Structural Characterization of Native RNA Polymerase II Transcription Complexes and Nucleosomes in *Drosophila melanogaster*

Natalie L. Venette-Smith[1, 2, 3, 4, 5], Rishi K. Vishwakarma[1, 3, 4, 5, 7], Varun Venkatakrishnan[1, 5, 6], Roberta Dollinger[1, 2, 8], Josie Schultz [1, 2, 9], Paul Babitzke[1, 3, 5], Ganesh Anand[1, 5, 6], David S. Gilmour[1, 2, 5], Jean-Paul Armache[1, 2, 4, 5], Katsuhiko S. Murakami[1, 3, 4, 5]

[1] Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA 16802
[2] Center for Eukaryotic Gene Regulation, Penn State University, University Park, PA 16802
[3] Center for RNA Molecular Biology, Penn State University, University Park, PA 16802
[4] Center for Structural Biology, Penn State University, University Park, PA 16802
[5] Huck Institutes of Life Sciences, Penn State University, University Park, PA 16802
[6] Department of Chemistry, Penn State University, University Park, PA 16802
[7] Current address: Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105
[8] Current address: Department of Cell and Regenerative Biology, Wisconsin Blood Cancer Research Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI 53792
[9] Current address: Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD 21701

* Corresponding authors. Email: jza449@psu.edu and kum14@psu.edu

Abstract

Structural studies of eukaryotic RNA polymerase II (Pol II) transcription complexes often depend on *in vitro* assembly by mixing purified Pol II with synthetic DNA/RNA scaffolds, recombinant transcription factors, and/or histones, followed by stalling transcription at defined positions by adding selected nucleotide triphosphate substrates. These studies have yielded remarkable results for understanding nucleosome transcription by Pol II with elongation factors but may fail to represent transcription in native conditions. To investigate Pol II transcription within metazoan cells, we developed an approach to isolate the native transcription complexes from *Drosophila melanogaster* embryos. Utilizing one-step FLAG-tag affinity purification and mild chromatin treatment with Micrococcal Nuclease (MNase), we preserved the native transcription complex for cryo-EM and proteomics studies. *In silico* purification through the cryo-EM classifications determined structures of multiple forms of native transcription complex, nucleosome and other macromolecules. Remarkably, we determined the structures of metazoan Rpb4/Rpb7 stalk-less elongation complex as well as hexameric nucleosome lacking an H2A/H2B dimer, revealing that diverse elongation complexes and nucleosomes are involved in active transcription *in vivo*. Nucleosome is positioned only downstream of Pol II in the nucleosome elongation complex, underscoring it as a major energy barrier and a time-consuming step during Pol II progression through nucleosomal DNA. Proteomics identified co-purified factors responsible for initiation and elongation stages of transcription, as well as RNA modification factors. This study lays the groundwork for structural study of native transcription in eukaryotes, with future work focused on studies of transient and minor populations of transcription complexes.

Introduction

In all eukaryotes except for plants, three DNA-dependent RNA polymerases can be found in the nucleus: RNA polymerase I, II and III, abbreviated as Pol I, II and III, respectively. Pol II is responsible for transcribing messenger RNA (mRNA), long noncoding RNA (lncRNA) and a number of small RNAs (1). While Pol II alone has the capacity to synthesize RNA from a DNA template, binding of various factors allows it to proceed through different stages of transcription (initiation, elongation and termination), acquire diverse functions and provide opportunities for regulation.

Transcription initiation involves the assembly of Pol II with general transcription factors (GTFs) and the Mediator complex, followed by CDK7-mediated phosphorylation of the Serine-5 residues in the C-terminal domain (CTD) heptapeptide repeats of Pol II (2). Binding the Pol II with negative elongation factor (NELF) and Spt4/Spt5 (as known as DSIF) establishes promoter proximal pausing, a post initiation regulatory sub-step. Active transcription begins after the removal of NELF from Pol II, and bindings of elongation factors such as TFIIS, elongation factor 1 (ELOF1; aka Elf1), Spt6 and the PAF complex (3). Furthermore, histone chaperones such as facilitates chromatin transcription (FACT) to enhance Pol II transcription through the nucleosome DNA. When Pol II reaches the poly-A termination sequence, a series of factors terminate transcription by releasing RNA and Pol II from DNA.

Structural biology has been one of the major ways of probing the transcriptional machinery. It reveals how different factors interact with Pol II to elucidate their functions. The first high-resolution structure of Pol II was obtained by X-ray crystallography in 2000 (4). Since then, over 200 structures of Pol II complexes have been determined by X-ray crystallography or cryo-electron microscopy (cryo-EM) (5). Currently, cryo-EM is the primary method for structural biology, as it does not require crystallization, does not require a large quantity of sample, and allows visualization of heterogeneity and conformational dynamics of targeted macromolecule. This makes it well-suited for determining the structures of transcriptional complexes containing multiple factors as well as elusive intermediates during nucleosomal DNA transcription (6, 7).

In many of these studies, the transcription complexes are assembled *in vitro* with purified Pol II, synthetic DNA/RNA as well as recombinant transcription factors and/or histones, and some cases, adding limited nucleotide triphosphate (NTP) substrate to move Pol II on DNA template at defined locations. Recent studies of the nucleosome elongation complex (EC) have provided snapshots of Pol II moving through nucleosomal DNA, showing disassembly of a nucleosome at downstream DNA followed by reassembly at upstream DNA, either in the absence (8, 9) or in the presence of the histone chaperone (10). However, the insights from these studies are derived from experimental conditions specifically designed to capture transient intermediates, leaving it uncertain whether they faithfully represent native transcription processes within cells. Recently, cryo-electron tomography (cryo-ET) in combination with cryo-Focused Ion-beam Milling (cryo-FIB) is employed to visualize thinned flash-frozen biological samples in their near-native environment (11). This *in situ* approach allows for atomic-resolution structural characterization of

gigantic macromolecules such as ribosomes (~4.5 MDa molecular weight) as they can be easily located in crowded cellular tomograms followed by averaging in 3D to improve resolution (12). The technique is in continuous development, however, and since Pol II (~500 kDa molecular weight) is ~9 times smaller than ribosome, identifying their positions and orientations within tomograms remains a technical issue. This leads to the question: How does one characterize native Pol II transcription complexes?

To address this gap, we establish a method for visualizing endogenous transcription complexes directly isolated from *Drosophila melanogaster* (Drosophila) embryos. We used large population of CRISPR/Cas9-modified Drosophila containing a double FLAG-tag on Rbp1-CTD, to obtain sufficient Pol II complexes for cryo-EM and proteomics studies. Although the affinity purified Pol II complexes remain heterogenous and contain impurities, *in silico* purification of targeted macromolecules through cryo-EM data processing allowed us to determine multiple forms of transcriptional complexes, nucleosomes and other macromolecules co-purified with Pol II.

RESULTS

Isolation of native Pol II transcription complexes from FLAG-tagged Rpb1 Drosophila embryos

In this study, we used a double FLAG-tagged Rpb1 Drosophila strain for isolating native Pol II transcription complex (13). Using populations of ~10,000 adults, we collected ~75 g of 0-12 hr embryos for nuclei preparation **(Fig. 1A)**. Next, we used MNase, an endo- and exo-nuclease that preferentially targets the linker DNA between nucleosomes (14), to digest chromatin while preserving nucleosome bound with Pol II and other factors. MNase digestion at 0.5 U/µL for 10 mins at room temperature produced DNA fragments of ~150 bp and ~250 bp, corresponding to mono- and di-nucleosomes, respectively (**Fig. 1B**). Nuclei lysis followed by removing insoluble materials, Pol II transcription complex was isolated by using one-step FLAG-affinity purification. The presence of Pol II in elution fractions was confirmed by colloidal Coomassie-stained SDS-PAGE analysis, which showed well-defined bands of Rpb1, Rbp2 and Rpb3 subunits, with impurities (**Fig. 1C**). Western blot analysis confirmed presence of Rpb3 and histone H3 in the elution fractions (**Fig. 1D**). Quality and quantity of the Pol II transcription complex were estimated by negative stain EM.

In silico purification of native Pol II and other macromolecules isolated from Drosophila embryos

Cryo-EM data were collected using a Talos Arctica 200 kV microscope equipped with a Falcon 4 detector (**SI Appendix Table 1**). Despite the heterogeneity of the affinity-purified Pol II complex containing several impurities (**Fig. 1C, SI Appendix Figs 1A**), computational particle picking and

2D/3D classifications during cryo-EM data processing enabled us to enrich particles belonging to each targeted macromolecule, a process we term "in silico purification." From 30,103 micrographs, we extracted ~15 million particles, which underwent multiple rounds of 2D and 3D classifications, yielding subsets of macromolecules (**Fig. 2; SI Appendix Figs. 1-2**). These subsets include Pol II, nucleosome, a putative U1 spliceosome, and an octahedral density attributed to the dihydrolipoyllysine-residue succinyltransferase (DLST) component of the 2-oxoglutarate dehydrogenase complex. Further 3D classification separated 876,764 Pol II particles into four major subunits including: 1) EC with stalk; 2) EC without stalk; 3) nucleic acid-free apo-form and 4) nucleosome EC.

Cryo-EM structural analysis of the native Drosophila Pol II elongation complex

The cryo-EM map of the apo-form Pol II was solved at a lower resolution (4.24 Å) and was visibly anisotropic (**SI Appendix Figs. 1C and 3A**) due to preferred particle orientation. Thus, we could not reliably construct its model. In contrast, the EC with and without the Rpb4/7 stalk show well-defined isotropic cryo-EM densities (**Fig. 3A, SI Appendix Figs. 1D and E, SI Appendix Table 2, SI Movie 1**). The cryo-EM map shows Pol II, 10 bases of DNA/RNA hybrid and 25 bases of double-stranded downstream DNA. The 3' end of the RNA is in the post-translocated site (i site) and nucleotide binding site (i+1 site) is vacant (**Fig. 3B**). The EC shows no sign of RNA backtracking (15), nor a tilted state of the DNA/RNA hybrid (16). The cryo-EM density for DNA/RNA hybrid is strong and traceable until 10 bases from the 3' end of RNA, which provides the first direct evidence showing that the native Pol II EC accommodates 10 bases of DNA/RNA hybrid. The lid, rudder (Rpb1) and flap (Rpb2) form the RNA exit channel to guide nascent RNA toward outside of Pol II but cryo-EM density of single-strand RNA beyond DNA/RNA hybrid is not resolved (**Fig. 3B**). The density corresponding to upstream DNA is weak but traceable, however, single-stranded non-template DNA within a transcription bubble is not visible (**Fig. 3A**). Since DNA and RNA sequences within the EC are mixed due to *in situ* Pol II transcription at different locations in genome, we modeled DNA and RNA with the following sequences: template DNA as poly(dT); non-template DNA as poly(dA), and the nascent RNA as poly(A). The overall structure of Pol II is nearly identical to the structures of other eukaryotic Pol II, including those from yeast (17, 18) and mammal (19, 20) (**Fig. 3C**).

The cryo-EM density of the Rpb4/Rpb7 stalk of Pol II was less well-defined compare with the main body of Pol II (see Pol II-overall in **Fig. 2, SI Appendix Fig. 1B**), suggesting its mobility and/or sub-stoichiometric occupancy. Focused 3D variability analysis (3DVA) in cryoSPARC revealed two subsets in which stalk was either fully present (EC with stalk, **Fig. 3A** top; **SI Appendix Fig. 1D**) and completely absent (EC without stalk, **Fig. 3A** bottom; **SI Appendix Fig. 1E**), and these cryo-EM maps are determined as overall resolutions of 3.44 and 3.37 Å, respectively. Not only the EC with stalk, but also EC without stalk show strong densities of DNA and RNA, indicating that the stalk-less Pol II is actively participated in transcription within cells. The numbers of particles corresponding to the EC with and without the stalk are about the same

(**Fig. 2; SI Appendix Table 2**) indicating that a significant proportion of the stalk-less Pol II participates in gene expression in metazoan cells.

Cryo-EM analysis of the native *Drosophila* nucleosome

The in silico purification also enriched a large pool of nucleosome particles and yielded subsets including: 1) octameric nucleosome (called "nucleosome" throughout this manuscript); 2) hexameric nucleosome (hexasome) and 3) di-nucleosome (**Fig. 2**). We obtained a 3.29 Å reconstruction of a nucleosome (**SI Appendix Fig. 2A; SI Appendix Table 3**), which shows an asymmetric arrangement of entry and exit sites DNA, where one site is ~11 bases shorter compared to the other (**Figs. 4 A,B**). Throughout the manuscript, we refer to the shorter and longer DNA strands outside from nucleosome core as the entry and exit sites, respectively. This high-resolution reconstruction of nucleosome visualizes features such as side-chain densities and the DNA phosphate backbone (**SI Appendix Fig. 4A)**. As these nucleosomes contain a mix of native DNA sequences, the base assignment was not possible, thus we modeled them using a 601 Widom sequence (21). These nucleosomes were very similar in structure to recombinant octameric nucleosomes (22, 23).

We also obtained a 6.33 Å reconstruction of a hexasome **(Figs. 4C,D; SI Appendix Fig. 2B; SI Appendix Table 3**), which lacks a single copy of the H2A/H2B histone dimer as compared with nucleosome (**Fig. 4E**). The quality of the hexasome reconstruction allowed for unambiguous positioning of histone subunits and the DNA path around the hexasome core. Approximately 100 base pairs of DNA wrapping around the hexasome (**Fig. 4D**), which is ~50 base pairs shorter than the ~150 base pairs of DNA wrapping around the nucleosome (**Fig. 4B**). In the hexasome, the DNA contacting with histone core starts at the superhelical location SHL(-3), continues through SHL(0), and ends at SHL(7) at the exit site. Due to the absence of one H2A/H2B dimer in the hexasome, the DNA prior to the SHL(-3) site unwraps from histone core, thus the DNA at the entry and exit sites point in opposite directions (**Fig. 4E**). These observations are consistent with previous characterizations of hexasome by structural and genome wide nucleosome profiling (24-27).

Intriguingly, we also obtained a subset representing di-nucleosome, in which two nucleosomes are partially stacked in a stair-wise manner, while also being connected by the linker DNA (**Fig. 4F; SI Appendix Fig. 2C**). A similar arrangement of di-nucleosome in recombinant experiments has been previously observed through X-ray crystallography (28) and cryo-EM (29) and was also reported in complex with a chromatin remodeling enzyme, CHD1 (26), wherein the two nucleosomes were bridged by the remodeler.

Cryo-EM structure of Drosophila Nucleosome Elongation Complex

After confirming the presence of both Pol II and nucleosome in the elution fractions of FLAG-affinity purification (**Fig. 1D**) and through the cryo-EM data processing to obtain reconstructions

of the Pol II and nucleosome, we employed further *in silico* purification from these subsets to enrich the nucleosome EC particles (**Fig. 2**). In particular, careful analyses of the Pol II and nucleosome 2D classes revealed some showing extra density beyond their well-defined densities. Using particles belonging to these 2D classes, we created an ab initio cryo-EM map as a reference of the nucleosome EC followed by carrying out the 3D classification of particles belonging to Pol II and nucleosome to enrich particles belonging to nucleosome EC. The isolated nucleosome EC was well-defined and could be refined to an overall resolution of 7.79 Å, limited by the Pol II and nucleosomes moving each other. To address this, we refined them independently (focused/local refinement), resulting in a stalk-less Pol II and nucleosome at resolutions of 4.31 Å and 6.12 Å, respectively **(SI Appendix Fig. 3F).** We combined these maps to create a composite map of nucleosome EC for a model building and refinement (**Fig. 5; SI Appendix Table 2**). The cryo-EM map clearly shows template DNA and endogenous RNA in the Pol II and downstream DNA extending from the DNA binding channel of Pol II wraps around histones in nucleosome. The downstream nucleosome has clear DNA features such as sugar-phosphate backbones, and major and minor grooves, as well as an identifiable the histone core (**SI Movie 2**). Notably, the nucleosome EC is observed in one state only, with Pol II located at the nucleosome position SHL(-5). About 20 bases DNA unwrapped from the nucleosome for Pol II moving toward it, but maintain the DNA interacting with H2A/H2B. The jaw (Rpb1), protrusion (Rpb2) and jaw (Rpb9) domains of Pol II face to nucleosome but there is no physical contact between Pol II and nucleosome (**Fig. 5C**). The cryo-EM structure of native nucleosome EC determined in this study, which is derived from Pol II transcriptions starting from promoters, is similar to one of previously reported nucleosome EC structures that were assembled and prepared in vitro (8) (PDB: 6A5P, **SI Appendix Figs. 3 D, E)**. It is likely that other previously reported states of the *in vitro* reconstituted nucleosome ECs are transient and represent minor states during the Pol II transcription through nucleosome DNA *in vivo*.

Mass spectrometry identification of proteins co-purified with Pol II

To identify factors co-purified with Pol II, we analyzed them by mass spectroscopy. Trypsin digested samples were analyzed by bottom-up liquid chromatograph-tandem mass spectrometry (LC-MS/MS), a technique commonly applied to identify proteins from macromolecular assemblies (30). Peptide search against the entire *D. melanogaster* genome was used for protein identification with searches for post-translational modifications (PTM) such as phosphorylation on Serine, Threonine and Tyrosine, and lysine methylation. We identified a total of 28 proteins participating to the Pol II transcription, nucleosome formation, chromatin remodeling and RNA processing, and other proteins of interest with a Sequest HT Score greater than 10 (**Table 1**).

While 10 of the 12 subunits of Pol II could be identified, only Rpb 1, 2, 3, 5, 8, 9 subunits met the cut-off criteria of Score Sequest HT greater than 10 (**Table 1**). Nucleosomes were confirmed due to the presence of histones H2A, H2A variant, and H4. Three subunits of mediator (MED 1, 15, 17) and subunit 1 of TFIIF were identified, as well as DSIF (including Spt4 and Spt5

subunits) and metazoan-specific Pol II binding factor Gdown1 were identified (31). Although its score was just below the arbitrary cut-off, the presence of NELF was also confirmed (**SI Dataset 1**). The proteomics study suggests that an endogenous promoter-proximal pausing complex including EC, DSIF and NELF can possibly be isolated and enriched by the affinity and *in silico* purifications for future cryo-EM characterization. Most factors identified by mass spectrometry are involved in transcriptional elongation underscoring the most Pol II participate in the active RNA synthesis. The histone chaperone FACT (Ssrp and Spt16) (10, 32) was also identified, although it was not found in the cryo-EM sturdy of nucleosome EC (**Fig. 2**). Other proteins of interest are DNA topoisomerase 1 and 2, RNA processing factors, chromatin/chromosome factors and the U1 snRNP spliceosome (U1 snRNP and snRNP A subunits) (**Fig. 2, SI Appendix Fig. 2D; Table 1**).

DISCUSSION

In this study, we used a FLAG-tagged Rpb1-CTD of Pol II to isolate native transcriptional complex from Drosophila embryos. The high-yield Drosophila population system allowed us to harvest sufficient embryos for isolating endogenous Pol II complex for cryo-EM and proteomics. In silico purification facilitated the isolation, enrichment, and construction of 3D cryo-EM maps of individual macromolecules, achieving separations not possible with other methods. For examples, distinguishing Pol II with or without the stalk, and nucleosome or hexasome, are challenging using chromatography purification due to their chemical and molecular weight similarities. While affinity tagging at specific site can achieve some separations, in silico purification offers a more straightforward approach without requiring genome editing. This method is powerful and convenient for studying heterogeneous and dynamic transcription processes in a cellular context. In addition, all transcription complexes characterized here originated from native promoters, enabling the visualization of endogenous, rather than artificially assembled, complexes.

The Rpb4-Rpb7 heterodimeric stalk binds around the RNA exit channel of Pol II. Rpb7 is essential in budding yeast *S. cerevisiae*, but Rpb4 is only essential under stress conditions and during starvation for retaining viability (33, 34). Rpb4 is essential in fission yeast *S. pombe* (35), *Drosophila* (36) and for zebrafish development (37). The stalk functions in closing the DNA binding clamp (17, 18), providing binding platform of numerous transcription factors (38-41) and for forming Pol II dimer (42, 43). The stalk also functions in RNA transport and translation in the cytoplasm (44). In *S. cerevisiae*, several studies suggested that the stalk can dissociate from the EC (45), but the stalk separation from the transcribing Pol II remains controversial in other studies (46). One challenge was to observe the stalk-less Pol II actively participating transcription *in vivo*. In this study, *in silico* purification allowed us to isolate it (**Fig. 2**), giving us a direct evidence of the stalk-less Pol II involved in the active transcription. This finding aligns with chromosomal immunostaining of *Drosophila* third instar larvae, showing that not all transcriptionally-active Pol II co-localize with Rpb4, as seen on polytene chromosome (36).

Numerous transcription factors associate with Pol II in each stage of transcription (38, 39). The cryo-EM analysis of *in vitro* reconstituted transcription complexes revealed how these factors influence the Pol II activity and providing functions to Pol II for carrying out stage specific activities. However, we did not observe Pol II particles unambiguously associated with any factors, despite their identification by proteomics study. Through further data processing, we identified small subsets of EC showing extra densities around the Spt5 and Spt6 binding sites (**SI Appendix Figs. 3 F,G**). However, they resulted in low resolution reconstructions and thus could not be unequivocally identified. Perhaps, most Pol II engaging in expressing transcription during rapid cell cycle embryogenesis in Drosophila do not require elongation factors (Spt4/5, Spt6) and/or histone chaperones (e.g., FACT) due to lower densities of nucleosomes at highly expressing genes (47) and coupling between DNA replication and transcription (48).

We also isolated native metazoan nucleosomes including nucleosome, hexasome, and di-nucleosome (**Fig. 4**). Nucleosomes co-eluted with Pol II from FLAG-affinity column may result from their abundance in the nuclear extract or association with apo-form Pol II, as shown previously (49). Additionally, we co-purified spliceosomes and DLST, a mitochondrial enzyme. Spliceosomes may associate with Pol II to couple transcription and splicing (50). DLST, an enzyme localized in mitochondria, may be isolated because of contamination. However, it was reported that a small percentage of DLST remains in the nucleus to provide succinyl-CoA to histone acetyl/succinyltransferase, KAT2A, a component of the eukaryotic transcription coactivator Spt-Ada-Gcn5-Acetyltransferase (SAGA) (51).

Affinity-purified Pol II from Drosophila embryos enabled cryo-EM and proteomic studies, revealing diverse Pol II and nucleosome complexes actively engaged in transcription. Our findings likely represent the most stable and abundant transcription complexes due to the lack of cross-linking prior cryo-EM specimen preparation. Crossing FLAG-tagged Rpb1 lines with those encoding tags for GTFs, elongation factors, or histones could enrich minor Pol II complexes for future studies. CRISPR/Cas9 could further expand this system for isolating diverse complexes.

The Drosophila transcription system has uncovered key mechanisms in eukaryotic gene regulation, such as promoter-proximal pausing (52). While differences in transcriptional activity between Drosophila and humans exist, the similarities in transcription factors and the advantages of the Drosophila genetic system make it a powerful model for studying transcription. This approach can also be extended to other developmental stages, offering a versatile model for structural and proteomic analyses.

Experimental Procedures

Growth and collection of *Drosophila* embryos

A *D. melanogaster* strain (Bloomington Drosophila Stock Center number 99015) expressing Rpb1 with a double FLAG-tag at the C-terminus was used in this study (53). The sequences expressing the double FLAG-tagged Rpb1 were introduced into the endogenous Rpb1 gene in Lu and Gilmour

(13). Drosophila populations of ~10,000 adults were housed in a 30 cm x 60 cm plastic cylinder pipe for 10 days. Starting on day 3, 0-12 hr embryos were collected on grape juice/agar plates and stored for a maximum 3 days at 4 °C. Embryos were strained through 2 layers of mesh to remove any adult flies, washed with water, dechorionated by mixing with 50% bleach for 30 sec and washed with deionized water before being stored at -80 °C. These steps were repeated until enough embryos were collected for the subsequent experiments.

Generation of nuclear extract

We prepared nuclear extract using a published procedure with modifications (54). 75 g of partially thawed, dechorionated embryos were resuspended in 120 mL Buffer A (1 M sucrose, 10 mM HEPES-HCl (pH = 7.5), 4 mM $MgCl_2$, 0.1 mM EGTA, 1 mM β-mecaptoethanol (BME), 1 mM sodium bisulfite, 0.2 mM PMSF, 1 mM benzamidine HCl, and protease inhibitors (1.6 µg/mL benzamidine-HCl, 1 µg/mL aprotinin, 1 µg/mL pepstatin A, and 2 µg/mL leupeptin), homogenized with Potter Elvehjem tissue homogenizer, strained through a layer of MiraCloth, divided evenly into 6 Sorvall tubes, and centrifuged at 1,200 rpm at 4 °C for 10 mins to pellet debris. Supernatant was transferred to fresh Sorvall tubes and centrifuged at 6,000 rpm at 4 °C for 20 mins. The supernatant was aspirated off the nuclei pellets, taking care to remove all the fatty layer. Nuclei were resuspended in 5 mL buffer A, and dounced in a 7 mL homogenizer with first a loose pestle then a tighter pestle. Homogenized nuclei were poured over 10 mL of high sucrose buffer (1.75 M sucrose, 10 mM HEPES-HCl (pH = 7.5), 2 mM $MgCl_2$, 0.1 mM EGTA, 1 mM BME, and protease inhibitors), mixed gently to generate a rough sucrose gradient and prevent nuclei aggregation at the interphase, and centrifuged at 12,500 rpm at 4 °C for 30 mins. The supernatant was aspirated off the nuclei pellets, taking care to remove all the fatty layer. Nuclei were resuspended in 1.5 mL of nuclear resuspension buffer (0.3 M sucrose, 10 mM HEPES-HCl (pH = 7.5), 2 mM $MgCl_2$, 0.1 mM EGTA, 0.1 M NaCl, 1 mM DTT, 2 mM $CaCl_2$, and protease inhibitors) and combined in a 15 mL Falcon tube before undergoing MNase Digestion.

Chromatin digestion was performed with 0.5 U/µL concentration MNase (#LS004797, Worthington Biochemical Corporation) at room temperature for 10 mins. As per the manufacturer's definition, a single unit of MNase is the amount of enzyme necessary to change the optical density at 260 nm by 1.0 when incubated at 37 °C, pH 8.0 and DNA used as substrate. The reaction was quenched by adding EGTA to a final concentration of 15 mM and placed on ice. Triton X-100 was added for a final concentration of 0.5 % and incubated on ice for 10 mins before nuclei were mechanically lysed by passing sample through a 21-gauge needle 4 times. Lysed nuclei were gently mixed at 4 °C for 20 mins and then centrifuged at 16,000x g at 4 °C for 10 mins to separate insoluble and soluble material. The supernatant was transferred into clean 1.5 mL microfuge tubes, flash frozen in liquid nitrogen and then stored at -80 °C.

Purification of RNA Polymerase II Complex

Nuclear lysate was thawed and centrifuged at 16,000x g at 4 °C for 10 mins to remove any additional insoluble material before overnight dialysis at 4 °C into FLAG purification buffer (10

mM HEPES-HCl (pH = 7.5), 150 mM NaCl, 5% glycerol, 1 mM EDTA). Post-dialysis nuclear lysate was centrifuged at 16,000x g at 4 °C for 10 mins again to remove any remaining insoluble material. Protease inhibitors and 1/1000th Proteoloc Protease Inhibitor (100X Expedeon Proteoloc Protease Inhibitor Cocktail - Expedeon 42516) were added directly to sample before affinity purification. All steps were performed at 4 °C, unless stated otherwise.

Sepharose 4B resin (Sigma-Aldrich 4B1200-100ML) at a ratio of 100 µL slurry per 1 mL nuclear lysate was prepared with two 5 min washes with FLAG purification buffer (supplemented with 1/1000x concentration protease inhibitors) before batch binding with sample for 10 mins. Supernatant was separated by centrifugation at 2,500 x g at 4 °C for 2.5 mins. Supernatant was transferred to anti-FLAG resin (Sigma-Aldrich ANTI-FLAG M2 Affinity Gel #A2220) at a ratio of 60 µL slurry per 1 mL nuclear lysate, which was prepared as described above, and gently mixed for batch binding for 3 hours. Resin slurry was transferred to a column and allowed to settle before the flow-through passed. The column was washed with FLAG purification buffer (supplemented with 1/1000th protease inhibitors) with a 5 min incubation period. Pol II complexes were eluted from the column using elution buffer (10 mM HEPES-HCl (pH = 7.5), 150 mM NaCl, 5% glycerol, 1 mM EDTA, 350 µg/mL 3x FLAG peptide, 1/1000th protease inhibitor) in 1 column volume (CV) fractions. Elution buffer was incubated on column before collection: fractions 1-2 at 30 mins, fractions 3-4 at 20 mins and fractions 5-8 at 10 mins.

SDS-PAGE followed by Western blot analysis using rabbit anti-Rbp3 (serum made in-house as described in Lu and Gilmour (53)) and rabbit anti-histone H3 (abcam #AB1791) antibodies confirmed presence of both Pol II and nucleosomes in the elution fractions. Alexa-488 fluorescent secondary antibodies (Invitrogen #A11008) and Amersham Typhoon Fluorescent Image Analyzer (GE Healthcare Bio-sciences) were used for Western blot analysis. Elution fractions containing both Pol II and nucleosomes were combined, dialyzed overnight in Flag Purification Buffer to remove any remaining 3x Flag Peptide, and concentrated by filtration (Amicon Ultra 0.5 mL 10000 NMWL UF501096). Qualitative assessment of the Pol II transcription complex was determined by negative stain (0.75% Uranyl Formate) before proceeding with vitrification.

Cryo-electron microscopy data collection and in silico purification of targeted macromolecules

Cryo-EM grids were prepared as follows. Quantifoil Au 2/2 grids (200 mesh, 2-µm hole size) were glow-discharged for 10 s at 15 mA in a PelCo easiGlow system (PelCo). 3.5 µl of the sample (1.8 mg/mL) was applied on the grid surface, and the grids were blotted and plunge-frozen in liquid ethane maintained at liquid-nitrogen temperatures, using a Vitrobot IV (FEI) maintained at 4 °C and 100% humidity. The data was recorded on a Talos Arctica G2 system at Penn State Huck Cryo-EM Core Facility, operated at 200 kV in nanoprobe mode, using a Falcon 4 camera at a nominal magnification of 150,000x corresponding to a calibrated physical pixel size of 0.944 Å. The collection resulted in 30,103 images, each fractionated into 42 frames, accumulating 50 e^- per $Å^2$ total dose (0.0825 s per frame, 1.25 $e^-/Å^2$/frame).

Raw movies were processed by cryoSPARC version 4.4.1 (55). We used *"Patch Motion Correction"*, followed by *"Patch CTF Estimation (Multi)"* to preprocess the images. 18,270,461 particles were picked by *"Blob picker"* and extracted 14,859,077 particles in a 440x440 box Fourier-cropped to 96x96. *"2D classification"* into 250 classes revealed multiple forms of macromolecules as seen on the raw images (**SI Appendix Fig. 1A**). The 2D classes representing Pol II, nucleosomes, spliceosome and DLST were selected independently and used to generate initial 3D templates using *"Ab initio"* reconstruction. We also selected 2D classes representing ice and other contaminants and used them to generate *ab initio* 3D templates attracting particles of impurities. These initial 3D templates were used to enrich particles into independent subsets (*in silico* purification) by using *"Heterogeneous Refinement"*. We then optimized the quality and resolution of the individual subsets and quantified their heterogeneity using a combination of *"Non-Uniform Refinement"* (56), *"2D classification"*, *"Heterogeneous Refinement"* and *"3D Variability"*. All final reconstructions were calculated using particles in box 440x440 (0.944 Å/pixel) Fourier cropped to 360x360 (corresponding to 1.1538 Å/pixel), following a gold-standard refinement by *"Non-Uniform Refinement"* and are reported at Fourier Shell Correlation = 0.143 (57) (**SI Appendix Table 2,3**).

Refinement and classification of Pol II particle subset

After refining the initial Pol II particles, we obtained an anisotropic reconstruction due to particle orientation bias. Using "*Heterogeneous Refinement"* in combination with *"2D classification"* and selection, we reduced bias and generated an isotropic reconstruction. Further optimization through signal-based selection using CryoSieve (58) yielded the final resolution of 2.86 Å. Upon careful examination of the data, we found that the Rpb4/7 stalk was not well-resolved compare with Pol II containing 10 subunits. We performed a series of global and local classifications using "*Heterogeneous Refinement"* and *"3D Variance"*, resulting in three distinct Pol II reconstructions including: 1) apo-form, 2) EC with the stalk and 3) EC without the stalk. The individual subsets of EC with and without stalk were further refined, yielding 3.44 Å and 3.37 Å final resolution reconstructions, respectively (**Fig 2; Appendix Figs. 1B,D,E; Appendix Table 2**).

Refinement and classification of octameric nucleosome particles

We refined the octameric nucleosome subset using *"Non-Uniform Refinement"*. The reconstruction indicated that one side of the nucleosomal DNA, when displayed at different thresholds, disappears earlier than the other side. To investigate this further, we used *"Heterogeneous Refinement"* to classify the nucleosome particles into two subsets, varying in DNA length on one side. The highest quality nucleosome reconstruction was obtained in C1 symmetry at 3.34 Å, which was then subjected to signal-based selection using CryoSieve, yielding a reconstruction at 3.29 Å (**Fig. 2; SI Appendix Fig. 2A; SI Appendix Table 3)**.

Refinement and classification of hexameric nucleosome particles

Hexameric nucleosomes were refined using *"Homogeneous Refinement"*, and the corresponding particles were reextracted at 1.1538 Å/pixel and re-refined using *"Non-Uniform Refinement"*. The data was further processed and homogenized using *"Heterogeneous Refinement"*. The final reconstruction was refined to 6.33 Å using *"Non-Uniform Refinement"* (**Fig 2; SI Appendix Fig. 2B; SI Appendix Table 3)**.

Refinement and classification of stacked nucleosome particles

Stacked nucleosomes were refined using *"Homogeneous Refinement"*, and the corresponding particles were reextracted at 1.1538 Å/pixel and re-refined using *"Non-Uniform Refinement"*. The data was further processed and homogenized using selection through 2D classification and *"Heterogeneous Refinement"*. The final reconstruction was obtained at 9.80 Å using *"Non-Uniform Refinement"* (**Fig 2; SI Appendix Fig. 2C)**.

Refinement and classification of nucleosome EC particles

A subset representing nucleosome EC was generated through iterative 2D classification and selection for enriching nucleosome EC particles from the Pol II and nucleosome subsets. We then used cryoSPARC's *"Ab initio"* reconstruction to generate a 3D template, which was used *"Heterogeneous Refinement"* of particles from the Pol II and nucleosome subsets. Resulting particles were further analyzed using 2D classification to eliminate impurities, such as ice, and refined using *"Homogeneous Refinement"*. We then re-extracted the particles at 1.1538 Å/pixel and refined the data using *"Non-Uniform Refinement"* to obtain the final overall reconstruction of 7.79 Å. In this map, Pol II and nucleosome were not defined to high quality; therefore, we refined them individually in focused masks using *"Local Refine"*, resulting in the final reconstructions at 4.31 and 6.12 Å, respectively. For modeling and refinement of the nucleosome EC structure, we combined these maps to generate a composite map using Phenix **(Fig 2; SI Appendix Fig. 1F; SI Appendix Table 2)**.

Refinement and classification of putative spliceosome particles

Particles corresponding to the putative spliceosome in our data were refined using *"Homogeneous Refinement"*, reextracted at 1.1538 Å/pixel, and re-refined. We performed extensive 2D classification analysis, which further confirmed that this subset suffered from a severe orientation bias, which drastically limited the quality of the final reconstruction (**Fig. 2; SI Appendix Fig. 2D)**. CryoSPARC reported the final resolution using particles at 1.1538 Å/pixel at 7.00 Å, however the features of this reconstruction do not reflect this quality.

Refinement and classification of DLST particles

Throughout our processing, particles exhibiting a potential in-plane 4-fold -symmetry were found using 2D classification in every reported subset. We selected and pooled all the particles together. We then ran 2D classification on all these particles to further eliminate any contamination. Selected

particles were used to generate ab initio 3D templates. The refinement against a selected 3D Coulomb potential density was first run in C1-symmetry. We then reextracted the particles at 1.1538 Å/pixel and re-refined the data using *"Non-Uniform Refinement"* in C1, C2, C4, D4 and O symmetries. The octahedral symmetry yielded a reconstruction at 3.66 Å, with clear secondary structure features (**Fig. 2; SI Appendix Figs. 2E, 5; SI Appendix Table 3)**. To identify the macromolecule, we used ModelAngelo (59) using the *Drosophila* proteome database (Uniprot: UP000000803), which suggested that our Coulomb potential density map represented dihydrolipoyllysine-residue succinyltransferase component of mitochondrial 2-oxoglutarate dehydrogenase complex. To further confirm the assignment, we used AlphaFold, as well as an existing structure from *H. sapiens* (PDB: 6H05, (60)).

Model building

Pol II EC

A model of the Pol II EC was constructed by assembling AlphaFold 2-modeled (61) Pol II subunits (from Rpb1 to 12) using *Sus scrofa* Pol II (PDB: 6GML, (16)) as a guide. EC subunits were manually fitted into the cryo-EM density map using UCSF ChimeraX v1.8 (62), followed by real-space refinements using Coot (63) and Phenix (*phenix.real_space_refine*)(64). Modeling errors were corrected by DAQ refine (65) (**SI Appendix Tables 2, 4)**.

Nucleosomes and hexasomes

The model of the *Drosophila* octameric nucleosomes were constructed using histone protein models from AlphaFold 2 with the 601 Widom DNA sequence used as a reference (PDB: 3LZ0, (23)) which was extended and adjusted to the Coulomb potential density map. This model was then used to construct the hexameric nucleosome. Both models were then refined using *phenix.real_space_refine* **(SI Appendix Tables 3,4).**

Nucleosome EC

To model the nucleosome EC, we manually placed the *Drosophila* EC and octameric nucleosome models into the composite map of nucleosome EC describe above. Using Coot, we further optimized the local fit and connected the DNA. We then fixed the model using *phenix.real_space_refine* **(SI Appendix Tables 2, 4)**.

DLST

The model for DLST was constructed into the 3.66 Å Coulomb potential density map. First, we used an AlphaFold 2 model and rigid-body fit it into one of the subunits. This monomeric model was then adjusted to the density and used to create a 24-subunit model. This model was refined through *phenix.real_space_refine* using "*Non-crystallographic symmetry*" restraints (**SI Appendix Tables 3, 4; SI Appendix Fig. 4B,C)**.

Protein Identification through Mass Spectrometry

The protein identity was confirmed by the Huck Proteomics and Mass Spectrometry Core Facility (RRID: SCR_024462). The purified endogenous Pol II complex sample was prepared by reduction-alkylation of cysteines followed by a standard trypsin digestion (MS-grade trypsin, Pierce, Thermo Fisher Scientific, USA, MAN0011821). The sample was then subjected to liquid chromatography tandem mass spectrometry (LC-MS/MS) in an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific). Data analysis was performed using Proteome Discoverer v2.5 against the *Drosophila* proteome (Uniprot: UP000000803) with default settings using Sequest HT search with fixed modifications set to carbamidomethylation of Cysteines and oxidation of Methionine and posttranslational modifications set to Serine, Threonine and Tyrosine phosphorylation, and Lysine methylation.

Competing interests

The authors declare no competing interests.

Author Contributions

Fly embryo preparation (NLVS, RD, JS, DSG), Pol II preparation (NLVS, RD, JS, DSG), proteomics experiment (VV, GA), cryo-EM data collection and processing (NLVS, RKV, JPA and KSM), model building and refinement (JPA and KSM), manuscript writing (NLVS, RKV, VV, RD, PB, GA, DSG, JPA, KSM).

Funding

Funding: National Institutes of Health grants R35 GM131860 and R21AI168948 (to K.S.M.), R01 GM047477 (to D.S.G.), and R01 GM098399 (to P.B.). This project is funded, in part, under a Grant with the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions.

Acknowledgements

We thank Dr. Sung Hyun Cho at the Penn State Huck Cryo-Electron Microscopy Facility for assistance with cryo-EM grid preparation and operating microscope for cryo-EM grid screening and cryo-EM data collection, and we thank Dr. Tatiana Laremore at Penn State Huck Proteomics and Mass Spectrometry Facility for technical assistance in protein identification. We thank Olivia Abboud for *D. melanogaster* maintenance. Research reported in this publication was supported by the Office of the Director, NIH, under award number S10OD026822-01 (S.L.H.). The results generated in this paper were computed on a hardware generously provided to us through the NVIDIA Academic Grant by NVIDIA Corporation. Figure 1A was created with BioRender.com.

Data and materials availability

All data needed to evaluate the conclusions of this paper are present in the Main Text and/or the Supplementary Materials. The cryo-EM maps and the refined models were deposited in the Electron Microscopy DataBank (EMDB, www.ebi.ac.uk/emdb/) and Protein Data Bank (PDB,

www.rcsb.org), respectively. Pol II EC 2.86 Å overall: EMD-48624; Pol II-EC with stalk: EMD-48622 and 9MU7; Pol II EC without stalk: EMD-48623 and 9MU8; nucleosome: EMD-48619 and 9MU4; hexasome: EMD-48620 and 9MU5; nucleosome EC: EMD-48625 (overall), EMD-48626 and 9MU9 (composite); DLST: EMD-48621 and 9MU6.

Supporting information

This article contains supporting information.

References

1. S. R. Archuleta, J. A. Goodrich, J. F. Kugel, Mechanisms and Functions of the RNA Polymerase II General Transcription Machinery during the Transcription Cycle. *Biomolecules* **14** (2024).
2. S. Malik, R. G. Roeder, Regulation of the RNA polymerase II pre-initiation complex by its associated coactivators. *Nat Rev Genet* **24**, 767-782 (2023).
3. J. Li, D. S. Gilmour, Promoter proximal pausing and the control of gene expression. *Curr Opin Genet Dev* **21**, 231-235 (2011).
4. P. Cramer *et al.*, Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640-649 (2000).
5. S. Osman, P. Cramer, Structural Biology of RNA Polymerase II Transcription: 20 Years On. *Annu Rev Cell Dev Biol* **36**, 1-34 (2020).
6. A. A. Mohamed, R. Vazquez Nunez, S. M. Vos, Structural advances in transcription elongation. *Curr Opin Struct Biol* **75**, 102422 (2022).
7. S. I. Sekine, H. Ehara, T. Kujirai, H. Kurumizaka, Structural perspectives on transcription in chromatin. *Trends Cell Biol* **34**, 211-224 (2024).
8. T. Kujirai *et al.*, Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595-598 (2018).
9. M. Filipovski, J. H. M. Soffers, S. M. Vos, L. Farnung, Structural basis of nucleosome retention during transcription elongation. *Science* **376**, 1313-1316 (2022).
10. H. Ehara, T. Kujirai, M. Shirouzu, H. Kurumizaka, S. I. Sekine, Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. *Science* **377**, eabp9466 (2022).
11. A. Doerr, Cryo-electron tomography. *Nature Methods* **14**, 34-34 (2017).
12. W. Jin, Y. Zhou, A. Bartesaghi, Accurate size-based protein localization from cryo-ET tomograms. *J Struct Biol X* **10**, 100104 (2024).
13. F. Lu, D. S. Gilmour, Genetic analysis of the RNA polymerase II CTD in Drosophila. *Methods* **159-160**, 129-137 (2019).
14. M. Noll, Subunit structure of chromatin. *Nature* **251**, 249-251 (1974).
15. A. C. Cheung, P. Cramer, Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249-253 (2011).
16. S. M. Vos, L. Farnung, H. Urlaub, P. Cramer, Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **560**, 601-606 (2018).
17. D. A. Bushnell, R. D. Kornberg, Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**, 6969-6973 (2003).
18. K. J. Armache, S. Mitterweger, A. Meinhart, P. Cramer, Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280**, 7131-7134 (2005).
19. Y. Chen *et al.*, Allosteric transcription stimulation by RNA polymerase II super elongation complex. *Mol Cell* **81**, 3386-3399 e3310 (2021).
20. Y. He *et al.*, Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-365 (2016).
21. P. T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19-42 (1998).

22. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260 (1997).
23. D. Vasudevan, E. Y. D. Chua, C. A. Davey, Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1-10 (2010).
24. Y. Arimura, H. Tachiwana, T. Oda, M. Sato, H. Kurumizaka, Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry* **51**, 3302-3309 (2012).
25. J. Yang, K. Yen, S. Mahony, Size-based expectation maximization for characterizing nucleosome positions and subtypes. *Genome Res* **34**, 1334-1343 (2024).
26. M. Engeholm *et al.*, Resolution of transcription-induced hexasome-nucleosome complexes by Chd1 and FACT. *Mol Cell* **84**, 3423-3437 e3428 (2024).
27. H. Wu *et al.*, Reorientation of INO80 on hexasomes reveals basis for mechanistic versatility. *Science* **381**, 319-324 (2023).
28. D. Kato *et al.*, Crystal structure of the overlapping dinucleosome composed of hexasome and octasome. *Science* **356**, 205-208 (2017).
29. M. Shimizu *et al.*, Asymmetric fluctuation of overlapping dinucleosome studied by cryoelectron microscopy and small-angle X-ray scattering. *PNAS Nexus* **3**, pgae484 (2024).
30. A. I. Nesvizhskii, O. Vitek, R. Aebersold, Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787-797 (2007).
31. X. Hu *et al.*, A Mediator-responsive form of metazoan RNA polymerase II. *Proc Natl Acad Sci U S A* **103**, 9506-9511 (2006).
32. G. Orphanides, G. LeRoy, C. H. Chang, D. S. Luse, D. Reinberg, FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105-116 (1998).
33. M. Choder, R. A. Young, A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol Cell Biol* **13**, 6984-6991 (1993).
34. K. McKune, K. L. Richards, A. M. Edwards, R. A. Young, N. A. Woychik, RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. *Yeast* **9**, 295-299 (1993).
35. H. Sakurai, H. Mitsuzawa, M. Kimura, A. Ishihama, The Rpb4 subunit of fission yeast Schizosaccharomyces pombe RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Mol Cell Biol* **19**, 7511-7518 (1999).
36. T. Pankotai, Z. Ujfaludi, E. Vamos, K. Suri, I. M. Boros, The dissociable RPB4 subunit of RNA Pol II has vital functions in Drosophila. *Mol Genet Genomics* **283**, 89-97 (2010).
37. M. Maeta *et al.*, RNA polymerase II subunit D is essential for zebrafish development. *Sci Rep* **10**, 13213 (2020).
38. Y. Chen, P. Cramer, RNA polymerase II elongation factors use conserved regulatory mechanisms. *Curr Opin Struct Biol* **84**, 102766 (2024).
39. E. Nogales, R. K. Louder, Y. He, Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* **46**, 59-83 (2017).
40. V. Babbarwal, J. Fu, J. C. Reese, The Rpb4/7 module of RNA polymerase II is required for carbon catabolite repressor protein 4-negative on TATA (Ccr4-not) complex to promote elongation. *J Biol Chem* **289**, 33125-33130 (2014).
41. Y. Zeng, H. W. Zhang, X. X. Wu, Y. Zhang, Structural basis of exoribonuclease-mediated mRNA transcription termination. *Nature* **628**, 887-893 (2024).

42. S. Aibara, C. Dienemann, P. Cramer, Structure of an inactive RNA polymerase II dimer. *Nucleic Acids Res* **49**, 10747-10755 (2021).
43. M. Carminati, J. B. Rodriguez-Molina, M. C. Manav, D. Bellini, L. A. Passmore, A direct interaction between CPF and RNA Pol II links RNA 3' end processing to transcription. *Mol Cell* **83**, 4461-4478 e4413 (2023).
44. M. Farago, T. Nahari, C. Hammel, C. N. Cole, M. Choder, Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol Biol Cell* **14**, 2744-2755 (2003).
45. M. Choder, Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem Sci* **29**, 674-681 (2004).
46. D. Schulz, N. Pirkl, E. Lehmann, P. Cramer, Rpb4 subunit functions mainly in mRNA synthesis by RNA polymerase II. *J Biol Chem* **289**, 17446-17452 (2014).
47. S. Ercan, M. J. Carrozza, J. L. Workman, Global nucleosome distribution and the regulation of transcription in yeast. *Genome Biol* **5**, 243 (2004).
48. C. Y. Cho, J. P. Kemp, Jr., R. J. Duronio, P. H. O'Farrell, Coordinating transcription and replication to mitigate their conflicts in early Drosophila embryos. *Cell Rep* **41**, 111507 (2022).
49. B. W. Baer, D. Rhodes, Eukaryotic RNA polymerase II binds to nucleosome cores from transcribed genes. *Nature* **301**, 482-488 (1983).
50. S. Zhang *et al.*, Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* **371**, 305-309 (2021).
51. Y. Wang *et al.*, KAT2A coupled with the alpha-KGDH complex acts as a histone H3 succinyltransferase. *Nature* **552**, 273-277 (2017).
52. D. S. Gilmour, J. T. Lis, RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Mol Cell Biol* **6**, 3984-3989 (1986).
53. F. Lu, B. Portz, D. S. Gilmour, The C-Terminal Domain of RNA Polymerase II Is a Multivalent Targeting Sequence that Supports Drosophila Development with Only Consensus Heptads. *Mol Cell* **73**, 1232-1242 e1234 (2019).
54. D. W. Miller, S. C. Elgin, Transcription of heat shock loci of Drosophila in a nuclear system. *Biochemistry* **20**, 5033-5042 (1981).
55. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
56. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221 (2020).
57. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745 (2003).
58. J. Zhu *et al.*, A minority of final stacks yields superior amplitude in single-particle cryo-EM. *Nat Commun* **14**, 7822 (2023).
59. K. Jamali *et al.*, Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450-457 (2024).
60. B. Nagy *et al.*, Structure of the dihydrolipoamide succinyltransferase (E2) component of the human alpha-ketoglutarate dehydrogenase complex (hKGDHc) revealed by cryo-EM and cross-linking mass spectrometry: Implications for the overall hKGDHc structure. *Biochim Biophys Acta Gen Subj* **1865**, 129889 (2021).

61. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
62. E. C. Meng *et al.*, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **32**, e4792 (2023).
63. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
64. P. V. Afonine *et al.*, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).
65. G. Terashi, X. Wang, D. Kihara, Protein model refinement for cryo-EM maps using AlphaFold2 and the DAQ score. *Acta Crystallogr D Struct Biol* **79**, 10-21 (2023).

List of Figures and Tables

Figure 1. Methodology and purification of native Pol II transcription complex from Drosophila embryo
(A) Diagram showing the processes of Drosophila embryo collection, MNase treatment and nuclear extract preparation, purification of Pol II transcription complexes and characterization, and vitrification for cryo-EM data collection. **(B)** Chromatin DNA digested by MNase. DNA fragments from mono- and di-nucleosomes are indicated in red box. Lanes: (1,2) 2 U/µL MNase digestion, (3,4) 1 U/µL MNase digestion; (5,6) 0.5 U/µL MNase digestion, (7,8) 0.2 U/µL MNase digestion, (1,3,5,7) insoluble fractions after MNase digestion, (2,4,6,8) soluble fractions after MNase digestion. **(C)** Coommassie stained SDS-PAGE protein gel of affinity-purified samples. Lanes: (1,2) input, (3) Sepharose bead flow-through, (4) FLAG bead flow-through, (5) FLAG bead wash, (6-13) FLAG bead elution. Subunits of Pol II are indicted. **(D)** Western blot detection of Pol II by anti-Rpb3 antibody (top) and nucleosomes by anti-histone H3 antibody (bottom). Lanes are labeled as in panel **(C)**.

Figure 2. In silico purification of the Pol II transcription complexes and nucleosomes
A simplified scheme of *in silico* purification of the macromolecules by cryo-EM data processing (for details, see ***Materials and Methods***). The preprocessed macromolecule particles were subjected to 3D classification, yielding six unique subsets, representing Pol II (gray), putative U1 spliceosome (orange), DLST (dark gray), stacked nucleosomes (pink), hexasome (green) and nucleosome (blue). The first row represents the initial subsets that were further processed to obtain the final reconstructions (the second row). Pol II particles were classified further into three subsets, containing: EC with Rpb4/7 stalk, EC without Rpb4/7 stalk, and apo-form Pol II. Further individual analysis of Pol II and nucleosome subsets yielded enriched particles of nucleosome EC (middle, violet). Pol II and nucleosome within the nucleosome EC were refined by focused/local refinement, and combine these maps to generate a composite map of nucleosome EC.

Figure 3. Cryo-EM structure of the Pol II EC.
(A) Orthogonal views of the cryo-EM density map of the Drosophila EC with stalk (top) and without stalk (bottom). Pol II subunits (Rpb1-12), DNA and RNA are colored and labeled (dDNA, downstream DNA; uDNA, upstream DNA). Several Pol II domains are also indicated. **(B)** Cryo-EM density map (transparent cyan) of DNA and RNA in the EC is shown with DNA/RNA (stick model, tDNA, template DNA; ntDNA, non-template DNA) and motifs of Pol II (cartoon model, BH, bridge helix) in the refined EC. NTP binding site of the active site is indicated. **(C)** Comparison of the structures of the Pol II within Drosophila EC (green, main body; cyan, stalk) and mammalian EC (black) (PDB: 6GML, (16)). Structures were superimposed via main body of Pol II with an RMSD of 0.795 Å.

Figure 4. Cryo-EM structures of nucleosome, hexasome and di-nucleosome.
Cryo-EM density maps of Drosophila nucleosome **(A)** and hexasome **(C)**. Super Helical Location (SHL) are marked starting from SHL(0) at the dyad axis of nucleosome. Refined models of the nucleosome **(B)** and hexasome **(D)** are shown using the color scheme from (A) and (C). **(E) (left)** A nucleosome model (dark blue) is superimposed in the cryo-EM density map of the hexasome. A H2A/H2B dimer of nucleosome, which is missing in hexasome, is indicated and colored (light

yellow and light red); (**right**) Comparative schematic representation of the nucleosome and hexasome. Positions of the entry and exit sites DNA and SHL are labeled. **(F)** Cryo-EM density map of an overall reconstruction of the di-nucleosomes.

Figure 5. Cryo-EM structure of the native nucleosome EC.
(A, B) Orthogonal views of the cryo-EM densities of the Drosophila nucleosome EC paused at SHL(-5) position of nuclesomal DNA. Pol II (gray), DNA, RNA and histones are colored. **(B)** Orthogonal views of the cryo-EM densities of nucleosome EC highlighting the nucleosome core, DNA and RNA with transparent Pol II density (white). (**C**) A magnified view of the interface between Pol II and nucleosome. Domains of Poll II and SHL positions are labeled.

Table 1. Proteins identified in purified endogenous Drosophila Pol II through liquid chromatography tandem mass spectrometry. In gray, Pol II subunits below the selected Score Sequest HT threshold of 10.

Supplementary Figure Legends:

SI Appendix Figure 1. Cryo-EM micrograph and analysis.
(A) Representative cryo-EM micrographs showing heterogeneity of the affinity-purified Pol II complex. Examples of different types of particles are marked: polymerases (green circle), nucleosomes (blue circles), nucleosome EC (pink circle), DLST (orange square) **(B-F)** Statistics on final reconstructions of apo-form Pol II (**B**) and ECs (**C-F**) including selected 2D class averages (left), Euler angle distribution of particles used in final reconstructions (second to left), local resolution maps with color ranging from blue to red representing high to low resolutions (second from right), and FSC curves calculated in CryoSPARC reported at the 0.143 FSC cutoff (right). In the Euler angle plot, the length of the cylinders is correlated to the distribution of particles in that direction. In the local resolution visualization, the maps are colored locally according to the scale above.

SI Appendix Figure 2. Cryo-EM micrograph and analysis, continued.
(A-E) Statistics on final reconstructions of nucleosome (**A**), hexasome (**B**), di-nucleosomes (**C**), putative spliceosome (**D**) and DLST (**E**) including selected 2D class averages (left), Euler angle distribution of particles used in final reconstructions (second to left), local resolution maps with color ranging from blue to red representing high to low resolutions (second from right), and FSC curves calculated in CryoSPARC reported at the 0.143 FSC cutoff (right). In the Euler angle plot, the length of the cylinders is correlated to the distribution of particles in that direction. In the local resolution visualization, the maps are colored locally according to the scale above.

SI Appendix Figure 3. Variability of cryo-EM densities of native EC.
(A-C) Cryo-EM densities of the apo-form Pol II (**A**), EC with the stalk **(B)** and EC without the stalk **(C)** with each subunit colored as described in **Figure 3. (D-E)** Comparison of nucleosome ECs from *Drosophila* (green, this study) and yeast paused at SHL(–5) (red, PDB: 6A5P) (8) with structures aligned on Pol II **(D)** and the nucleosome **(E)**. **(F)** Ribbon model of mammalian Pol II with elongation factor DSIF in red and histone chaperone SPT6 in blue (PBD: 6GMH (16)). **(G)** 3D classification of the Drosophila Pol II EC data focused on the DSIF or Spt6 binding sites reveal

low resolution densities showing that some binding sites are occupied (red, DSIF; blue, Spt6), which may indicate co-purification of DSIF and Spt6 with endogenous Pol II. Mass spectrometry detected DSIF and Spt6 in high abundance.

SI Appendix Figure 4. Quality of cryo-EM of nucleosome reconstruction; Cryo-EM map and model of DLST.
(A) Selected representative 3.29 Å resolution cryo-EM densities of core histones and DNA in nucleosome superimposed on the refined models. **(B)** Two views of the cryo-EM density of *Drosophila* dihydrolipoyllysine-residue succinyltransferase component of mitochondrial 2-oxoglutarate dehydrogenase complex at 3.66 Å resolution, with individual subunits colored separately. **(C)** Two views of the refined model of DLST with individual subunits colored separately.

SI Appendix Table 1. Cryo-EM data collection, interpretation and visualization details.

SI Appendix Table 2. Data processing and model refinement statistics for EC and nucleosome EC.

SI Appendix Table 3. Data processing and model refinement statistics for nucleosome, hexasome and DLST.

SI Appendix Table 4. Protein sequences used for model building.

SI Appendix Dataset 1. Protein identification by mass spectrometry.

SI Appendix Movie 1. Cryo-EM density map of EC with and without Rpb4/7 stalk.
Overview of the cryo-EM maps of the EC with and without the stalk. Related to Fig. 3A

SI Appendix Movie 2. Cryo-EM density map of nucleosome EC.
Overview of the cryo-EM composite map of the nucleosome EC. Related to Fig. 5.

Table 1. Proteins identified in the FLAG affinity-purified samples by liquid chromatography tandem mass spectrometry.

In gray, Pol II subunits below the selected Score Sequest HT threshold of 10.

	Uniprot ID	Gene Name	Description	Coverage [%]	# Peptides	# PSMs	MW [kDa]	Score Sequest HT	Phosphorylation (S; T; Y)	Methylation (K)
RNA Polymerase II subunits	P04052	Polr2A	DNA-directed RNA polymerase II subunit RPB1	33	50	106	209	255.91	S1130; T1132	K902; K906; K1125; K1127
	A0A0B4LI86	Polr2B	DNA-directed RNA polymerase II subunit RPB2	38	38	92	134	219.95		K131
	O97183	Polr2C	DNA-directed RNA polymerase II subunit RPB3	40	7	20	31.2	52.39		
	Q9VEA5	Polr2D	DNA-directed RNA polymerase II subunit RPB4	19	2	2	16.2	5.53		
	Q7JZF5	Polr2E	DNA-directed RNA polymerase II subunit RPB5	27	5	12	24.5	29.11	S113; S117	K124
	Q9VFB5	Polr2G	DNA-directed RNA polymerase II subunit RPB7	13	2	3	19.1	6.24		
	Q9VNZ3	Polr2H	DNA-directed RNA polymerases II subunit RPB8	35	4	12	17.2	20.66		
	P36958	Polr2I	DNA-directed RNA polymerase II subunit RPB9	60	6	15	15.1	37.6		
	Q9VJE4	Polr2J	DNA-directed RNA polymerase II subunit RPB11	9	1	1	13.4	2.13		
	Q9VC49	Polr2L	DNA-directed RNA polymerases II subunit RPB10	18	1	1	7.7	2.59		
Histon	Q4AB57	His2A	Histone H2A	60	5	19	13.4	39.76		
	A0A0B4KH25	His2Av	Histone H2A variant	27	4	16	15	33.87		
	P84040	His4	Histone H4	57	8	28	11.4	48.81	Y89	K32
Transcription Factors	Q9VP05	MED1	Mediator of RNA polymerase II transcription subunit 1	6	6	6	149.4	16.15	T396; S399; S400; Y513; S667	
	Q9Y149	MED15	Mediator of RNA polymerase II transcription subunit 15	10	4	6	80.5	12.92	S521; T531; T538; T642; S644	
	Q9VEC1	MED17	Mediator of RNA polymerase II transcription subunit 17	17	5	7	71.5	14.12	S558; S559; S564; S567	
	Q05913	TfIIFalpha	General transcription factor IIF subunit 1	26	8	11	64.4	28	Y513	K196; K364; K365; K525
	Q9TVQ5	Spt4	Transcription elongation factor SPT4	30	2	4	13.3	11.95		
	Q9V460	Spt5	Transcription elongation factor SPT5	27	22	39	119.4	107.02		
	N0D8I3	Spt6	Transcription elongation factor spt6	22	26	37	208.5	95.28	S124; S335	K332
	Q05344	Ssrp	FACT complex subunit Ssrp1	35	20	54	81.5	117.2	S443; S628	K550; K554; K556
	Q8IRG6	dre4	FACT complex subunit spt16	26	24	57	123.5	118.78	S994	K423; K424; K910
	Q9VDY6	Polr2M	Gdown1	20	7	8	41.7	19.08		
Other	P15348	Top2	DNA topoisomerase 2	12	14	18	164.3	49.34		
	A0A0B4K6K1	Rm62	RNA helicase	16	6	10	52.5	24.51		
	Q494M1	Top1	DNA topoisomerase I	9	7	9	111.8	20.38	T815; T821	K820
	P25171	Rcc1	Regulator of chromosome condensation	12	4	6	58.8	12.64	T286; T290	K289
	Q9V9T4	Acf	ATP-dependent chromatin assembly factor large subunit	4	4	4	170.3	11.24	Y519; Y640	
	Q9VW52	Prp3	Precursor RNA processing 3, isoform A	10	3	5	67.3	10.69	S23; S25; S62; S64; S65; T110	K18; K69; K92; K115
	Q9VEN2	cal1	Chromosome alignment defect 1	9	4	6	109.5	10.53	S488; S634; T637; T667; S681; T872; S888	K469; K470; K658; K672; K679; K898

	M9PC89	snRNP-U1-70K	U1 small nuclear ribonucleoprotein 70 kDa	8	3	6	52.9	19.69		
	P43332	snf	U1 small nuclear ribonucleoprotein A	15	2	4	24.5	11.07		

Supplementary Info Appendix

SI Appendix Table 1. Cryo-EM data collection, interpretation and visualization details

Data Collection	
Microscope	Talos Arctica
Voltage (kV)	200
Camera	Falcon 4
Magnification	150,000
Electron exposure ($e^-/Å^2$)	50.65
Defocus range (μm)	-0.8 to -2.0
Pixel size (Å)	0.944
Number of Frames	42
Number of Images	31,103
Images/Hole	2
Data Collection Software	EPU v. 3.4
Data Processing	
Motion Correction	CryoSPARC v 4.3.1 *Patch Motion Correction (Multi)*
Data Processing	CryoSPARC v 4.3.1 – 4.5.1; Relion v 5.0
Data Interpretation	
Initial Models	AlphaFold 2; AlphaFold 3
Rigid Body Fit	UCSF ChimeraX v.1.8; Coot v 0.9.8
Model Building	Coot v 0.9.8
Model Refinement	Coot v 0.9.8 Real Space Refine; phenix.real_space_refine
Data Visualization	UCSF ChimeraX v.1.8
Figure Preparation	Biorender; Apple Keynote; MS PowerPoint; Adobe Illustrator

SI Appendix Table 2. Data processing and model refinement statistics for EC and Nucleosome EC

Refinements	**Pol II overall** (EMD-48624)	**EC with stalk** (EMD-48622) (PDB: 9MU7)	**EC without stalk** (EMD-48623) (PDB: 9MU8)	**Nucleosome EC** ** (EMD-48625) (EMD-48626) (PDB: 9MU9**)
Initial # of particles	876,764	876,764	876,764	36,552
Final # of particles	68,716	70,116	96,409	36,552
Map resolution (Å) at 0.143 FSC threshold	2.86	3.44	3.37	7.79/4.31/6.12*
Reference models (PDB code)	-	6GML	6GML	6GML/3LZ0
Model composition Non-hydrogen atoms Protein residues Nucleic acid residues Ligands	-	 33,887 3,992 99 Zn:8	 31,485 3,689 99 Zn:8	 42,423 4,446 347 Zn:8
B factors (Å^2) Protein DNA/RNA Ligand	-	 13.6/491.5/117.2 39.0/536.2/171.1 127.7/181.2/156.5	 57.2/403.6/112.2 39.0/536.2/171.1 127.7/181.2/156.5	 30.0/1229.4/457.3 200.1/1104.7/352.9 286.1/944.0/570.4
r.m.s. deviations Bond lengths (Å) Bond angles (°)	-	 0.006 0.754	 0.006 0.746	 0.003 0.728
Validation MolProbity score Clash score Q-score Rotamer outliers (%)	-	 1.48 4.54 0.393 0.00	 1.48 4.57 0.403 0.00	 1.48 5.07 0.202 0.00
Ramachandran plot Favored (%) Allowed (%) Disallowed (%)	-	 96.29 3.71 0.00	 96.26 3.74 0.00	 96.61 3.39 0.02

* left, overall resolution; middle, focused on Pol II; right, focused on nucleosome
** validation performed against the nucleosome EC composite map

SI Appendix Table 3. Data processing and model refinement statistics for nucleosome, hexasome and DLST

Refinements	Nucleosome (EMD-48619) (PDB: 9MU4)	Hexasome (EMD-48620) (PDB: 9MU5)	DLST (EMD-48621) (PDB: 9MU6)
Initial # of particles	1,848,096	231,425	70,106
Final # of particles	66,988	38,738	23,543
Map resolution (Å) at 0.143 FSC threshold	3.29	6.33	3.66
Reference models (PDB code)	3LZ0	Octameric nucleosome	AlphaFold
Model composition			
Non-hydrogen atoms	12,794	9,666	43,134
Protein residues	770	532	5,616
Nucleic acid residues	328	266	-
Ligands	-	-	-
B factors (Å^2) (mean)			
Protein	23.6/168.3/67.5	124.0/1398.6/616.9	40.6/144.90/78.1
DNA	15.7/472.4/123.3	188.7/1043.1/444.0	-
Ligand	-	-	-
r.m.s. deviations			
Bond lengths (Å)	0.006	0.004	0.005
Bond angles (°)	0.837	0.722	0.983
Validation			
MolProbity score	1.28	1.45	1.86
Clash score	2.50	5.89	7.51
Q-score	0.442	0.206	0.39
Rotamer outliers (%)	0.00	0.00	0.00
Ramachandran plot			
Favored (%)	96.29	97.31	93.09
Allowed (%)	3.71	2.69	6.91
Disallowed (%)	0.00	0.00	0.00

SI Appendix Table 4. Protein sequences used for model building

Protein Name	Gene Name	Uniprot ID
DNA-directed RNA polymerase II subunit RPB1	RPB1	P04052
DNA-directed RNA polymerase II subunit RPB2	RPB2	P08266
DNA-directed RNA polymerase II subunit RPB3	RPB3	O97183
DNA-directed RNA polymerase II subunit RPB4	RPB4	Q9VEA5
DNA-directed RNA polymerase II subunit RPB5	RPB5	Q7JZF5
DNA-directed RNA polymerase II subunit RPB6	RPB6	Q24320
DNA-directed RNA polymerase II subunit RPB7	RPB7	Q9VFB5
DNA-directed RNA polymerase II subunit RPB8	RPB8	Q9VNZ3
DNA-directed RNA polymerase II subunit RPB9	RPB9	P36958
DNA-directed RNA polymerase II subunit RPB10	RPB10	Q9VC49
DNA-directed RNA polymerase II subunit RPB11	RPB11	Q9VJE4
DNA-directed RNA polymerase II subunit RPB12	RPB12	Q6IGE3
Histone H2A	H2A	P84051
Histone H2B	H2B	P02283
Histone H3	H3	P02299
Histone H4	H4	P84040
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	Q9VGQ1

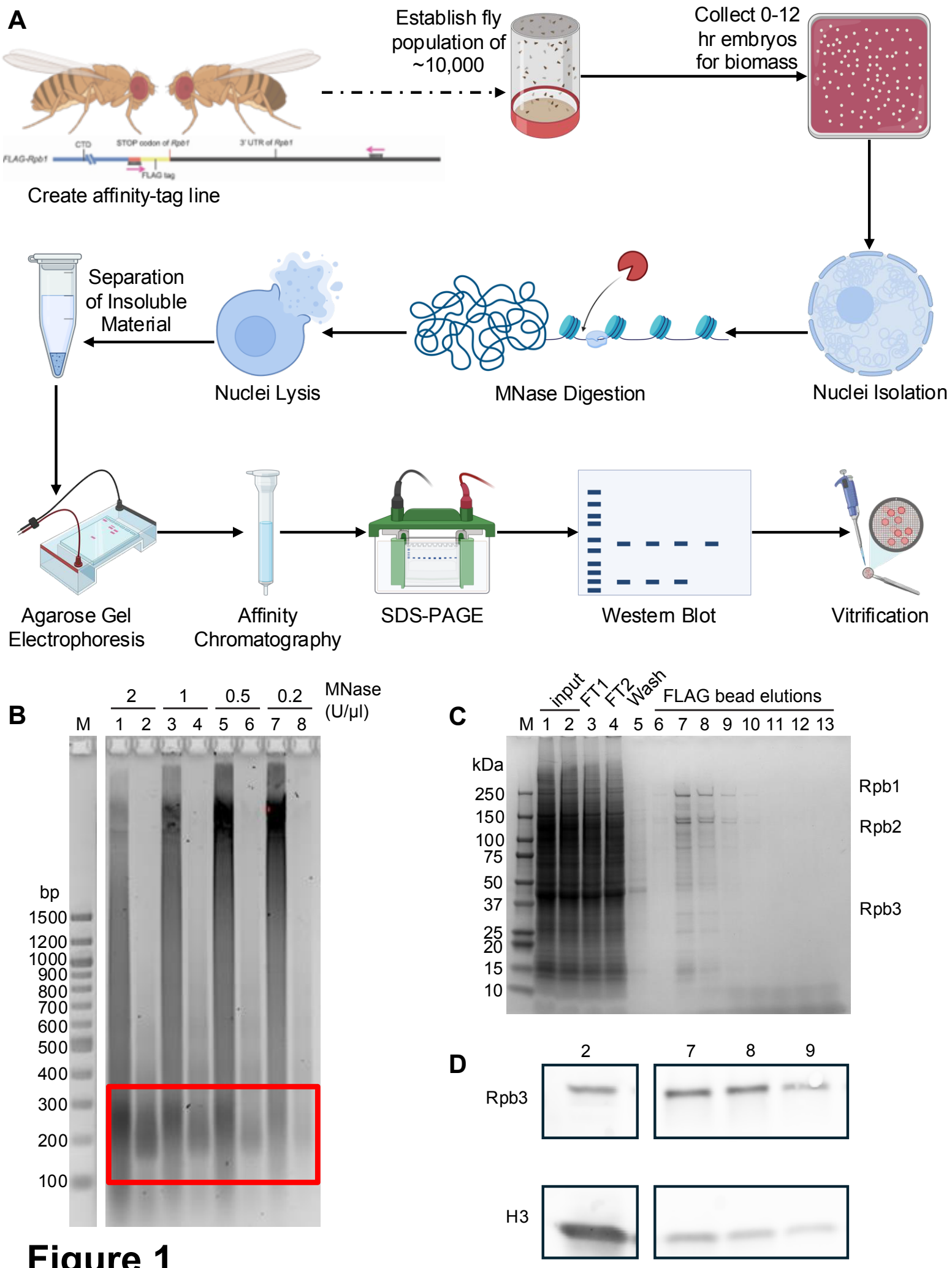

A
Establish fly population of ~10,000
Collect 0-12 hr embryos for biomass
CTD
STOP codon of Rpb1
3' UTR of Rpb1
FLAG-Rpb1
FLAG tag
Create affinity-tag line
Separation of Insoluble Material
Nuclei Lysis
MNase Digestion
Nuclei Isolation
Agarose Gel Electrophoresis
Affinity Chromatography
SDS-PAGE
Western Blot
Vitrification
B
2
1
0.5
0.2
MNase (U/µl)
M
1 2 3 4 5 6 7 8
bp
1500
1200
1000
900
800
700
600
500
400
300
200
100
C
input
FT1
FT2
Wash
FLAG bead elutions
M 1 2 3 4 5 6 7 8 9 10 11 12 13
kDa
250
150
100
75
50
37
25
20
15
10
Rpb1
Rpb2
Rpb3
D
2
7
8
9
Rpb3
H3

Figure 1

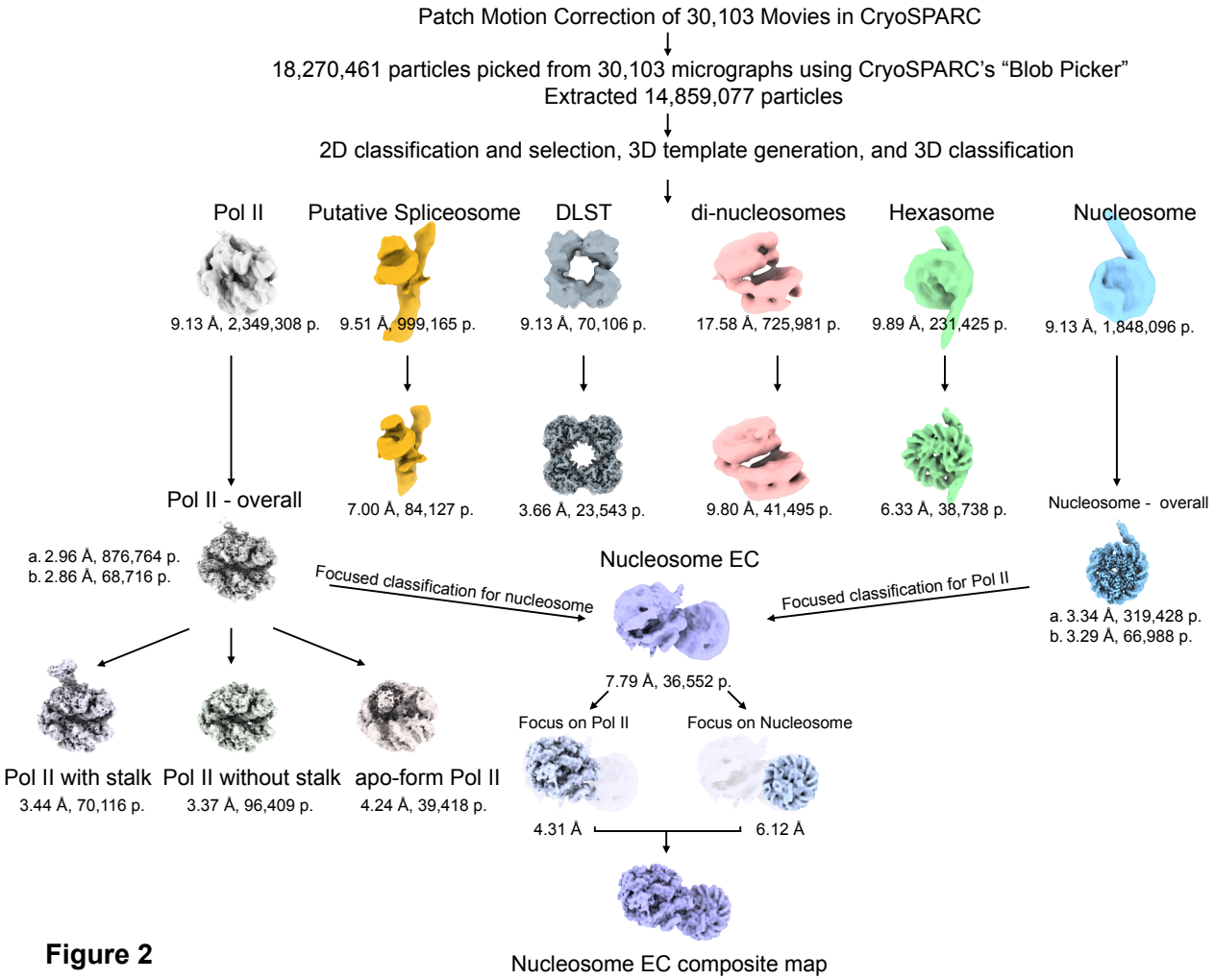

Patch Motion Correction of 30,103 Movies in CryoSPARC
18,270,461 particles picked from 30,103 micrographs using CryoSPARC's "Blob Picker"
Extracted 14,859,077 particles
2D classification and selection, 3D template generation, and 3D classification
Pol II
9.13 Å, 2,349,308 p.
Putative Spliceosome
9.51 Å, 999,165 p.
DLST
9.13 Å, 70,106 p.
di-nucleosomes
17.58 Å, 725,981 p.
Hexasome
9.89 Å, 231,425 p.
Nucleosome
9.13 Å, 1,848,096 p.
7.00 Å, 84,127 p.
3.66 Å, 23,543 p.
9.80 Å, 41,495 p.
6.33 Å, 38,738 p.
Pol II - overall
a. 2.96 Å, 876,764 p.
b. 2.86 Å, 68,716 p.
Focused classification for nucleosome
Nucleosome EC
Focused classification for Pol II
Nucleosome - overall
a. 3.34 Å, 319,428 p.
b. 3.29 Å, 66,988 p.
7.79 Å, 36,552 p.
Pol II with stalk
3.44 Å, 70,116 p.
Pol II without stalk
3.37 Å, 96,409 p.
apo-form Pol II
4.24 Å, 39,418 p.
Focus on Pol II
Focus on Nucleosome
4.31 Å
6.12 Å
Nucleosome EC composite map

Figure 2

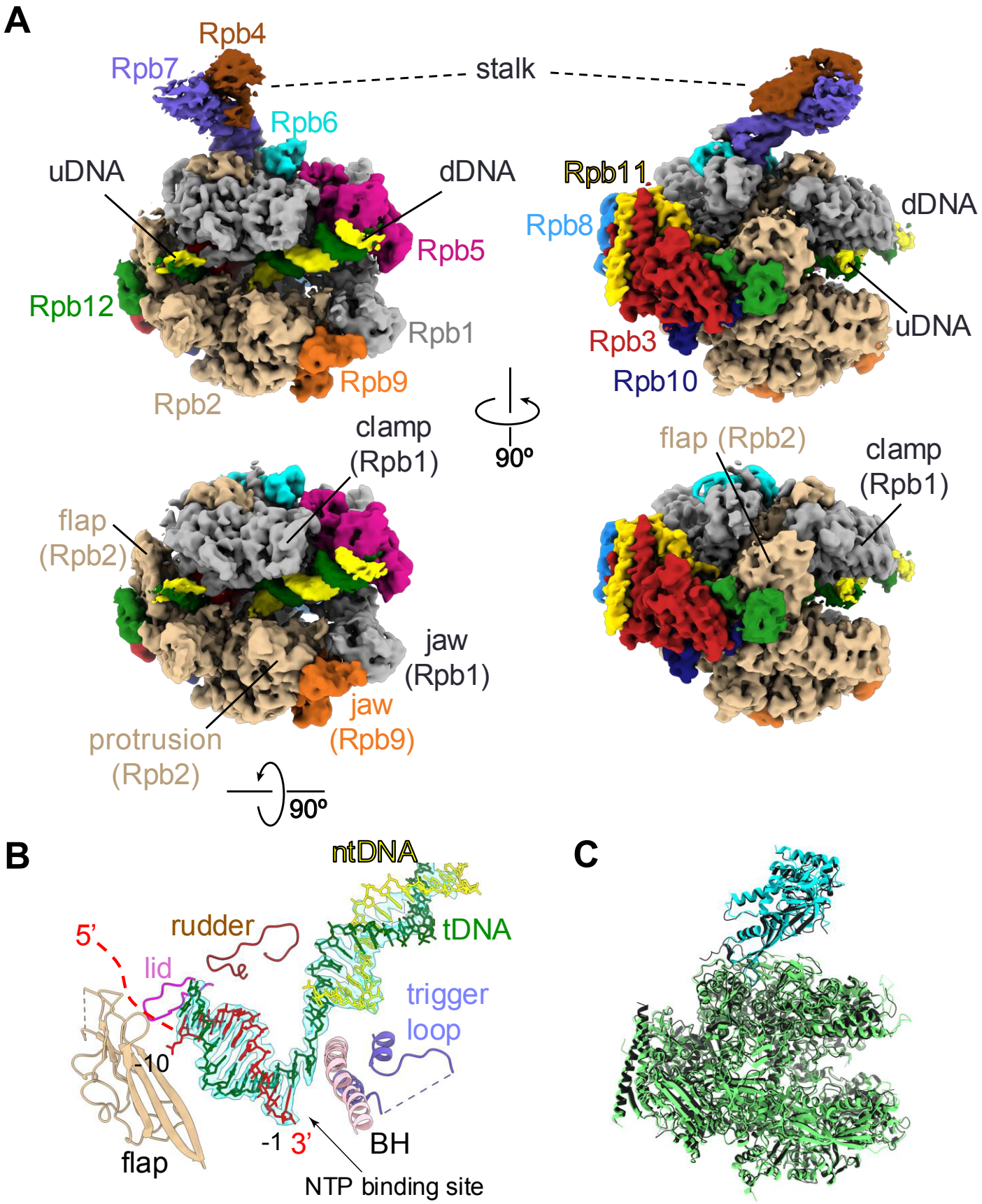

A
Rpb4
Rpb7
stalk
Rpb6
uDNA
dDNA
Rpb11
Rpb8
dDNA
Rpb5
Rpb12
Rpb1
uDNA
Rpb3
Rpb9
Rpb10
Rpb2
90º
clamp
(Rpb1)
flap (Rpb2)
clamp
(Rpb1)
flap
(Rpb2)
jaw
(Rpb1)
jaw
(Rpb9)
protrusion
(Rpb2)
90º
B
ntDNA
5'
rudder
tDNA
lid
trigger
loop
-10
-1
3'
BH
flap
NTP binding site
C

Figure 3

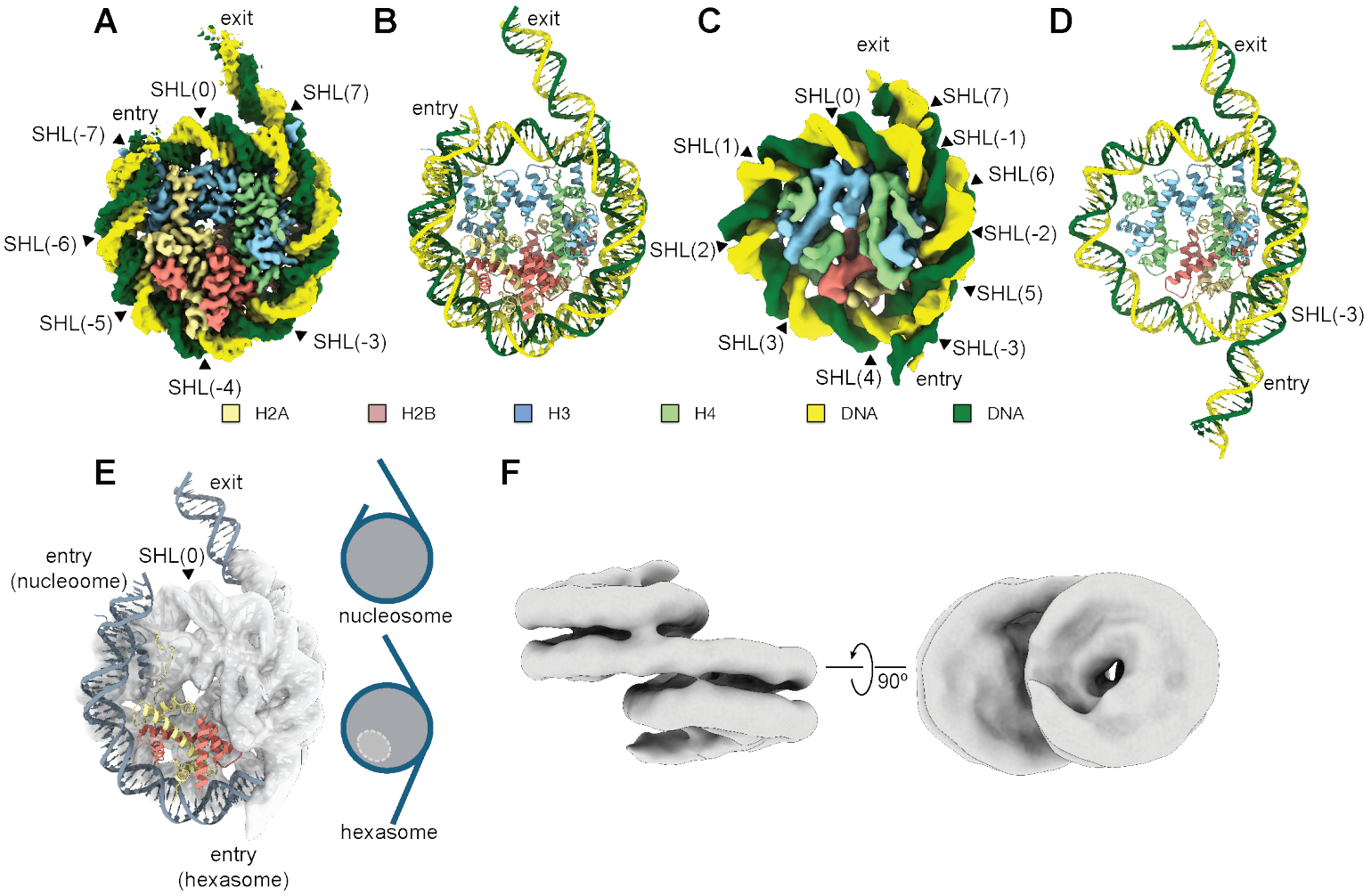

A
exit
SHL(0)
SHL(7)
entry
SHL(-7)
SHL(-6)
SHL(-5)
SHL(-3)
SHL(-4)
B
exit
entry
C
exit
SHL(0)
SHL(7)
SHL(1)
SHL(-1)
SHL(6)
SHL(-2)
SHL(2)
SHL(5)
SHL(3)
SHL(-3)
SHL(4)
entry
D
exit
SHL(-3)
entry
H2A
H2B
H3
H4
DNA
DNA
E
exit
entry
(nucleoome)
SHL(0)
nucleosome
hexasome
entry
(hexasome)
F
90°

Figure 4

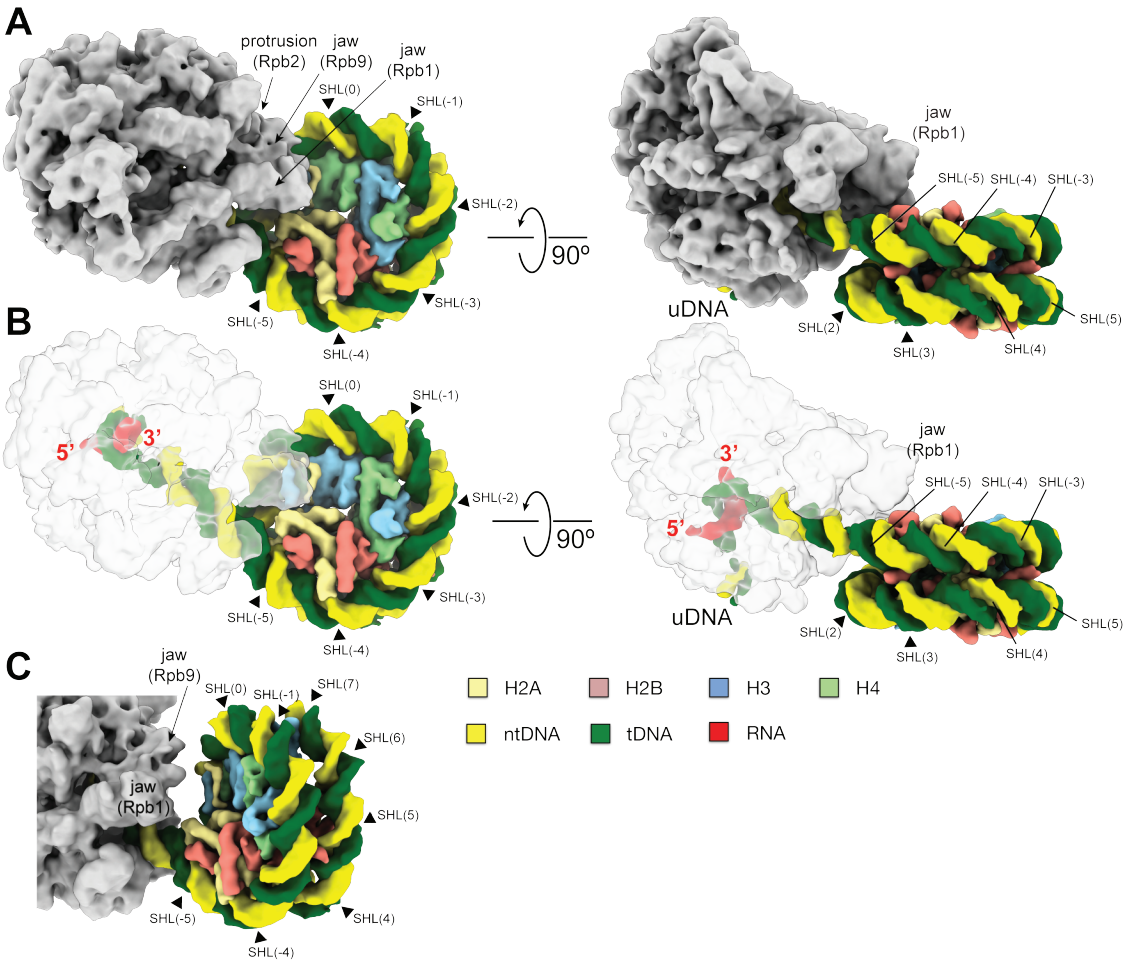
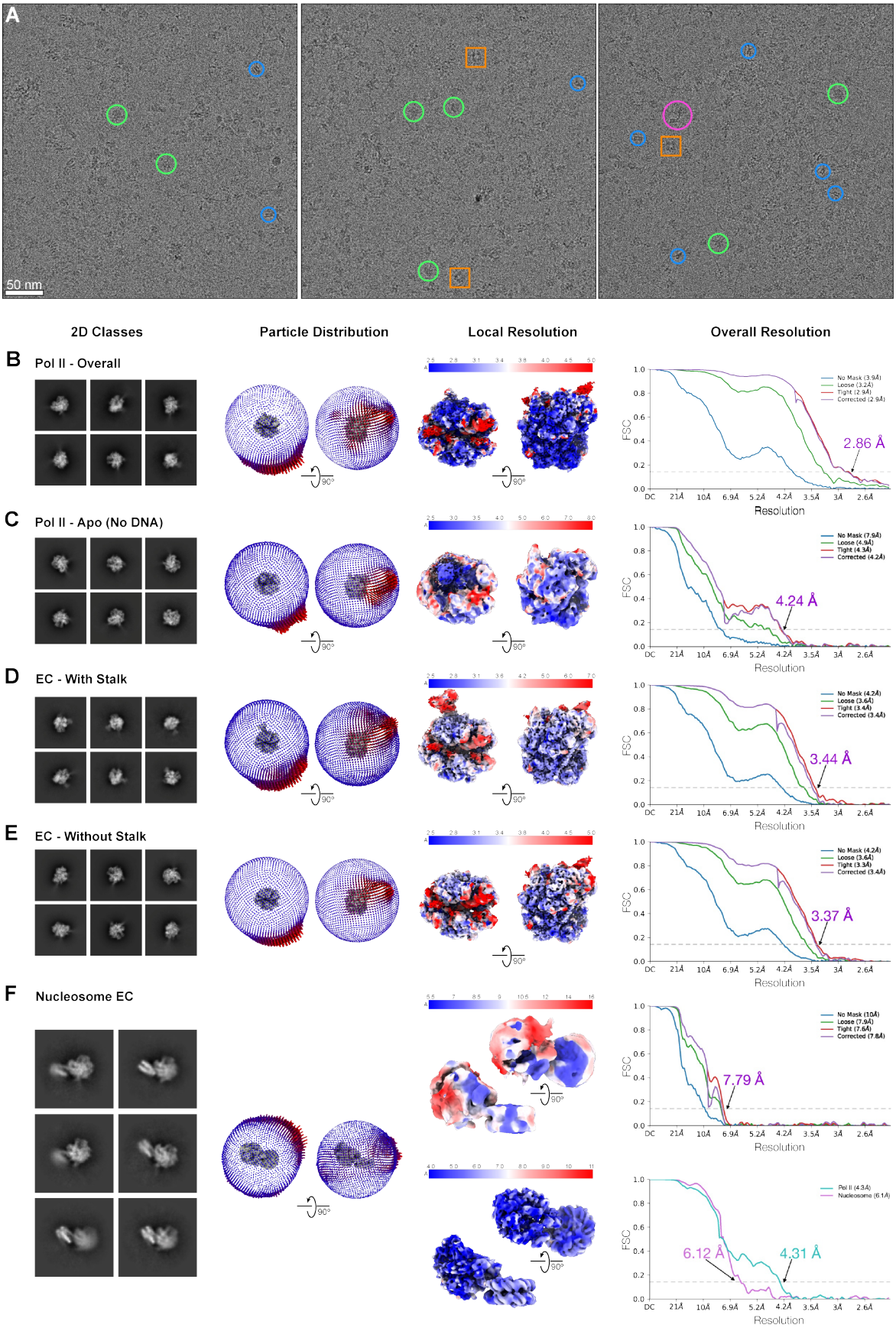
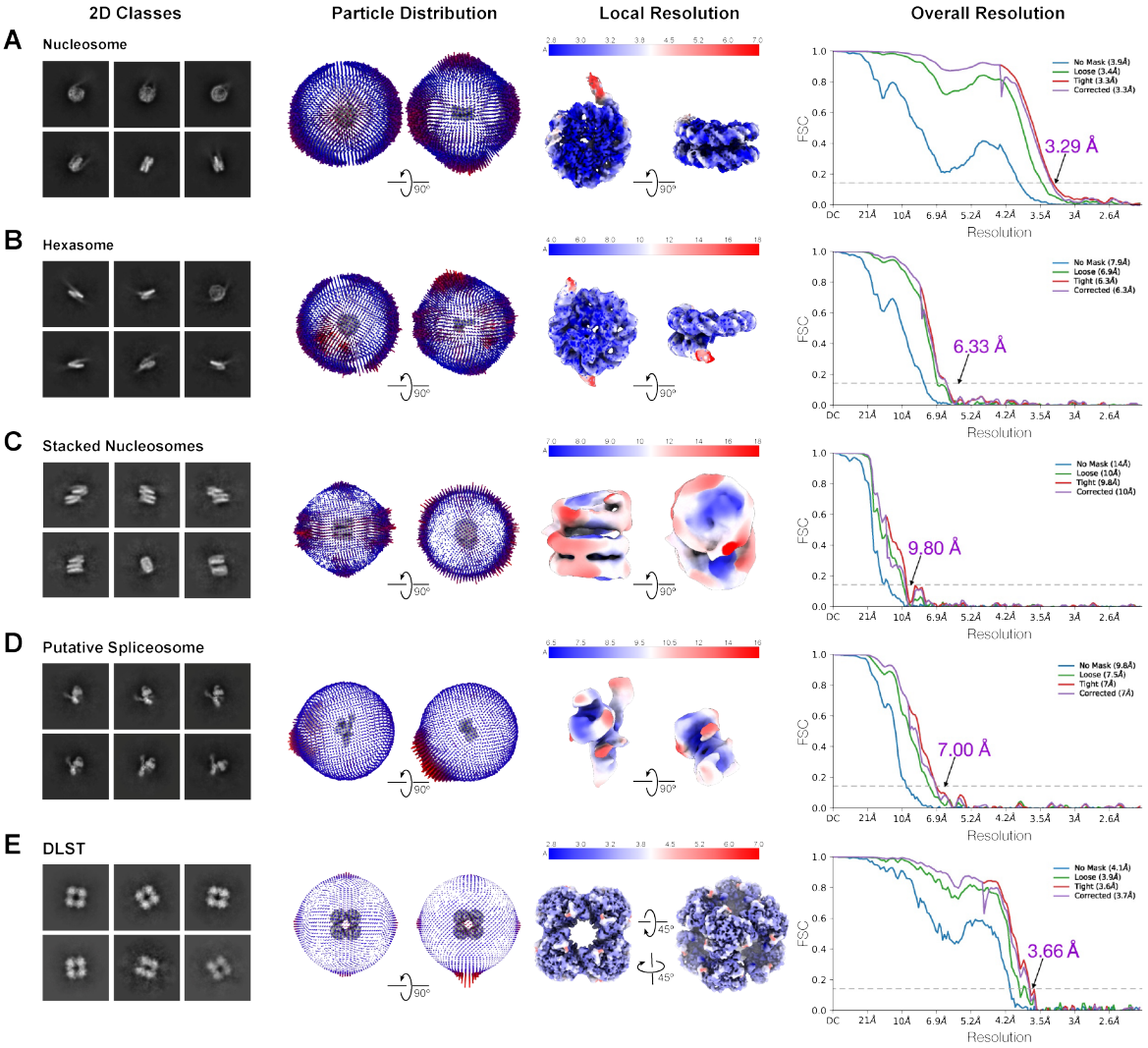
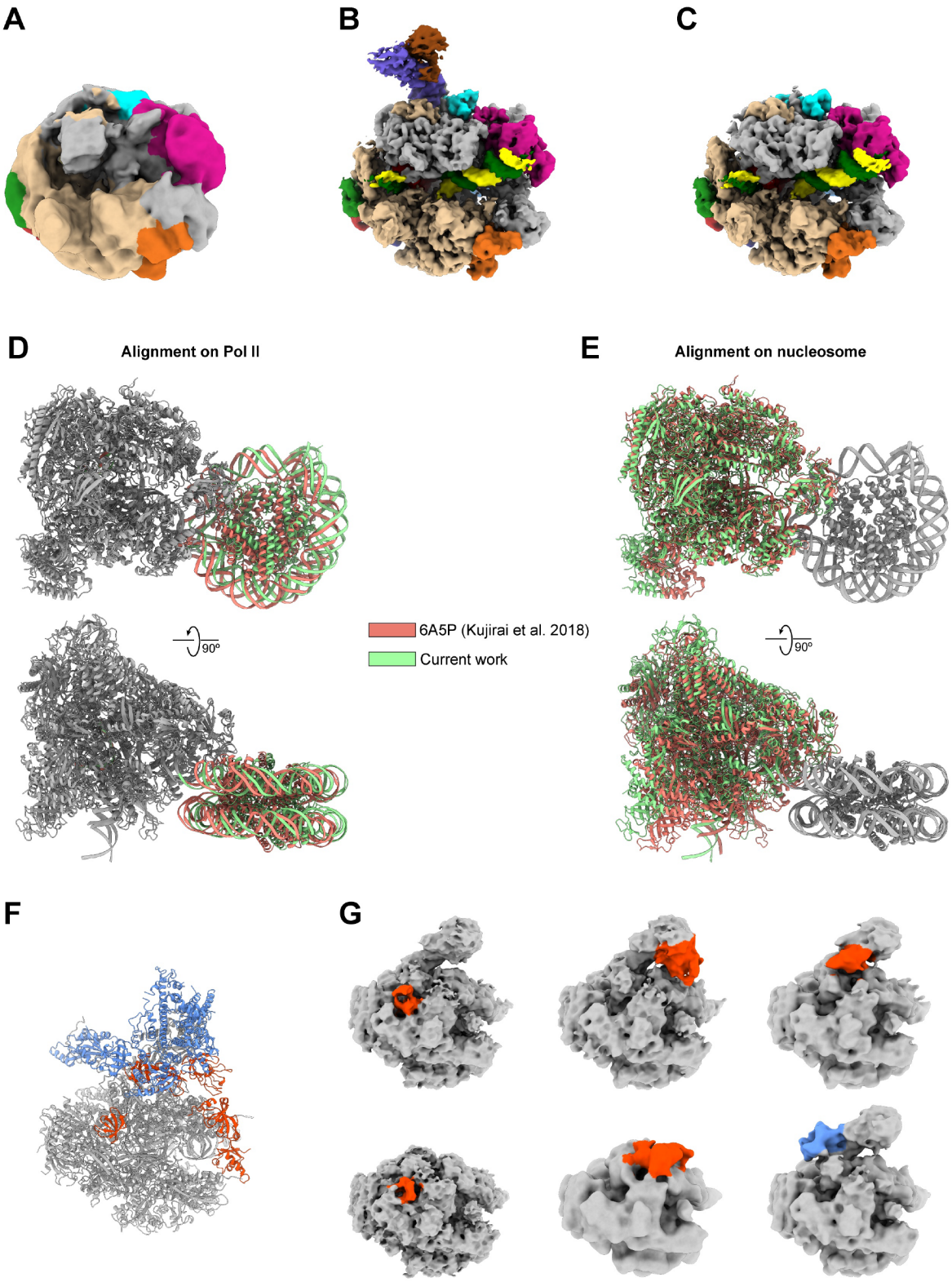
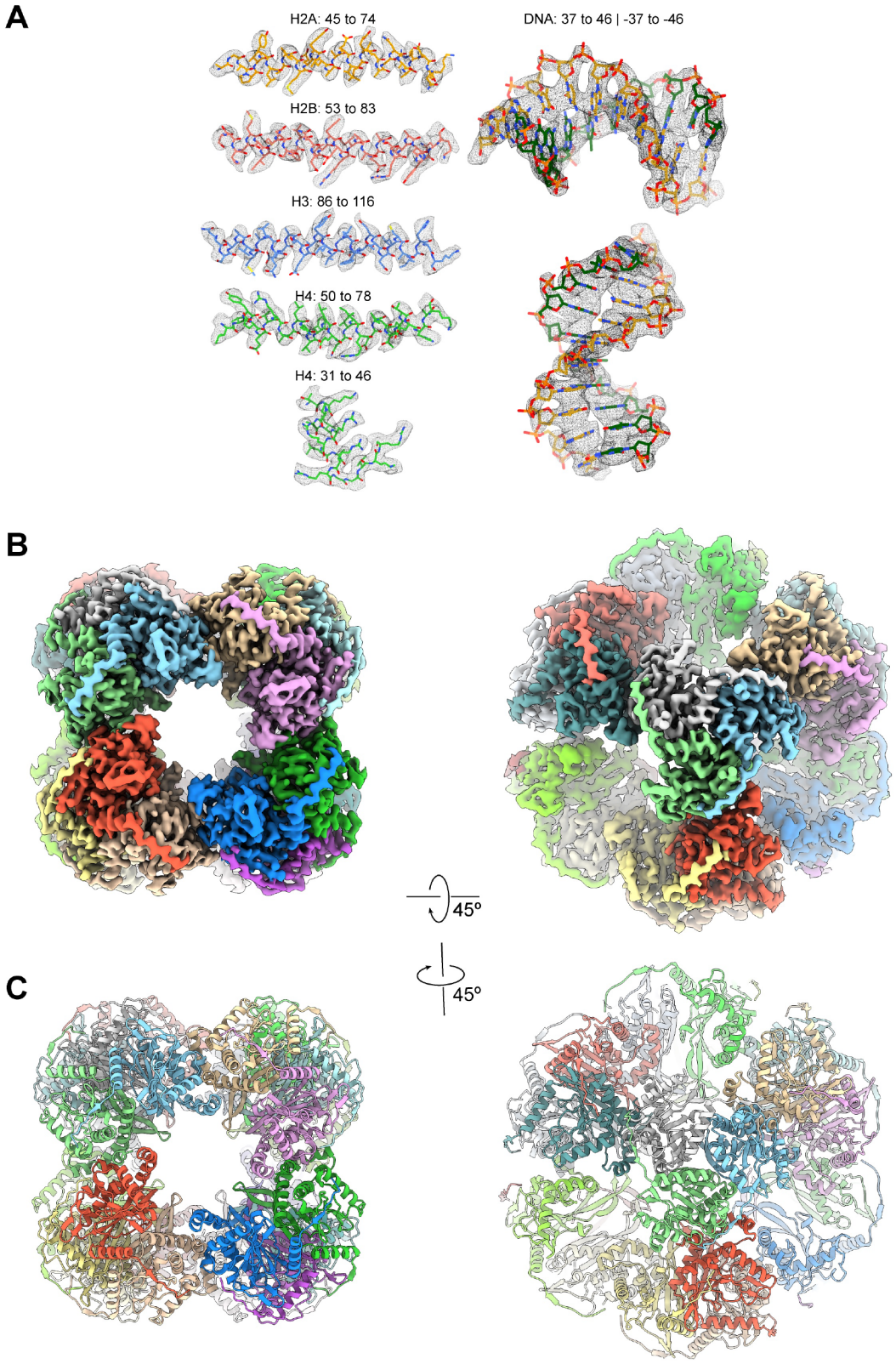

A
protrusion
(Rpb2)
jaw
(Rpb9)
jaw
(Rpb1)
SHL(0)
SHL(-1)
SHL(-2)
SHL(-3)
SHL(-4)
SHL(-5)
90°
jaw
(Rpb1)
SHL(-5) SHL(-4) SHL(-3)
uDNA
SHL(2)
SHL(3)
SHL(4)
SHL(5)
B
5'
3'
SHL(0)
SHL(-1)
SHL(-2)
SHL(-3)
SHL(-4)
SHL(-5)
90°
3'
5'
jaw
(Rpb1)
SHL(-5) SHL(-4) SHL(-3)
uDNA
SHL(2)
SHL(3)
SHL(4)
SHL(5)
C
jaw
(Rpb9)
jaw
(Rpb1)
SHL(0)
SHL(-1)
SHL(7)
SHL(6)
SHL(5)
SHL(4)
SHL(-5)
SHL(-4)
H2A
H2B
H3
H4
ntDNA
tDNA
RNA

Figure 5


A
50 nm
2D Classes
Particle Distribution
Local Resolution
Overall Resolution
B
Pol II - Overall
2.86 Å
C
Pol II - Apo (No DNA)
4.24 Å
D
EC - With Stalk
3.44 Å
E
EC - Without Stalk
3.37 Å
F
Nucleosome EC
7.79 Å
6.12 Å
4.31 Å
FSC
Resolution
90°


2D Classes
Particle Distribution
Local Resolution
Overall Resolution
A
Nucleosome
B
Hexasome
C
Stacked Nucleosomes
D
Putative Spliceosome
E
DLST
90°
45°
FSC
Resolution
No Mask (3.9Å)
Loose (3.4Å)
Tight (3.3Å)
Corrected (3.3Å)
3.29 Å
No Mask (7.9Å)
Loose (6.9Å)
Tight (6.3Å)
Corrected (6.3Å)
6.33 Å
No Mask (14Å)
Loose (10Å)
Tight (9.8Å)
Corrected (10Å)
9.80 Å
No Mask (9.8Å)
Loose (7.5Å)
Tight (7Å)
Corrected (7Å)
7.00 Å
No Mask (4.1Å)
Loose (3.9Å)
Tight (3.6Å)
Corrected (3.7Å)
3.66 Å


A
B
C
D
Alignment on Pol II
E
Alignment on nucleosome
90°
6A5P (Kujirai et al. 2018)
Current work
90°
F
G


A
H2A: 45 to 74
H2B: 53 to 83
H3: 86 to 116
H4: 50 to 78
H4: 31 to 46
DNA: 37 to 46 | -37 to -46
B
45º
C
45º