CTD with interacting subunits of MedHead and MedMiddle. S^5 makes close contacts with α^2 of Med31, preventing binding of phosphorylated repeats in this location. E) Model of $_{MED}$CTD in the yeast MedHead crystal structure shows a more extensive interface between $_{MED}$CTD and MedHead than in the Med-PIC, likely due to the presence of MedMiddle in the Med-PIC. F) View of $_{CDK}$CTD and $_{MED}$CTD within the human Med-PIC structure. Based on the directionality of the CTD, $_{CDK}$CTD is C-terminal to $_{MED}$CTD, and the gap between them would require at least two repeats of the CTD. MedMiddle is hidden for easier visibility. Models are colored as in Figure 3.8. Annotated domains of Mediator are labeled in black.


A
cyclin H
Mat1
CDK7
T loop
RPB1
CTD
B
cyclin H
Mat1
CDK7
T loop
Y1
P6
Y1
S5
RPB1
CTD
C
cyclin A
substrate peptide
1QMZ (H. sapiens)
P
S
CDK2
ATP
D
Med8
Med17
Y1
S5
S5
Med6
α2
RPB1
CTD
Med7
Med31
E
4GWQ (S. cerevisiae)
Y1
S5
RPB1
CTD
S5
Y1
F
cyclin H
Mat1
Med8
Med22
MED CTD
(SPSTPSY)X SPST
48 Å
CDK7
CDK CTD
Med6
Med17

CHAPTER 4:

Conclusions

4.1. Summary of Findings

4.1.1. Structure of PIV5 L-P complex

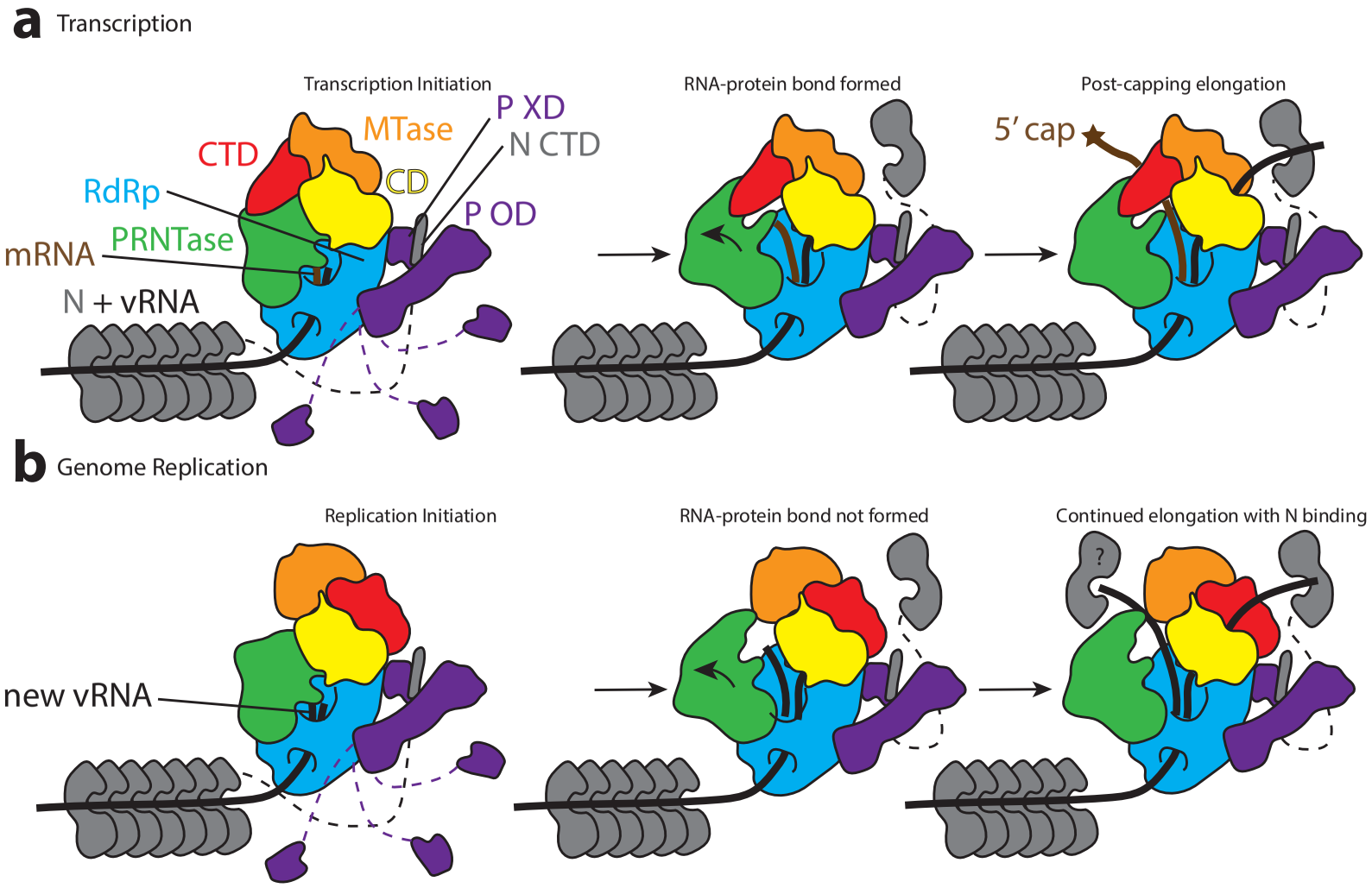
NNS viruses require both L and P to form competent polymerase complexes to function on the genomic RNA wrapped around the N protein. Until recently, only two NNS virus polymerase structures have been determined with key modules missing from these studies, leading to many questions about how L and P cooperate during viral RNA synthesis. My structure represents the first of a complete viral L protein, with all domains visible, in complex with its full-length co-factor P. Major differences in the extent of the L-P interface and a rearrangement of the MTase-CTD module help elucidate the mechanistic cycle of these essential polymerase complexes.

4.1.2. Mechanism of L-P procession along the N-coated RNA genome

The differences in the extent of the L-P interfaces and differences in the P-CTD fold among NNS viruses lead to many questions about how L-P interacts with the viral genome. Recent biochemical studies have shown that one copy of P-XD is essential and sufficient for L-P or N-P interaction, and the MoRE motif-containing N-tail is not essential for the preliminary binding of the polymerase to the RNA template [58, 182]. These results are consistent with a model in which repeated association and separation of P-XD and N-tail allows P-XD to cartwheel along the N-coated genome to enable the polymerase to scan the RNA template [183]. I hypothesize that P-XD might play a role in bridging N to L through two non-overlapping interfaces (Figure 4.1). Anchoring of an N molecule by this single copy of P-XD would limit the movement of the globular domain of N, positioning N to re-capture the emerging template genomic RNA. This leads me to hypothesize a mechanism where once the N-RNA interaction has been disrupted upon the genome entering the L template entry tunnel, N is captured through the MoRE-P-XD interaction and eventually brought to the anchoring position on L. Capturing N monomers after dissociation

from the genomic RNA is likely critical during initial infection when no new N protein has been synthesized to recycle all existing N protein to coat the genome after it has been transcribed (Figure 4.1A). This agrees with existing evidence that transcription of mRNA precedes genome replication, which would require additional copies of N to coat the new (anti)-genomic RNA strand in addition to the template strand [14, 15]. With four strands of P per polymerase in paramyxoviruses, up to four monomers of N could be retained during transcription, allowing up to 24 bases to be displaced from the N-encapsidated genome at a time.

Figure 4.1. **Model of transcription and genome replication by the L-P complex.** A) In transcription initiation, the MTase-CTD module is positioned directly above the PRNTase domain, as in the PIV5 structure. This positions the active site of the MTase as that the covalently linked RNA-PRNTase is pushed up into it, leading to productive capping and methylation. This requires an outward movement of the PRNTase domain to accommodate the growing RNA strand. I hypothesize that a P-XD captures the monomer of N that is no longer bound to the genomic RNA, keeping it close so it can be used to recapture the genomic RNA re-emerging from the template exit channel. Additional P-XDs are shown in the first panel but removed for clarity from subsequent panels. B) In genome replication, the MTase-CTD module is positioned further away from the PRNTase domain. No covalent linkage is formed between the RNA and conserved histidine, but an outward movement of the PRNTase domain is still required to accommodate the growing RNA strand. The newly released monomer of N is captured in the same way as in (A), and a second copy of N is used to coat the newly synthesized anti-genome by an unknown mechanism.


a Transcription
Transcription Initiation
RNA-protein bond formed
Post-capping elongation
P XD
MTase
N CTD
CTD
CD
RdRp
P OD
PRNTase
mRNA
5′ cap
N + vRNA
b Genome Replication
Replication Initiation
RNA-protein bond not formed
Continued elongation with N binding
new vRNA
?

Interactions between P-XD and N or L are thought to be dynamic, leading to three possibilities for how N eventually ends up bound through P-XD to L. One possibility involves the rotation of the P tetramer relative to L, allowing the different P-XD domains to cycle through interactions with N and L as the RNA-N complex disassembles and reassembles. A second mechanism occurs without rotation, where an N monomer could be passed from one P-XD to another until it arrives at the P-XD bound to L due to the transient interactions between N, P, and L. The third mechanism does not require rotation or the handoff of N from one P-XD to another but just relies on the length of the flexible linker between P-OD and P-XD such that each copy of P-XD is capable of reaching over to the binding site of P-XD on L. In all cases, once N is positioned to coat the genome emerging from the template exit site, it is now furthest away from the flexible P-XD domains, and transient dissociation could lead to complete dissociation from the polymerase complex. Despite our structural observation, biochemical studies have suggested that a single copy of P-XD in MeV cannot interact with both L and N at the same time [58]. Further studies are necessary to rationalize these competing observations.

Even though the P-XD fold is not conserved across NNS viruses, the C-terminus of VSV P adopts an entirely different fold that also binds N [41, 184]. Pneumovirus P proteins have also been shown to bind to N, but no structure of this interaction exists yet [185, 186]. Given that the chain of RSV P that binds in the same place as P-XD does not adopt a globular fold, it is less likely that L, N, and P could all bind together as appears possible in paramyxoviruses. Even though significant differences exist in these interfaces, their shared location and binding abilities suggest a common mechanism of P binding to N during the transcriptional cycle. The highly divergent nature of P both structurally and sequence-wise suggests that differences in regulatory mechanisms likely do exist.

4.1.3. Location of the N-terminus of P

The N-terminus of P interacts with nascent N^0-P to prevent premature encapsidation of RNA [2]. The angle at which the P-OD projects away from L positions its N-terminus far away from L. Some NNS viruses, including RSV and VSV, have much shorter P-ODs, leading to a wide variation in the positioning of the N-terminus relative to L [26, 49]. There are ~200 unstructured residues between the PIV5 P-OD and the N-terminal domain, making it difficult to predict the three-dimensional organization of the N-terminus. The N-termini could interact with the transiently tethered N monomers to prevent premature encapsidation of RNA during both transcription and genome replication. Alternatively, it could function only during genome replication as a hub for coating the newly synthesized (anti-)genome using the N monomers recycled from the template.

4.1.4. Coupling of conformational rearrangements and the transcription/replication switch

Based on steric clashes and a poorly positioned active site in published structures, it has been hypothesized that the PRNTase domain needs to undergo a substantial conformational change to accommodate the growing RNA strand [11]. The resulting conformational change also likely forms the complete active site for capping and methylation. This observation is corroborated by the strict requirement for 31 nucleotides to be synthesized before capping and methylation can occur [37]. Here, I provide a structural basis for this hypothesis where the arrangement of the MTase-CTD dimer relative to the CD in my PIV5 structure positions the active site of the MTase domain directly adjacent to the PRNTase domain.

This conformation of the MTase-CTD module exists during transcription of mRNA to allow proper processing of the nascent transcripts (Figure 4.1A). Following initiation, the PRNTase domain is required to open to accommodate the synthesis of 31 nucleotides to proceed to a pre-

capping elongation state. Continued transcription could push the flexible intrusion loop with RNA bound to the conserved histidine toward the active site of the MTase-CTD. This would create the active site for capping, followed immediately by methylation.

This model then leads to the conclusion that the conformation of the MTase-CTD dimer observed in the existing rhabdovirus structures is likely utilized during genome duplication (Figure 4.1B) [11, 27, 28]. The transition from initiation to the equivalent of the transcriptional pre-capping elongation state still requires the PRNTase domain to open. However, no capping or methylation occurs, and RNA synthesis continues into the equivalent of the post-capping elongation phase. Domain exchange experiments with different strains of VSV showed that swapping the MTase and CTD domains from the New Jersey strain into the Indiana strain caused a preference for replication over transcription, suggesting that the interface between the CD-MTase-CTD module and RdRp-PRNTase module might bias towards one conformation or the other and that these conformations favor transcription or genome replication [38].

Surprisingly, only a single conformation is present in both the PIV5 and VSV data sets. All my attempts to look for a conformation of the PIV5 complex with the MTase-CTD dimer in the same position as in the VSV structure failed. While I cannot conclude that there are no complexes that adopt that configuration, if they are present, they are likely well below 5% occupancy of the full data set and, therefore, too difficult to sort out from the rest of the data.

Because transcription precedes genome replication in the viral life cycle, questions remain as to (1) what is the mechanism that favors folding L into the conformation seen in PIV5; (2) what triggers the change to the state seen in VSV; and (3) can the same L-P complex transition between the two states. The CD-MTase-CTD module in the recently solved pneumovirus L-P complex structures is not visible due to its high flexibility [25, 26, 29]. *In vivo,* there are likely to be

other factors that could affect the folding pathways of these complexes, including the presence of large amounts of N protein [14-16].

4.1.5. Insights into repetitive CTD phosphorylation in Med-PIC

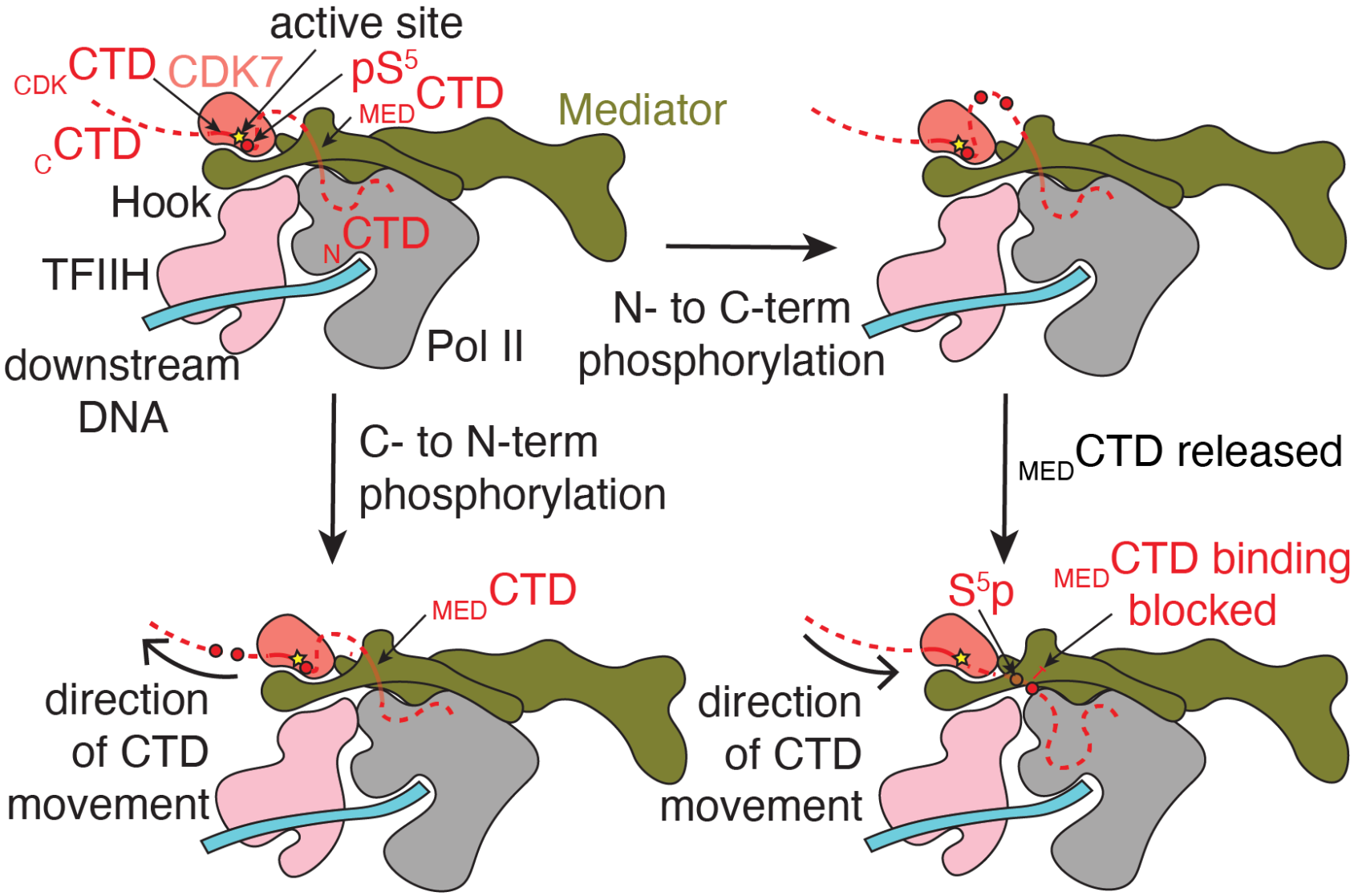
The role of $_{MED}$CTD binding is likely to capture the CTD and position it in the correct direction and close to the active site of CDK7 to facilitate pS^5 formation. Mass spectrometry experiments with both yeast and human complexes show that $pSer^5$ can be found within any repeat of the CTD except the final repeat [75, 76]. However, the phosphorylation patterns of individual CTD peptides and the direction that sequential phosphorylation can occur remain unknown. Two possibilities exist for the direction of sequential phosphorylation that generates different outcomes (Figure 3.18). If the CTD is phosphorylated in a C- to N-terminal direction, binding at $_{MED}$CTD precedes phosphorylation, and it is not clear how Pol II would dissociate from Mediator given that the CTD is threaded through a hole in Mediator formed by the hook, knob, and shoulder domains and the CAK module of TFIIH. Phosphorylated repeats would also be located far from the nascent RNA that needs to be capped.

If the CTD is phosphorylated in an N- to C-terminal direction, C-terminal phosphorylated repeats would not be able to bind at $_{MED}$CTD due to steric clashes that would arise with the added phosphates. Given that the CTD is important for Pol II-Mediator interaction and phosphorylation of the CTD leads to dissociation of Pol II and Mediator, we find this mechanism more likely [106, 187]. Separation of MedHead and Pol II would place the phosphorylated CTD close to the nascent RNA for capping to occur.

Given the large movements of MedMiddle and the CAK module of TFIIH relative to the PIC, we speculate that these conformational changes play an important role in the sequential phosphorylation of the CTD. The intrinsic flexibility of Mediator has been linked to the opening

and closing of the $_{MED}$CTD binding site on Mediator [108, 124], and if this movement is tied to binding and release of the CTD at $_{MED}$CTD, it could also facilitate the progression of CDK7 along the CTD.

Figure 3.18. **Model for phosphorylation of the Pol II CTD by CDK7.** $_{MED}$CTD binding positions the rest of the CTD in the CDK7 active site. Following phosphorylation, indicated by a red circle, translocation of the CTD towards the N-terminus (bottom) would place phosphorylated repeats further from the nascent RNA emerging from Pol II. Separation of Mediator and Pol II would be difficult without separation of the CAK module and Mediator. Translocation of the CTD towards the C-terminus would position phosphorylated repeats to block binding of the CTD at $_{MED}$CTD, a possible way to favor disassembly of Med-PIC. Phosphorylated repeats would also be significantly closer to the RNA exit tunnel of Pol II to recruit the capping complex properly. CAK = cyclin-activated kinase module; CTD = C-terminal domain of RPB1; pS^5 = phosphorylated serine 5 residue (red circle).


active site
CDK CTD
CDK7
pS5 CTD
MED
Mediator
C CTD
Hook
N CTD
TFIIH
Pol II
downstream DNA
N- to C-term phosphorylation
C- to N-term phosphorylation
MED CTD released
MED CTD
direction of CTD movement
S5p
MED CTD binding blocked
direction of CTD movement

4.2. Significance and Impact

4.2.1. Mononegavirus transcription

Parainfluenza virus 5 (PIV5) belongs to the family *Paramyxoviridae,* whose members pose significant health burdens to humans and animals. The viral RNA genome is packaged within a nucleoprotein complex, which serves as the template for genome replication and transcription. The L protein is responsible for RNA synthesis, capping, and methylation and requires a cofactor, the P protein, for RNA synthesis *in vivo*. This study provides a near-atomic resolution structure of a complete paramyxovirus L-P complex, an attractive target for drug design against paramyxoviruses. Comparisons to structures of other mononegavirus polymerases identify a significant conformational rearrangement of the MTase and CTD relative to the RdRp domain. These different locations elucidate the mechanism that allows the complex to switch between genome replication and transcription.

4.2.2. Eukaryotic transcription

Assembly of the PIC at eukaryotic promoters represents a critical regulatory mechanism for cells to dictate which genes to transcribe. This process is driven by transcription factors recruiting Mediator, which facilitates the assembly of the PIC and stimulates the activity of a crucial kinase, CDK7, within the PIC. Here, I solve the structure of the entire Mediator-bound PIC, with many parts of the complex, including MedTail, at sub-4 Å. This allowed me to build an atomic model of Med-PIC with a high degree of confidence. MedTail contains flexibly attached domains crucial for interactions with transcription factors, allowing for flexibility in how Med-PIC assembled *in vivo.* Visualizing multiple locations of CTD binding within the complex shows how binding of $_{MED}$CTD between MedHead and MedMiddle positions $_{CDK}$CTD in the proper location for phosphorylation by CDK7. I also visualize significant conformational flexibility of Mediator relative to Pol II, and

that leads me to hypothesize that this movement is responsible for sequentially phosphorylating the Pol II CTD.

4.3. Future Directions

4.3.1. Mononegavirus transcription

My structure of the PIV5 L-P complex provides valuable structural information that furthers our understanding of the mechanism of these polymerases. Differences in the positioning of the MTase and CTD lead to intriguing hypotheses regarding how the polymerase chooses between genome replication and transcription. Designing mutants that favor one conformation or another and testing the effect on capping and methylation are critical to interrogating the mechanism. Since the presence of newly transcribed N monomers has been implicated in the transition from transcription to genome replication, the addition of N to this complex to see if there is a stable complex that is formed that could explain this phenomenon. Two significant challenges remain: trapping an L-P complex at specific points during transcription to understand how the polymerase accommodates the transcription of 31 bases before capping and methylation occur, and assembling the L-P complex on an N-coated piece of RNA to understand how the polymerase acts on its native substrate.

4.3.2. Eukaryotic transcription

While the structure presented here shows two binding sites for the CTD within Med-PIC, it is still just a mostly static image of a movie. More remains to be done to visualize steps in this process directly. Treatment of purified Pol II with phosphatase is a promising technique to reduce the heterogeneity of the complex. Unfortunately, it is a finicky treatment and led to decreased stability of Pol II in certain instances. Incorporating phosphatase treatment into the Pol II

purification protocol would reduce the complexity of complex assembly and lead to a better sample for structural studies.

Improving the resolution of flexible portions of Med-PIC including MedMiddle-CAK and cTFIIH can be addressed through the use of denoising software to improve the signal-to-noise ratio of individual particles, critical for local refinements of small bodies.

Changing the DNA template to include promoter-proximal transcription factor binding sites would allow for the formation of the first activator-bound Med-PIC. Comparing that structure to this Med-PIC structure would show definitely what differences occur upon transcription factor binding as has been hypothesized previously. In addition, incorporation of TFIID into Med-PIC instead of TBP would help confirm the apparent compatibility of that structure based on our integrated modeling.

REFERENCES

1. Alayyoubi, M., et al., *Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein \textendash RNA complex.* Proceedings of the National Academy of Sciences, 2015. **112**(14): p. E1792--E1799.
2. Aggarwal, M., et al., *Structure of the Paramyxovirus Parainfluenza Virus 5 Nucleoprotein in Complex with an Amino-Terminal Peptide of the Phosphoprotein.* Journal of virology, 2018. **92**(5): p. 2242--14.
3. Welch, B.D., et al., *Structure of the parainfluenza virus 5 (PIV5) hemagglutinin-neuraminidase (HN) ectodomain.* PLoS Pathog, 2013. **9**(8): p. e1003534.
4. Welch, B.D., et al., *Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein.* Proc Natl Acad Sci U S A, 2012. **109**(41): p. 16672-7.
5. Poor, T.A., et al., *On the stability of parainfluenza virus 5 F proteins.* J Virol, 2015. **89**(6): p. 3438-41.
6. Ogino, T. and T.J. Green, *RNA Synthesis and Capping by Non-segmented Negative Strand RNA Viral Polymerases: Lessons From a Prototypic Virus.* Front Microbiol, 2019. **10**: p. 1490.
7. Calain, P. and L. Roux, *The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA.* J Virol, 1993. **67**(8): p. 4822-30.
8. Parks, R.A.L.G.D., *Paramyxoviridae: The Viruses and Their Replication*, in *Fields Virology*, D.M.K. Bernard N Fields, Peter M Howley, Editor. 2013, Lippincott, Williams, and Wilkins: Philadelphia. p. 957-995.
9. Kolakofsky, D., *Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: A review.* Virology, 2016. **498**: p. 94-98.
10. Kuo, L., et al., *Effects of mutations in the gene-start and gene-end sequence motifs on transcription of monocistronic and dicistronic minigenomes of respiratory syncytial virus.* J Virol, 1996. **70**(10): p. 6892-901.
11. Liang, B., et al., *Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy.* Cell, 2015. **162**(2): p. 314--327.
12. Noton, S.L., C.Z. Tremaglio, and R. Fearns, *Killing two birds with one stone: How the respiratory syncytial virus polymerase initiates transcription and replication.* PLoS pathogens, 2019. **15**(2): p. e1007548--8.
13. Noton, S.L. and R. Fearns, *Initiation and regulation of paramyxovirus transcription and replication.* Virology, 2015. **479-480**(C): p. 545--554.
14. Noton, S.L. and R. Fearns, *Initiation and regulation of paramyxovirus transcription and replication.* Virology, 2015. **479-480**: p. 545-54.
15. Plumet, S., W.P. Duprex, and D. Gerlier, *Dynamics of viral RNA synthesis during measles virus infection.* J Virol, 2005. **79**(11): p. 6900-8.
16. Hoffman, M.A. and A.K. Banerjee, *Precise mapping of the replication and transcription promoters of human parainfluenza virus type 3.* Virology, 2000. **269**(1): p. 201-11.
17. Qanungo, K.R., et al., *Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA.* Proc Natl Acad Sci U S A, 2004. **101**(16): p. 5952-7.
18. Vidal, S. and D. Kolakofsky, *Modified model for the switch from Sendai virus transcription to replication.* J Virol, 1989. **63**(5): p. 1951-8.

19. Kolakofsky, D., et al., *Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited.* J Virol, 1998. **72**(2): p. 891-9.
20. Sidhu, M.S., et al., *Canine distemper virus L gene: sequence and comparison with related viruses.* Virology, 1993. **193**(1): p. 50-65.
21. Poch, O., et al., *Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains.* J Gen Virol, 1990. **71 (Pt 5)**: p. 1153-62.
22. Ogino, T., et al., *Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA.* J Biol Chem, 2005. **280**(6): p. 4429-35.
23. Chandrika, R., et al., *Mutations in conserved domain I of the Sendai virus L polymerase protein uncouple transcription and replication.* Virology, 1995. **213**(2): p. 352-63.
24. Paesen, G.C., et al., *X-ray structure and activities of an essential Mononegavirales L-protein domain.* Nat Commun, 2015. **6**: p. 8749.
25. Pan, J., et al., *Structure of the human metapneumovirus polymerase phosphoprotein complex.* Nature, 2019.
26. Gilman, M.S.A., et al., *Structure of the Respiratory Syncytial Virus Polymerase Complex.* Cell, 2019. **179**(1): p. 193-204 e14.
27. Jenni, S., et al., *Structure of the Vesicular Stomatitis Virus L Protein in Complex with Its Phosphoprotein Cofactor.* Cell Rep, 2020. **30**(1): p. 53-60 e5.
28. Horwitz, J.A., et al., *Structure of a rabies virus polymerase complex from electron cryo-microscopy.* Proc Natl Acad Sci U S A, 2020. **117**(4): p. 2099-2107.
29. Cao, D., et al., *Cryo-EM structure of the respiratory syncytial virus RNA polymerase.* Nat Commun, 2020. **11**(1): p. 368.
30. Jordan, P.C., et al., *Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase.* PLoS pathogens, 2018. **14**(2).
31. Rahmeh, A.A., et al., *Molecular architecture of the vesicular stomatitis virus RNA polymerase.* Proc Natl Acad Sci U S A, 2010. **107**(46): p. 20075-80.
32. Gilman, M.S.A., et al., *Structure of the Respiratory Syncytial Virus Polymerase Complex.* Cell, 2019. **179**(1): p. 193--204.e14.
33. Bi, P., B. Shu, and P. Gong, *Crystal structure of the coxsackievirus A16 RNA-dependent RNA polymerase elongation complex reveals novel features in motif A dynamics.* Virol Sin, 2017. **32**(6): p. 548-552.
34. Nishio, M., et al., *Human parainfluenza virus type 2 L protein regions required for interaction with other viral proteins and mRNA capping.* J Virol, 2011. **85**(2): p. 725-32.
35. Chu, C., et al., *Structure of the guanylyltransferase domain of human mRNA capping enzyme.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10104-8.
36. Tao, Y., et al., *RNA synthesis in a cage--structural studies of reovirus polymerase lambda3.* Cell, 2002. **111**(5): p. 733-45.
37. Tekes, G., A.A. Rahmeh, and S.P. Whelan, *A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing.* PLoS Pathog, 2011. **7**(6): p. e1002073.
38. Ruedas, J.B. and J. Perrault, *Putative domain-domain interactions in the vesicular stomatitis virus L polymerase protein appendage region.* J Virol, 2014. **88**(24): p. 14458-66.

39. Zhou, Y., et al., *Structure and function of flavivirus NS5 methyltransferase.* J Virol, 2007. **81**(8): p. 3891-903.
40. Rahmeh, A.A., et al., *Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus.* Proceedings of the National Academy of Sciences, 2012. **109**(36): p. 14628--14633.
41. Green, T.J. and M. Luo, *Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P.* Proceedings of the National Academy of Sciences, 2009. **106**(28): p. 11713--11718.
42. Johansson, K., et al., *Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein.* J Biol Chem, 2003. **278**(45): p. 44567-73.
43. Morin, B., et al., *The polymerase of negative-stranded RNA viruses.* Current Opinion in Virology, 2013. **3**(2): p. 103--110.
44. Kingston, R.L., et al., *Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization.* J Mol Biol, 2008. **379**(4): p. 719-31.
45. Kingston, R.L., et al., *Structural basis for the attachment of a paramyxoviral polymerase to its template.* Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8301-6.
46. Bruhn, J.F., et al., *Crystal structure of the nipah virus phosphoprotein tetramerization domain.* J Virol, 2014. **88**(1): p. 758-62.
47. Cox, R., et al., *Structural and functional characterization of the mumps virus phosphoprotein.* J Virol, 2013. **87**(13): p. 7558-68.
48. Communie, G., et al., *Structure of the tetramerization domain of measles virus phosphoprotein.* J Virol, 2013. **87**(12): p. 7166-9.
49. Ding, H., et al., *Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus.* J Virol, 2006. **80**(6): p. 2808-14.
50. Tarbouriech, N., et al., *Tetrameric coiled coil domain of Sendai virus phosphoprotein.* Nat Struct Biol, 2000. **7**(9): p. 777-81.
51. Parks, G.D., *Mapping of a region of the paramyxovirus L protein required for the formation of a stable complex with the viral phosphoprotein P.* J Virol, 1994. **68**(8): p. 4862-72.
52. Cevik, B., et al., *The phosphoprotein (P) and L binding sites reside in the N-terminus of the L subunit of the measles virus RNA polymerase.* Virology, 2004. **327**(2): p. 297-306.
53. Horikami, S.M., et al., *An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex.* Virology, 1994. **205**(2): p. 540-5.
54. Holmes, D.E. and S.A. Moyer, *The phosphoprotein (P) binding site resides in the N terminus of the L polymerase subunit of sendai virus.* J Virol, 2002. **76**(6): p. 3078-83.
55. Malur, A.G., et al., *Role of a highly conserved NH(2)-terminal domain of the human parainfluenza virus type 3 RNA polymerase.* J Virol, 2002. **76**(16): p. 8101-9.
56. Chenik, M., et al., *Mapping the interacting domains between the rabies virus polymerase and phosphoprotein.* J Virol, 1998. **72**(3): p. 1925-30.
57. Smallwood, S., K.W. Ryan, and S.A. Moyer, *Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein.* Virology, 1994. **202**(1): p. 154-63.

58. Du Pont, V., Y. Jiang, and R.K. Plemper, *Bipartite interface of the measles virus phosphoprotein X domain with the large polymerase protein regulates viral polymerase dynamics.* PLoS Pathog, 2019. **15**(8): p. e1007995.
59. Shandilya, J. and S.G.E. Roberts, *The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling.* Biochimica et biophysica acta, 2012. **1819**(5): p. 391--400.
60. Thomas, M.C. and C.-M. Chiang, *The general transcription machinery and general cofactors.* Critical reviews in biochemistry and molecular biology, 2006. **41**(3): p. 105--178.
61. Soutourina, J., *Transcription regulation by the Mediator complex.* Nature Publishing Group, 2017. **34**(4): p. 1--13.
62. Jeronimo, C. and F. Robert, *The Mediator Complex: At the Nexus of RNA Polymerase II Transcription.* Trends in Cell Biology, 2017: p. 1--19.
63. Luse, D.S., *The RNA polymerase II preinitiation complex.* Transcription, 2013. **5**(1): p. e27050--10.
64. Nogales, E., R.K. Louder, and Y. He, *Structural Insights into the Eukaryotic Transcription Initiation Machinery.* Annual review of biophysics, 2017. **46**: p. 59--83.
65. Plaschka, C., et al., *Transcription initiation complex structures elucidate DNA opening.* Nature, 2016. **533**(7603): p. 353--358.
66. He, Y., et al., *Near-atomic resolution visualization of human transcription promoter opening.* Nature, 2016. **533**(7603): p. 359--365.
67. Schilbach, S., et al., *Structures of transcription pre-initiation complex with TFIIH and Mediator.* Nature Publishing Group, 2017. **551**(7679): p. 204--209.
68. Meinhart, A., et al., *A structural perspective of CTD function.* Genes Dev, 2005. **19**(12): p. 1401--1415.
69. Bernecky, C., et al., *Structure of transcribing mammalian RNA polymerase II.* Nature, 2016.
70. Armache, K.-J., et al., *Structures of complete RNA polymerase II and its subcomplex, Rpb4/7.* Journal of Biological Chemistry, 2005. **280**(8): p. 7131--7134.
71. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution.* Science, 2001. **292**(5523): p. 1863--1876.
72. Jeronimo, C., P. Collin, and F. Robert, *The RNA Polymerase II CTD: The increasing complexity of a low-complexity protein domain.* J Mol Biol, 2016. **428**(12): p. 2607--2622.
73. Napolitano, G., L. Lania, and B. Majello, *RNA Polymerase II CTD Modifications: How Many Tales From a Single Tail.* Journal of Cellular Physiology, 2014. **229**(5): p. 538--544.
74. Buratowski, S., *The CTD code.* Nature structural biology, 2003. **10**(9): p. 679--680.
75. Suh, H., et al., *Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II.* Molecular Cell, 2016. **61**(2): p. 297--304.
76. Schüller, R. and Forn, *Heptad-Specific Phosphorylation of RNA Polymerase~II CTD.* Molecular Cell, 2016. **61**(2): p. 305--314.
77. Schwer, B., et al., *Individual letters of the RNA polymerase II CTD code govern distinct gene expression programs in fission yeast.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(11): p. 4185--4190.

78. Heidemann, M., C. Hintermair, and Vo, *Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription.* Biochimica et biophysica acta, 2013. **1829**(1): p. 55--62.
79. Kim, M., et al., *Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7.* J Biol Chem, 2009. **284**(39): p. 26421--26426.
80. Srivastava, R. and S.H. Ahn, *Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function.* Biotechnology advances, 2015. **33**(6 Pt 1): p. 856--872.
81. Nilson, K.A., et al., *THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing.* Molecular Cell, 2015. **59**(4): p. 576--587.
82. Ghosh, A. and C.D. Lima, *Enzymology of RNA cap synthesis.* Wiley Interdisciplinary Reviews: RNA, 2010. **1**(1): p. 152--172.
83. Hengartner, C.J., et al., *Temporal Regulation of RNA Polymerase II by Srb10 and Kin28 Cyclin-Dependent Kinases.* Molecular Cell, 1998. **2**(1): p. 43--53.
84. Nair, D., Y. Kim, and L.C. Myers, *Mediator and TFIIH govern carboxyl-terminal domain-dependent transcription in yeast extracts.* Journal of Biological Chemistry, 2005. **280**(40): p. 33739--33748.
85. Esnault, C., et al., *Mediator-dependent recruitment of TFIIH modules in preinitiation complex.* Molecular Cell, 2008. **31**(3): p. 337--346.
86. Dotson, M.R., et al., *Structural organization of yeast and mammalian mediator complexes.* Proceedings of the National Academy of Sciences, 2000. **97**(26): p. 14307--14310.
87. Asturias, F.J., et al., *Conserved structures of mediator and RNA polymerase II holoenzyme.* Science, 1999. **283**(5404): p. 985--987.
88. Davis, J.A., et al., *Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction.* Molecular Cell, 2002. **10**(2): p. 409--415.
89. Naar, A.M., et al., *Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation.* Genes Dev, 2002. **16**(11): p. 1339--1344.
90. Taatjes, D.J., et al., *Structure, function, and activator-induced conformations of the CRSP coactivator.* Science, 2002. **295**(5557): p. 1058--1062.
91. Taatjes, D.J., T. Schneider-Poetsch, and R. Tjian, *Distinct conformational states of nuclear receptor-bound CRSP-Med complexes.* Nature structural \& molecular biology, 2004. **11**(7): p. 664--671.
92. Taatjes, D.J. and R. Tjian, *Structure and function of CRSP/Med2: a promoter-selective transcriptional coactivator complex.* Molecular Cell, 2004. **14**(5): p. 675--683.
93. Cai, G., et al., *Mediator structural conservation and implications for the regulation mechanism.* Structure (London, England : 1993), 2009. **17**(4): p. 559--567.
94. Bernecky, C., et al., *Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly.* PLoS Biol, 2011. **9**(3): p. e1000603.
95. Wang, X., et al., *Redefining the modular organization of the core Mediator complex.* Cell research, 2014. **24**(7): p. 796--808.
96. Tsai, K.-L., et al., *Subunit architecture and functional modular rearrangements of the transcriptional mediator complex.* Cell, 2014. **157**(6): p. 1430--1444.

97. Robinson, P.J.J., et al., *Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(44): p. 17931--17935.
98. Larivire, L., et al., *Structure of the Mediator head module.* Nature, 2012. **492**(7429): p. 448--+.
99. Imasaki, T., et al., *Architecture of the Mediator head module.* Nature, 2011. **475**(7355): p. 240--243.
100. Cai, G., et al., *Mediator head module structure and functional interactions.* Nature structural \& molecular biology, 2010. **17**(3): p. 273--279.
101. Koschubs, T., et al., *Identification, structure, and functional requirement of the Mediator submodule Med7N/31.* The EMBO journal, 2009. **28**(1): p. 69--80.
102. Larivire, L., et al., *Model of the Mediator middle module based on protein cross-linking.* Nucleic Acids Res, 2013. **41**(20): p. 9266--9273.
103. Baumli, S., S. Hoeppner, and P. Cramer, *A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer.* J Biol Chem, 2005. **280**(18): p. 18171--18178.
104. Thompson, C.M., et al., *A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast.* Cell, 1993. **73**(7): p. 1361--1375.
105. Nonet, M.L. and R.A. Young, *Intragenic and Extragenic Suppressors of Mutations in the Heptapeptide Repeat Domain of Saccharomyces-Cerevisiae Rna Polymerase-Ii.* Genetics, 1989. **123**(4): p. 715--724.
106. Robinson, P.J., et al., *Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex.* Cell, 2016. **166**(6): p. 1411--1422.e16.
107. Plaschka, C., et al., *Architecture of the RNA polymerase II-Mediator core initiation complex.* Nature, 2015. **518**(7539): p. 376-80.
108. Tsai, K.-L., et al., *Mediator structure and rearrangements required for holoenzyme formation.* Nature, 2017. **142**(7649): p. 749--201.
109. Cevher, M.A., et al., *Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit.* Nat Struct Mol Biol, 2014. **21**(12): p. 1028-34.
110. Han, S.J., et al., *Activator-specific requirement of yeast mediator proteins for RNA polymerase II transcriptional activation.* Mol Cell Biol, 1999. **19**(2): p. 979-88.
111. Park, J.M., et al., *In vivo requirement of activator-specific binding targets of mediator.* Molecular and Cellular Biology, 2000. **20**(23): p. 8709--8719.
112. Vojnic, E., et al., *Structure and VP16 binding of the Mediator Med25 activator interaction domain.* Nature structural \& molecular biology, 2011. **18**(4): p. 404--409.
113. Milbradt, A.G., et al., *Structure of the VP16 transactivator target in the Mediator.* Nature structural \& molecular biology, 2011. **18**(4): p. 410--415.
114. Bontems, F., et al., *NMR structure of the human Mediator MED25 ACID domain.* J Struct Biol, 2011. **174**(1): p. 245--251.
115. Plank, J.L. and A. Dean, *Enhancer function: mechanistic and genome-wide insights come together.* Molecular Cell, 2014. **55**(1): p. 5--14.
116. Bernecky, C. and D.J. Taatjes, *Activator-mediator binding stabilizes RNA polymerase II orientation within the human mediator-RNA polymerase II-TFIIF assembly.* J Mol Biol, 2012. **417**(5): p. 387-94.

117. Liu, Y., et al., *Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes.* Journal of Biological Chemistry, 2001. **276**(10): p. 7169--7175.
118. Hengartner, C.J., et al., *Association of an Activator with an Rna-Polymerase-Ii Holoenzyme.* Genes Dev, 1995. **9**(8): p. 897--910.
119. Borggrefe, T., et al., *A complex of the Srb8,-9,-10, and-11 transcriptional regulatory proteins from yeast.* J Biol Chem, 2002. **277**(46): p. 44202--44207.
120. Tsai, K.L., et al., *A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction.* Nat Struct Mol Biol, 2013. **20**(5): p. 611-9.
121. Jeronimo, C., et al., *Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In~Vivo.* Molecular Cell, 2016. **64**(3): p. 455--466.
122. Petrenko, N., et al., *Mediator Undergoes a Compositional Change during Transcriptional Activation.* Molecular Cell, 2016. **64**(3): p. 443--454.
123. Plaschka, C., K. Nozawa, and P. Cramer, *Mediator architecture and RNA polymerase II interaction.* J Mol Biol, 2016. **428**(12): p. 2569--2574.
124. El Khattabi, L., et al., *A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers.* Cell, 2019. **178**(5): p. 1145--1158.e20.
125. Luo, J., et al., *Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH.* Molecular Cell, 2015. **59**(5): p. 794--806.
126. He, Y., et al., *Structural visualization of key steps in human transcription initiation.* Nature, 2013. **495**(7442): p. 481--486.
127. Greber, B.J., et al., *The cryo-electron microscopy structure of the human CDK activating kinase. .* PNAS, 2020. **in press**.
128. Suloway, C., et al., *Automated molecular microscopy: the new Leginon system.* J Struct Biol, 2005. **151**(1): p. 41-60.
129. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* J Struct Biol, 2015. **192**(2): p. 216-21.
130. Voss, N.R., et al., *DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy.* J Struct Biol, 2009. **166**(2): p. 205-13.
131. Lander, G.C., et al., *Appion: an integrated, database-driven pipeline to facilitate EM image processing.* J Struct Biol, 2009. **166**(1): p. 95-102.
132. Ogura, T., K. Iwasaki, and C. Sato, *Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking.* J Struct Biol, 2003. **143**(3): p. 185-200.
133. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38-46.
134. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.
135. Zhang, K., *Gctf: Real-time CTF determination and correction.* J Struct Biol, 2016. **193**(1): p. 1-12.
136. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.
137. Kimanius, D., et al., *Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2.* Elife, 2016. **5**.

138. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-30.
139. Henderson, R., et al., *Outcome of the first electron microscopy validation task force meeting.* Structure, 2012. **20**(2): p. 205-14.
140. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
141. Afonine, P.V., et al., *Real-space refinement in PHENIX for cryo-EM and crystallography.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 531-544.
142. Goddard, T.D., et al., *UCSF ChimeraX: Meeting modern challenges in visualization and analysis.* Protein Sci, 2018. **27**(1): p. 14-25.
143. Bloyet, L.M., et al., *Regulation of measles virus gene expression by P protein coiled-coil properties.* Sci Adv, 2019. **5**(5): p. eaaw3702.
144. Suloway, C., et al., *Automated molecular microscopy: the new Leginon system.* J Struct Biol, 2005. **151**(1): p. 41--60.
145. Wang, F., et al., *General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM.* Proc Natl Acad Sci U S A, 2020. **117**(39): p. 24269-24273.
146. Mastronarde, D.N., *Automated electron microscope tomography using robust prediction of specimen movements.* J Struct Biol, 2005. **152**(1): p. 36-51.
147. Ogura, T., K. Iwasaki, and C. Sato, *Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking.* Journal of Structural Biology, 2003. **143**(3): p. 185--200.
148. Lander, G.C., et al., *Appion: An integrated, database-driven pipeline to facilitate EM image processing.* Journal of Structural Biology, 2009. **166**(1): p. 95--102.
149. Voss, N.R., et al., *DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy.* Journal of Structural Biology, 2009. **166**(2): p. 205--213.
150. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs*. 2015.
151. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38--46.
152. Zhang, K., *Gctf: Real-time CTF determination and correction.* J Struct Biol, 2016. **193**(1): p. 1--12.
153. Henderson, R., et al., *Outcome of the first electron microscopy validation task force meeting.* 2012: p. 205--214.
154. Sanchez-Garcia, R., et al., *DeepEMhancer: a deep learning solution for cryo-EM volume post-processing.* bioRxiv, 2020: p. 2020.06.12.148296.
155. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.
156. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
157. Miwa, K., et al., *Crystal Structure of Human General Transcription Factor TFIIE at Atomic Resolution.* J Mol Biol, 2016. **428**(21): p. 4258-4266.
158. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.* J Mol Biol, 2018. **430**(15): p. 2237-2243.

159. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Res, 2015. **43**(W1): p. W389-94.
160. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER.* Curr Protoc Bioinformatics, 2016. **54**: p. 5 6 1-5 6 37.
161. Lariviere, L., et al., *Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20.* Nat Struct Mol Biol, 2006. **13**(10): p. 895-901.
162. Croll, T.I., *ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 519-530.
163. Monte, D., et al., *Crystal structure of human Mediator subunit MED23.* Nat Commun, 2018. **9**(1): p. 3389.
164. Taschner, M., et al., *Structural basis of outer dynein arm intraflagellar transport by the transport adaptor protein ODA16 and the intraflagellar transport protein IFT46.* J Biol Chem, 2017. **292**(18): p. 7462-7473.
165. Yu, Y., et al., *Structural specializations of alpha(4)beta(7), an integrin that mediates rolling adhesion.* J Cell Biol, 2012. **196**(1): p. 131-46.
166. Kidmose, R.T., et al., *Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps.* IUCrJ, 2019. **6**(Pt 4): p. 526-531.
167. Greber, B.J., et al., *The cryoelectron microscopy structure of the human CDK-activating kinase.* Proc Natl Acad Sci U S A, 2020. **117**(37): p. 22849-22857.
168. Brown, N.R., et al., *The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases.* Nat Cell Biol, 1999. **1**(7): p. 438-43.
169. Greber, B.J., et al., *The complete structure of the human TFIIH core complex.* Elife, 2019. **8**.
170. Nakane, T., et al., *Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION.* eLife, 2018. **7**.
171. Louder, R.K., et al., *Structure of promoter-bound TFIID and model of human pre-initiation complex assembly.* Nature, 2016. **531**(7596): p. 604--609.
172. Greber, B.J., et al., *The cryo-electron microscopy structure of human transcription factor IIH.* Nature, 2017. **549**(7672): p. 414--417.
173. Takahashi, H., et al., *Human mediator subunit MED26 functions as a docking site for transcription elongation factors.* Cell, 2011. **146**(1): p. 92--104.
174. Belorusova, A.Y., et al., *Molecular determinants of MED1 interaction with the DNA bound VDR-RXR heterodimer.* Nucleic Acids Res, 2020. **48**(19): p. 11199-11213.
175. Hittelman, A.B., et al., *Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins.* EMBO J, 1999. **18**(19): p. 5380-8.
176. Malik, S., et al., *TRAP/SMCC/mediator-dependent transcriptional activation from DNA and chromatin templates by orphan nuclear receptor hepatocyte nuclear factor 4.* Mol Cell Biol, 2002. **22**(15): p. 5626-37.
177. Grontved, L., et al., *MED14 tethers mediator to the N-terminal domain of peroxisome proliferator-activated receptor gamma and is required for full transcriptional activity and adipogenesis.* Mol Cell Biol, 2010. **30**(9): p. 2155-69.
178. Meyer, K.D., et al., *p53 activates transcription by directing structural shifts in Mediator.* Nature structural \& molecular biology, 2010. **17**(6): p. 753--760.

179. Zhao, H., et al., *Structure of the Mammalian Mediator.* bioRxiv, 2020: p. 2020.10.05.326918.
180. Zhang, H., et al., *Mediator structure and conformation change.* Mol Cell, 2021.
181. Jasnovidova, O. and R. Stefl, *The CTD code of RNA polymerase II: a structural view.* Wiley Interdiscip Rev RNA, 2013. **4**(1): p. 1-16.
182. Kolakofsky, D., et al., *Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis.* Virology, 2004. **318**(2): p. 463-73.
183. Krumm, S.A., M. Takeda, and R.K. Plemper, *The measles virus nucleocapsid protein tail domain is dispensable for viral polymerase recruitment and activity.* J Biol Chem, 2013. **288**(41): p. 29943-53.
184. Paul, P.R., D. Chattopadhyay, and A.K. Banerjee, *The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype).* Virology, 1988. **166**(2): p. 350-7.
185. Tran, T.-L., et al., *The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein protomer.* Journal of General Virology, 2007. **88**(1): p. 196-206.
186. Slack, M.S. and A.J. Easton, *Characterization of the interaction of the human respiratory syncytial virus phosphoprotein and nucleocapsid protein using the two-hybrid system.* Virus Research, 1998. **55**(2): p. 167-176.
187. Sogaard, T.M. and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator.* J Biol Chem, 2007. **282**(19): p. 14113-20.