Title
Placing the preinitiation complex at the promoter

Author
Patel, Avinash Bharat

Publication Date
2019

Peer reviewed|Thesis/dissertation

Placing the preinitiation complex at the promoter

By

Avinash Bharat Patel

A dissertation submitted in partial satisfaction of the
requirements of the degree of
Doctor in Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eva Nogales, Chair
Professor James Hurley
Professor Andreas Martin
Professor Donald Rio

Summer 2019

Placing the preinitiation complex at the promoter

Abstract

Placing the preinitiation complex at the promoter
by
Avinash Bharat Patel

Doctor of Philosophy in Biophysics
University of California, Berkeley
Professor Eva Nogales, Chair

Transcription is the process of copying the genetic code and producing functional intermediates or readable instructions. The transcription of protein coding genes produces messenger RNA which the cell uses to synthesize proteins that carry out various functions. In this way the process of transcription initiation is a major determinant of cell fate as it controls which genes will be transcribed.

The main subject of this thesis will look at how human TFIID binds the core promoter and subsequently aids in the assembly of the transcription preinitiation complex. Early work on TFIID had shown that the complex binds the promoter at two distinct sites, a large region downstream of the transcription start site and a smaller region upstream. Subsequent studies found several conserved sequences within the core promoter such as the TATA-box, downstream promoter element, initiator and motif ten element. Structural studies of TFIID have revealed that the complex adopts a trilobed structure that undergoes a dramatic rearrangement upon promoter binding. Through these works and others, a great deal about the function of TFIID has been revealed, however the underlying mechanism by which TFIID binds the promoter and subsequently recruits RNA polymerase II remained missing. This thesis attempts to address these last two questions.

Using a combination of biochemical reconstitution, chemical crosslinking mass spectrometry, and electron microscopy I have been able to determine the complete architecture of TFIID and how it loads TBP onto the promoter. The structure of TFIID contains an asymmetric dimer at its core that forms a scaffold which the rest of the complex assembles upon. It was also found that TFIID initially binds the downstream promoter, and then subsequently rearranges to load TBP onto the TATA-box with the help of TFIIA. From this model we can begin to rationalize how TFIID integrates information from histone marks on the downstream +1 nucleosome and upstream activators.

While working with TFIID and other samples a common issue has been preferred sample orientation on the grid. This issue has led to artefacts in the reconstruction due to missing information. To overcome this issue, I have developed a reliable method to coat gold foil EM grids with graphene oxide. These grids can be used to collect tilted images which can produce high resolution isotropic reconstructions as they do not suffer large amounts of drift.

For my parents and teachers.

Table of Contents

List of Figures

Acknowledgements

I would like to thank all the people I have worked and learned from through my undergraduate and graduate studies.

I would like to first thank Eva Nogales for accepting me into her lab and allowing me to learn from the amazing research community she has built in her lab. In particular, I would like to thank Jie Fang for the many weeks, months and years of work she has put into purifying TFIID which without, this thesis would not be possible. Additionally, I would like to thank Jie for always looking out for me, especially for all the snacks she has given me these past 5 years. I would also like to thank Patricia Grob for all the many instructive and illuminating discussion about various EM topics, Teresa Tucker for keeping the lab running smoothly, Abhiram Chintangal and Tom Houweling for making sure the lab was always warm, and Lilia Garcia for providing a well-stocked and clean cabinet of lab equipment.

I would also like to thank the other members of the Nogales lab for their advice, training and mentorship. In particular, I would like to thank Robert Louder for training me on how to use the itan, advising on my work on TFIID and for the may discussions on transcription. I would also like to thank Basil Greber for the advice on data processing and model building, Stuart Howes for training me when I was rotating in the lab, Vignesh Kasinath for encouraging me to be more independent, and lastly Sarah Sterling and Ben LaFrance for making the lab a great place to be.

A lot of work that has not made it into my thesis but will hopefully be publish once I graduate was done in collaboration with the many amzing undergraduates I have been lucky enough to work with. I would like to thank all of them for helping keep me focused on research. Sarah Shattuck who worked on so many different projects but whose work helped elucidate the role TFIID has on Pol II recruitment, Anthony Liu for reconstituting the HAT module of SAGA and establishing native histone purification, Stefan Zukin for determining the structure of NuA4, Camille Moore for determining the structure of RSC and Audrey Litvak for reconstituting lobe C of TFIID.

I would also like to thank Goran Stjepanovic for his advice and training when it can to tissue culture and mass spectrometry, Lars-Anders Carlson for teaching me the fundamentals of EM, Tony Iavarone for advice on mass spectrometry, and Hitomi Asahara and Bob Lesch for training me on how to use the fermentor. And lastly, I would like to thank my undergraduate mentor Sheryl Tsai and her lab who help develop my interest in structural biology. I would especially like to thank Dave Jackson my first and best scientific mentor.

Chapter 1

Introduction

1.1 Eukaryotic transcription initiation

The human genome contains ~20,000 genes, with only a fraction being expressed in any given cell[1]. Differential gene expression allows for developmental patterning, response to environmental stimuli, control cell growth, cell type differentiation and maintaining homeostasis[2]. While gene expression can be controlled at many stages, the main point of control is at the first step, transcription initiation. The process of transcription initiation determines which genes are to be transcribed and at what level. This is accomplished by the assembly of the transcription preinitiation complex (PIC) which bring RNA polymerase II (Pol II) to the promoter of genes to begin the synthesis of messenger RNA (mRNA)[3,4].

From biochemical studies over the past few decades the process of eukaryotic PIC assembly has been found to proceed in a sequential manner. The first general transcription factor (GTF) to recognize the promoter is transcription factor IID (TFIID) which binds to promoter DNA with the aid of TFIIA[5]. This leads to the recruitment of a TFIIB – Pol II complex followed by TFIIF which stabilizes the interaction between Pol II and the promoter[6]. The subsequent addition of TFIIE further stabilizes the PIC and allows for the recruitment of TFIIH, which mediates the opening of the promoter DNA duplex[7].

In recent years many parts of the PIC assembly process have been visualized using cryo electron microscopy (cryo-EM). These studies have revealed the mechanism of PIC assembly from the initial binding of TATA-box binding protein (TBP) a component on TFIID (Figure 1.1.1)[8]. However, how TFIID contributes to this process remains unanswered. The primary focus of this thesis will be to elucidate how TFIID manages to bind the promoter and place TBP onto the TATA-box using cryo-EM.

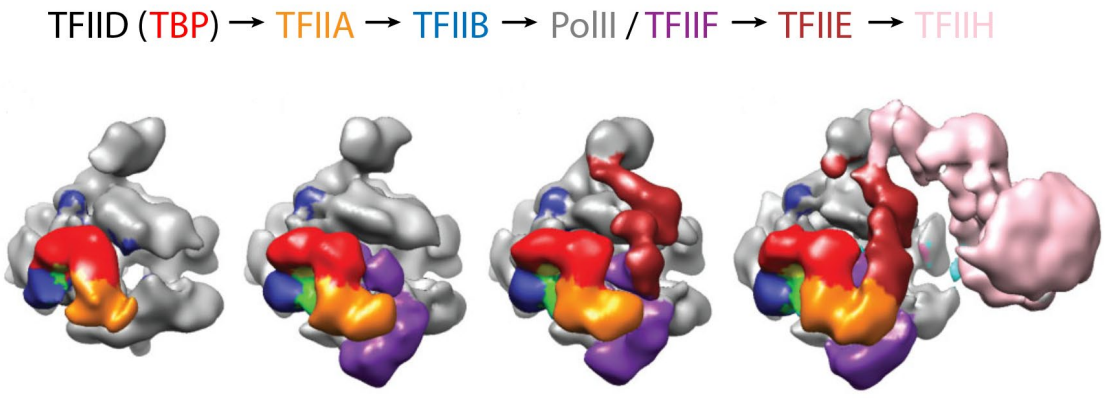

Figure 1.1.1 | Minimal PIC assembly pathway. Assembly of the human PIC onto the core promoter starting with TBP, TFIIA, TFIIB and Pol II. This is followed by the addition of TFIIF, TFIIE and TFIIH.

1.2 TFIID as an initiation factor for preinitiation complex assembly

In 1980, the lab of Robert Roeder began fractionating nuclear protein to identify what components were necessary to initiate Pol II transcription[9]. During that decade the Roeder lab continued to study these purified fractions and found that: TFIID and TFIIA were the first factors to bind to the promoter, TFIID bound the promoter both upstream and downstream of the transcriptional start site (TSS), and activators enhance transcription through TFIID[3,10,11]. Having determined that TFIID initiates transcription, his lab set out to clone yeast TFIID and produce it recombinantly. Analysis of the recombinantly purified TFIID by footprinting showed that it only protected the region downstream of the TSS where the TATA box is located[4]. The reason for this difference was solved by the labs of Robert Tjian and Arnold Berk not long after. Tjian's group showed that the protein thought to be TFIID (which he would rename TBP) pulled down with several other proteins or what he called TBP associated factors (TAFs)[12]. Soon thereafter Berk's group show that TFIID (now defined as TBP and TAFs) provided the same level of DNA protection in footprinting experiments as seen by the Roeder lab[13].

The discovery that TFIID is a multisubunit protein complex led to numerous studies of its function and structure. Studies on activator-TAF complexes showed that both TAF4 and TAF9 could bind activators[14–16]. Work on TAF-promoter binding showed that the complex of TAF1/2 and TBP could bind the initiator (Inr) and that a complex of TAF6/9 crosslinks to the downstream promoter element (DPE)[17,18]. Numerous studies have also mapped the interactions between the TAFs that make up TFIID and other GTFs[19–23]. Structural studies have also determined a few pieces of the complex by X-ray crystallography and nuclear magnetic resonance, but it was not until electron microscopy groups turned their attention to TFIID was any complete structure of the complex observed[24].

1.3 Electron microscopy studies on TFIID

Beginning in the late 1990s, the Schultz and Nogales labs' began to study the full complex by electron microscopy. Their labs discovered that TFIID forms a horseshoe like structure that is made up of three large lobes[24,25]. Follow up studies and revealed that TFIID is a flexible complex that predominantly exists in two major conformations[26].

In 2013 with the aided of new image processing tools and the creation of a super core promoter (SCP) sequence it became possible to computationally sort out flexible of TFIID[27]. It was shown that TFIID oscillates between the canonical and rearranged states with DNA binding only possible in the rearranged states (Figure 1.3.1)[28].

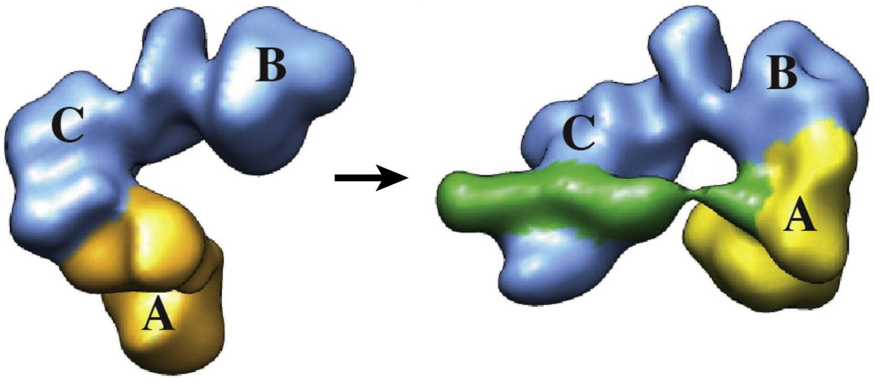

Figure 1.3.1 | Structural transition of human TFIID upon promoter binding. On the left the structure of TFIID in the canonical state and on the right the rearranged state which is promoter binding capable state. Each of the lobes are marked and the promoter DNA is colored in green.

In a subsequent 2016 study aided by new camera technology, processing softwares and sample preparation methods the structure of the DNA bound TFIID complex was determined. This was the first sub nanometer reconstruction of TFIID, which allowed for the docking of previously determined crystal structures and homology models unambiguously into the map. The structure showed that the Inr, DPE and MTE were recognized by TAF1 and TAF2 and that TBP with the help of TFIIA binds the TATA-box at the upstream promoter[29]. The structure revealed for the first time the positions of six different components of the TFIID complex (TAF1, TAF2, TAF6, TAF7, TAF8 and TBP) and revealed that the complex contains two copied of TAF6 which dimerizes through its HEAT repeat (Figure 1.3.2). Despite these advancements, a complete molecular structure of TFIID has still not been determined and neither has the mechanism of how TFIID binds the promoter been revealed.

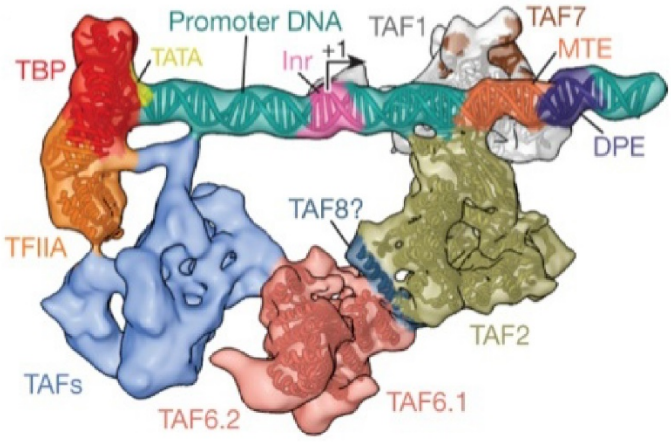

Figure 1.3.2 | Architecture of the promoter bound TFIID complex. The structure of the BC core of TFIID bound to the super core promoter. TAF1 and TAF2 of lobe C bind the Initiator (Inr), motif ten element (MTE) and downstream promoter element (DPE). The TATA box is bound by TBP which is connected to lobe B though TFIIA.

1.4 Thesis summary

To answer the question of how TFIID binds the promoter and help recruits Pol II, I aimed to determine the structure of TFIID. To accomplish this, I began by looking at the endogenously purified TFIID complex by cryo-EM. This was initially problematic and require some sample optimization, but by crosslinking the sample to limit flexibility and coating the TEM grid with a cationic polymer was I was I able to determine the complete evolutionarily conserved core of TFIID (Chapter 2, 3 and 6)[30].

Together with the previously determined promoter bound structure of TFIID I was able to propose a mechanism of how TFIID binds the promoter (Chapter 4)[30]. To better understand how TFIID recruits Pol II to the promoter I began attempts to capture the transition state intermediate between promoter bound TFIID and the early PIC containing Pol II, TFIIA, TFIIB and TBP. My preliminary results show that TFIID disengages from the promoter upon Pol II binding (Chapter 5)[30].

During my time preparing cryo-EM samples I encounter several reoccurring issues. One was that most samples experience some degree of preferential orientation bias. Another was that carbon support layers typically introduce a high amount of background to cryo-EM images. Methods to overcome both issues have been developed, however I found that implementing them are often difficult or impractical. At the end of this thesis I will describe a method to coat TEM grids with graphene oxide which produces very little background. Additionally, if this coating method is used with gold foil grids the sample can be tilted for data collection to mitigate the issue to preferential orientation bias (Chapter 7).

Chapter 2

Challenges in sample preparation of TFIID

2.1 Crosslinking to limit flexibility

The structure of TFIID has been described to have three lobes, termed lobe A, B and C[24]. Lobes B and C form what is called the BC core due to the rigid connection between these two lobes. Relative to the BC core Lobe A is highly flexible and exists in a continuum of states that has been shown to oscillate from one side of the complex to the other. However negative stain analysis of the motion of lobe A has shown that the range of motion is potentially even greater than previously realized (Figure 2.1.1). The main reason that lobe A appears more flexible is that the new analysis has been done with negative stain data rather than cryo-EM data. Negative stain data has much higher signal-to-noise ratio then cryo-EM data for low frequency information which with new image processing tools can better classify and align lobe A. Another potential reason is that the staining process can flatten the sample and potentially distort the shape of the complex.

The sample was crosslinked to limit the flexibility of lobe A and prevent any distortions to the complex due to its binding to the grid support layer. Crosslinking cryo-EM samples prior to grid application has been shown to preserve complex integrity[31]. To identify the best condition, I tested different crosslinkers (BS3, DSS and glutaraldehyde) at different concentrations (1mM and 5mM) all at the same temperature (4^{o}C) and for the same amount of time (5min). Due to the limited amount of sample I chose to screen these conditions by negative stain EM. I assessed the effectiveness of the crosslinking by performing 2D classification. The conditions that limited the flexibility of lobe A while still retaining the TFIID-like features were deemed suitable for structural studies. I found that both succinimide crosslinkers (BS3 and DSS) where not effective at crosslinking the complex. This could be due to the short incubation time and low temperature of the reaction, but also because these crosslinking agents more easily react with the peptide carried over from the immunopurifcation step. For the glutaraldehyde conditions I found the both concentrations effectively crosslinked the complex. Though 1mM (.01%) produced a greater number of 2D class averages with TFIID-like features. The final conditions that were used for preparing cryo-EM grids were prepared by crosslinking TFIID with 1mM glutaraldehyde for 5 minutes on ice before applying to the grid. The effect of this can be seen in Figure 2.1.1.

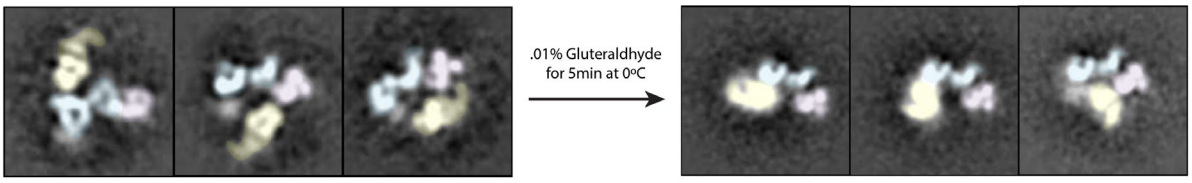

Figure 2.1.1 | Glutaraldehyde crosslinking reduces TFIID flexibility. On the left 2D class averages of uncrosslinked TFIID and on the right after crosslinking. Lobe A is colored in yellow, B purple and C blue.

2.2 Grid coating to mitigate preferential orientation bias

Initial attempts at preparing cryo-EM samples using the 1mM glutaraldehyde crosslinking condition were successful. Cryo-EM micrographs showed monodispersed particles with high contrast and produced 2D class averages that appeared to have alpha helical features. The resulting reconstructions showed a great deal of anisotropy, meaning that helices could only be distinguished from one orientation (Figure 2.2.1). To overcome this issue, I initially tried to collect more data to compensate for the rare views, but this was not effective. Instead I attempted to change the orientation on the grid by different grid treatments.

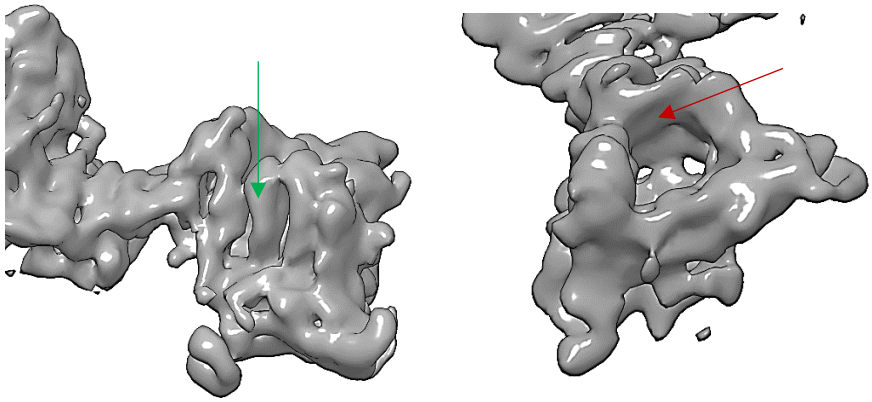

Figure 2.2.1 | Anisotropic reconstructions of TFIID. On left a helix is marked with a green arrow. On the right a region where a vertical helix should be is marked with a red arrow.

Existing methods to change particle orientations on the grid mostly relied on making the surface positively charged. This is the opposite charge the grids have when they are plasma cleaned in an atmosphere of air. The two main methods that exist to make grids positively charged are: to plasma clean grids in an atmosphere that contains amylamine or coat grids with poly-lysine after plasma cleaning[32,33]. The amylamine method in theory creates less background but the compound is highly toxic and requires special containment, and a dedicated plasma cleaner. Additionally, the grids tend to be hydrophobic which makes automated data collection difficult. Poly-lysine treatment by comparison is nontoxic and renders the grid hydrophilic, but it has not been used for preparing cryo-EM samples to my knowledge. This is likely due to the perception that it causes a greater amount of background. However due to the lack of a dedicated amylamine plasma cleaner I decided to try polymer coating.

I first decided to test three different polymers (poly-lysine, poly-arginine and polyethylenimine (PEI)) by negative stain to see if one would produce better results. I found that poly-arginine made most of the particles aggregate. Poly-lysine and PEI however both produced monodispersed samples, with poly-lysine producing better stain. However, I found that poly-lysine coated grids tended to be more variable when it came to particle concentration, with some areas/grids having far more particles then others. Also, poly-lysine

did not produce as many new views as PEI. In the end I decided to prepare and collect on PEI coated cryo grids (Figure 2.2.2). PEI coating the grids for cryo ended up producing far more views which helped overcome the issue to anisotropic reconstructions (Figure 2.2.3).

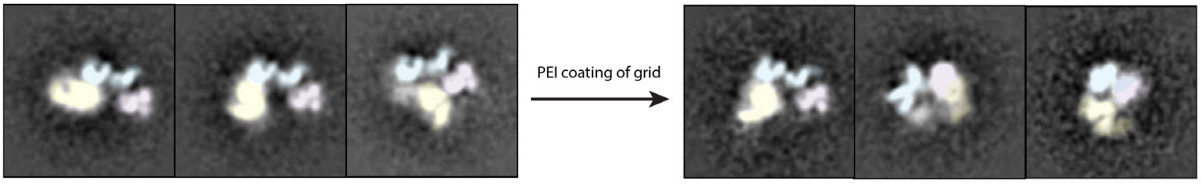

Figure 2.2.2 | PEI coating grids produces more views in negative stain. On the left 2D class averages of crosslinked TFIID on a plasma cleaned carbon support and on the right on a PEI coated grid. Lobe A is colored in yellow, B purple and C blue.

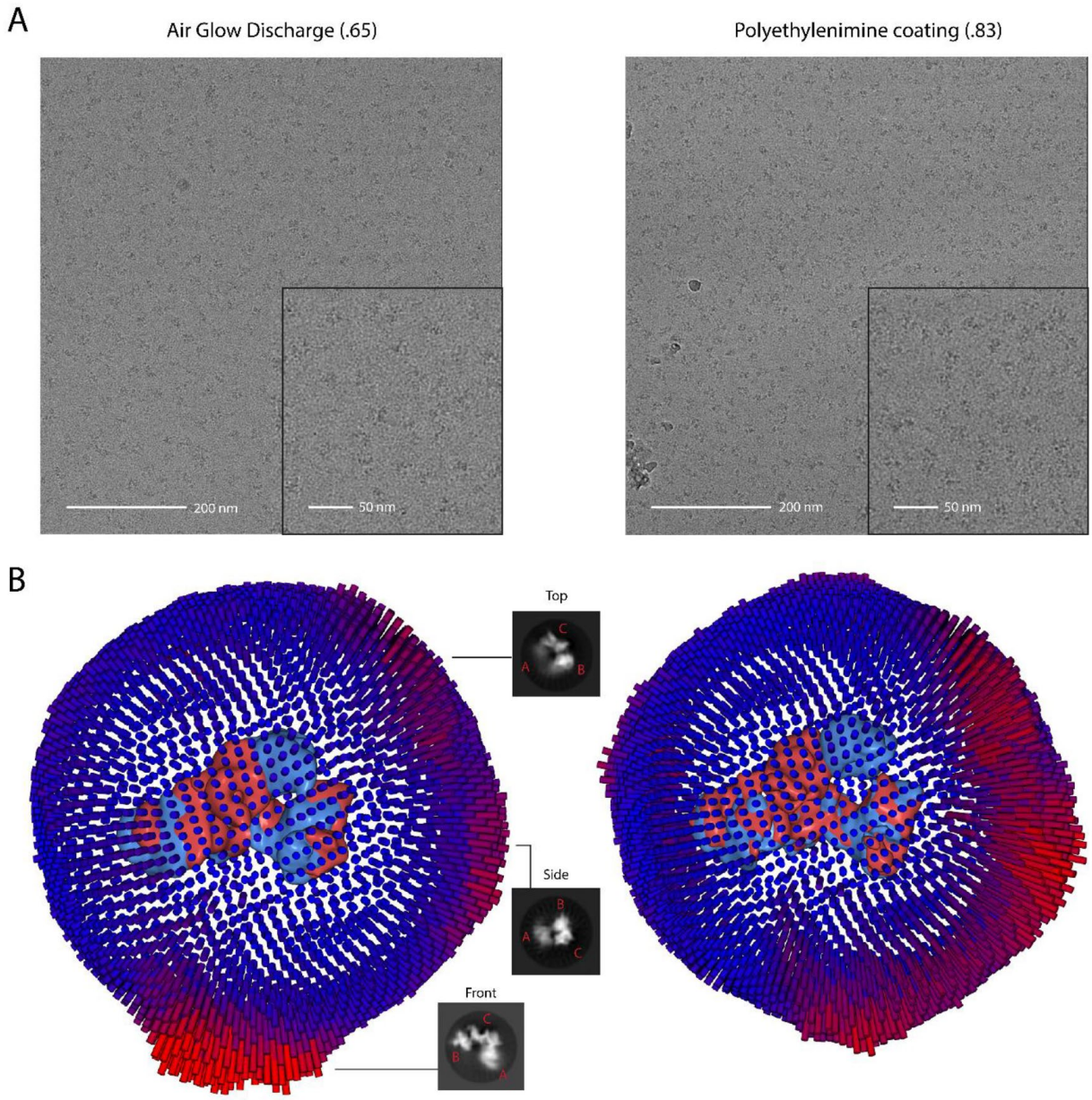

Figure 2.2.3 | Euler angle distribution of plot for plasma cleaned grids and PEI coat grids. (**A**) Representative cryo-EM micrographs, with an inset showing an enlarged view, of TFIID on either a glow discharged (left), and PEI coated grids (right). Values given in brackets are efficiencies derived from cryoEF. (**B**) Overlaid angular distribution plot from Relion for the both canonical and extended state of TFIID from glow discharged (left) and PEI coated grids (right). The major orientation peaks are marked with the corresponding 2D class averages.

Chapter 3

Structure of human TFIID

3.1 The structure of the TFIID core

The flexible nature of TFIID has long hampered a high-resolution structural description of the intact complex[28]. Previous studies showed how the distribution of positions of the flexibly attached lobe A shifts upon binding of promoter DNA and TFIIA[28]. Lobe A in apo-TFIID exists in a bimodal but continuous distribution of states, with roughly equal occupancy of two distinct, major states referred to as the canonical and extended states. Whereas in the canonical state lobe A is near lobe C, in the extended state lobe A is between lobes B and C (Figure 3.1.1). The displacement of lobe A between these two states is ~100 Å. By sorting a large cryo-EM dataset of free TFIID into two predominant states, refining them independently, and then combining the refined regions, we were able to extend the resolution of the BC core to 4.5 Å (range of 4.2 to 6.5 Å) and to generate a three-dimensional (3D) reconstruction of lobe A at 9.5 Å (range of 8.5 to 15 Å) (Figure 3.1.1). We then used a combination of cryo-EM, CX-MS, and structure prediction to generate a complete model of the complex (Figure 3.1.2).

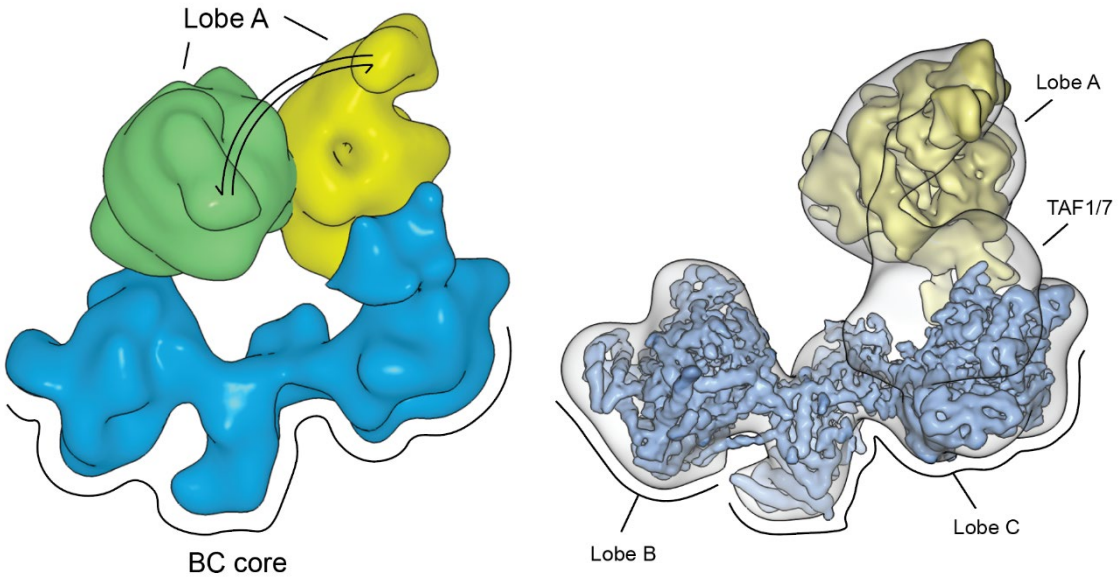

Figure 3.1.1 | Cryo-EM reconstructions of TFIID. On the left the cryo-EM reconstructions of TFIID, with the BC core in blue and lobe A in yellow (canonical state) and green (extended state). On the right a transparent cryo-EM map of TFIID in the canonical state with fitted cryo-EM maps from focused refinements of the BC core and lobe A in solid blue and yellow, respectively.

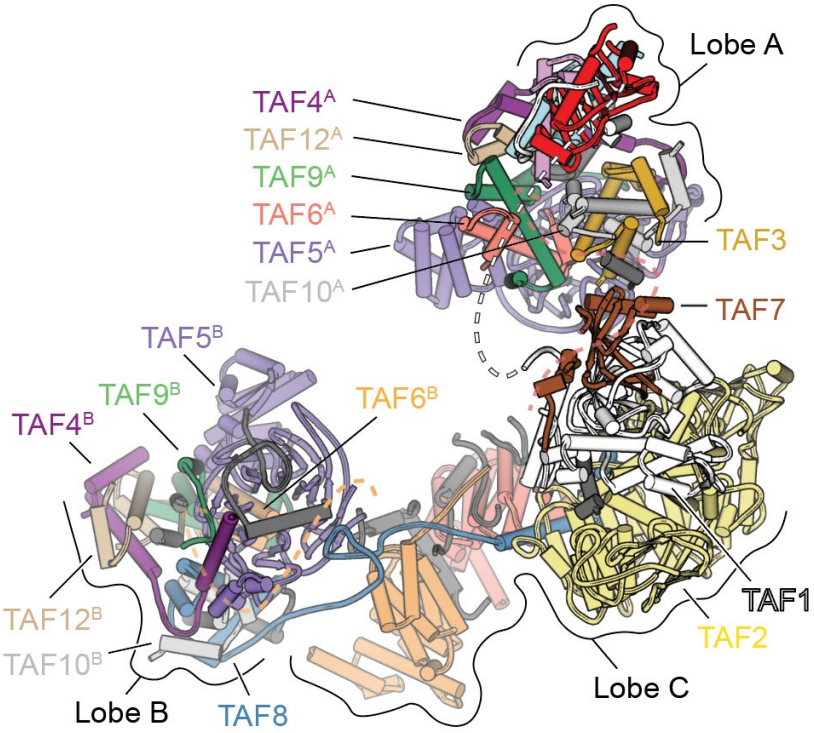

Figure 3.1.2 | Structure of TFIID. The structure of human TFIID in the canonical state.

Compared with that of the promoter-bound complex (IIDA-SCP) structure, the density corresponding to the TAF1-TAF7 subcomplex within lobe C in apo-TFIID is poorly defined, indicating that this module is flexible in the unbound TFIID, but stabilized upon binding to promoter DNA (Figure 3.1.3)[29]. For the rest of lobe C, it was possible to dock into the density the model of the TAF6 HEAT repeat dimer, a segment from the C-terminal region of TAF8, and the TAF2 aminopeptidase-like domain (APD) from the previous IIDA-SCP structure, with adjustments and extensions made to fit the observed density (Figure 3.1.4)[29].

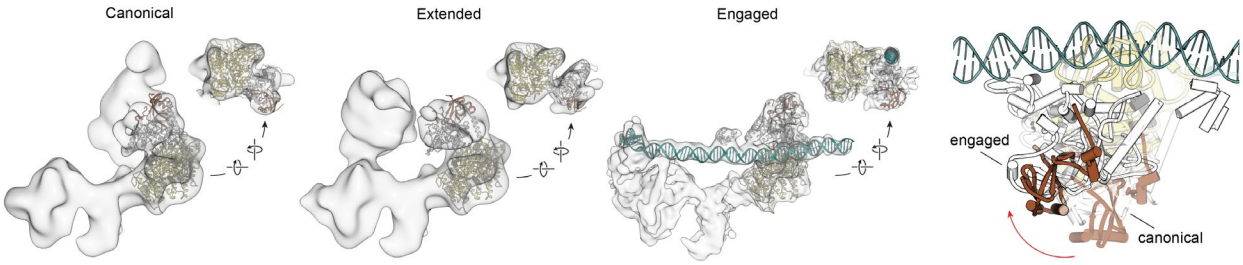

Figure 3.1.3 | Flexibility of the TAF1-TAF7 module. Docking of the TAF1/7 module into the canonical, extended and engaged state maps of TFIID. On the right the overlay of the TAF1/7 module for the canonical and engaged state, based on the alignment of TAF2.

Within lobe B, we were able to fit a homology model of the WD40 domain of TAF5, the crystal structures of the TAF5 NTD2 domain and the histone-fold domain (HFD) heterodimers of TAF6-TAF9, TAF4-TAF12, and TAF8-TAF10, as well as to extend the models where additional

densities were present in the cryo-EM map. The resulting atomic model for lobe B is consistent with our CX-MS data and in agreement with previous biochemical studies (Figure 3.1.4).

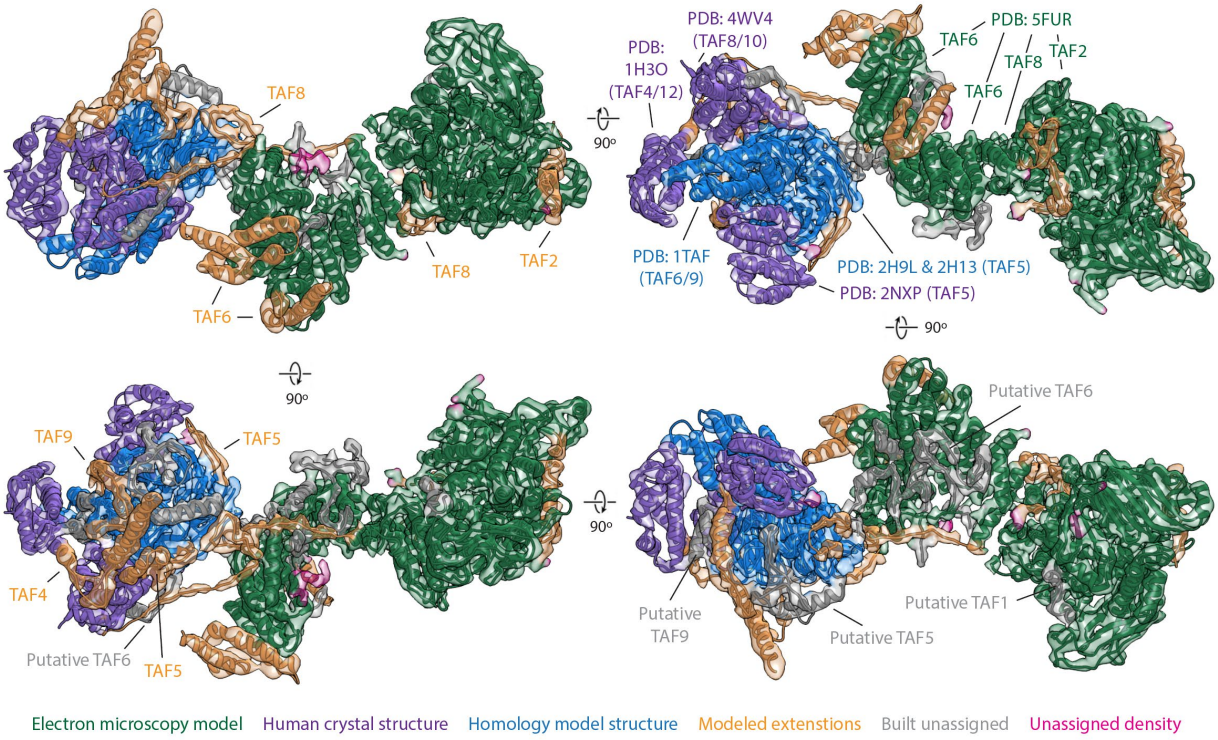

Figure 3.1.4 | Modeling of the BC core of TFIID. TFIID model colored according to the method used to interpret the corresponding density (rigid body docking of previous cryo-EM structure, green; docking of human crystal structures, purple; docking of homology models, blue; manually built extensions of docked protein cores, orange; unassigned poly-alanine fragments, grey; unassigned density, pink).

All the TAFs in lobe B, except for TAF8, have been proposed to exist in two copies within TFIID, suggesting that a similar architecture could exist within the flexible lobe A (Figure 3.1.5)[34,35]. We used a computational strategy based on automated docking of different combinations of TFIID subunits into the lobe A cryo-EM density to generate a complete model of lobe A. The core of the structure is equivalent to lobe B, except for the replacement of TAF8 with TAF3 as the histone-fold partner of TAF10. Additionally, lobe A includes the TAF11-TAF13 HFD pair and TBP. Our placement of TAF11-TAF13 adjacent to the TBP subunit is supported by the presence of chemical cross-links between TAF11 and TBP, as well as in vivo and in vitro studies showing that the HFDs of TAF11-TAF13 constitute the bridge between TBP and the rest of TFIID. Altogether, our structure defines the full architecture of human TFIID, revealing the complete evolutionarily conserved regions of all TAFs and TBP.

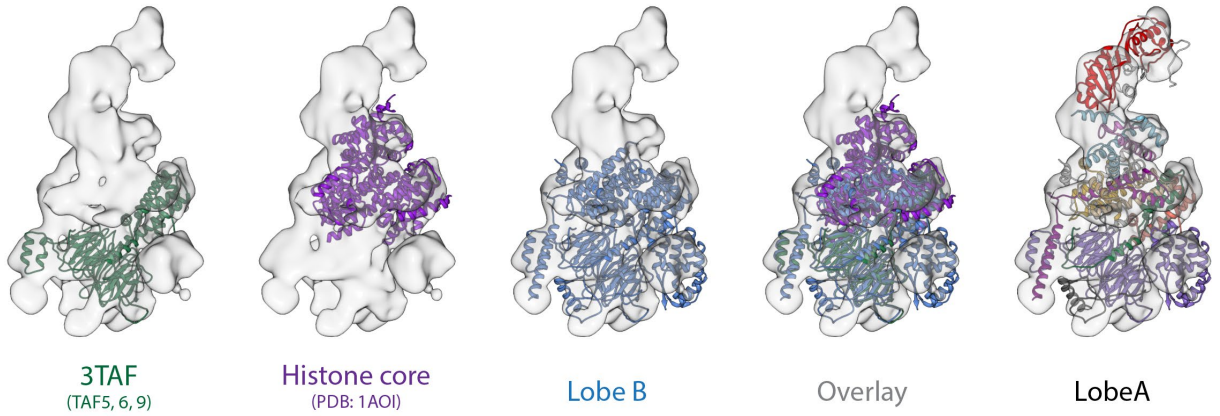

Figure 3.1.5 | Modeling of lobe A of TFIID. Top docking results from UCSF Chimera (using the FITMAP command starting from 128 random orientations and positions). The highest scoring result was used for docking of the 3TAF, histone core, and lobe B models. Overlay - three docked structures shown together reveal extensive overlap components shared between the three models. Lobe A - complete model of lobe A model fitted in the lobe A map.

3.2 TAFs that scaffold TFIID

Our structure of human TFIID shows that the complex assembles around a dimeric yet asymmetric arrangement of TAFs. Two copies of interacting TAF6 HEAT repeat domains are found at the center of the BC core, where they form a dimer with a 3_1 screw axis symmetry that bridges lobes B and C. The N-terminal HFDs of each copy of TAF6 are then separated between lobes A and B, and thus, TAF6, through the flexible connection between its HFD and HEAT repeat domain, tethers the entire complex together (Figure 3.2.1). This TAF6 connection is maintained throughout the various conformational states of TFIID. The HFD of TAF6 forms a heterodimer with the HFD of TAF9, which interacts with the WD40 and NTD2 regions of TAF5. The TAF6-TAF9 HFD pair then forms a tetramer with the TAF4-TAF12 HFD pair, and together these five subunits (TAF5, -6, -9, -4, -12) define the TAF subcomplex that is present in two copies within TFIID, one each in lobes A and B (Figure 3.2.2). The existence of a dimeric TAF-containing subcomplex has been previously proposed based on in vivo knockdown and in vitro biochemical studies[34,36]. However, the structure within the native TFIID complex does not exhibit the symmetry previously proposed for a reconstituted subcomplex containing the same subunits, likely due to the presence of additional symmetry-breaking TAFs in the fully formed, native complex[34].

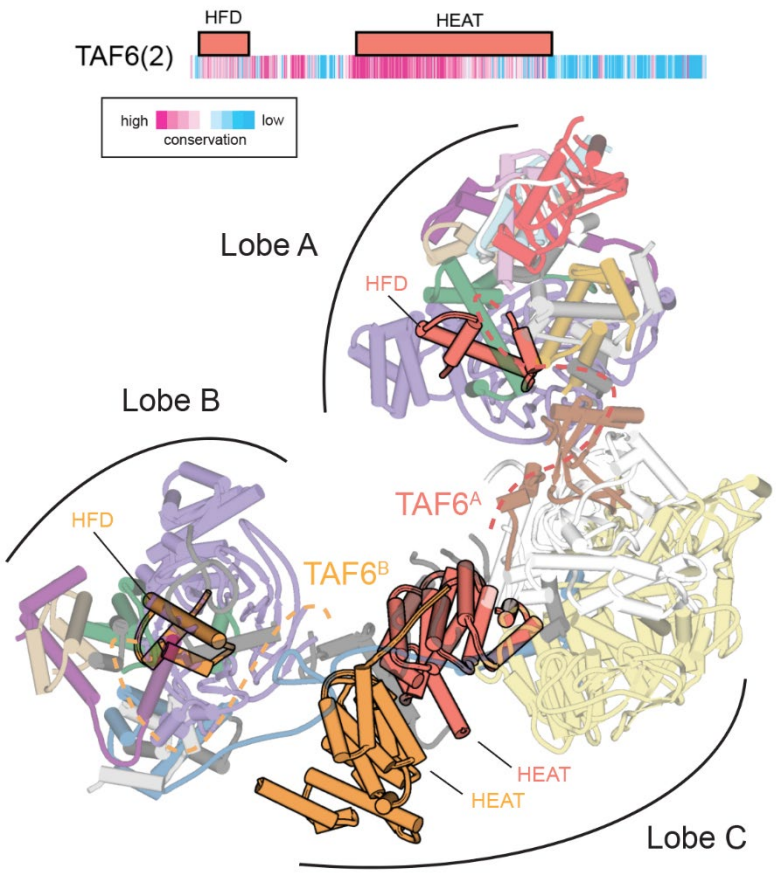

Figure 3.2.1 | Structural organization of human TFIID. Domain organization of TAF6, with sequence conservation colored according to ConSurf scores (top). Model of TFIID with the TAF6 dimer highlighted (bottom). The dimer of TAF6 HEAT repeats is centrally located within the complex. Dashed lines are shown connecting the TAF6 HEAT domains with their corresponding HFDs in lobes A and B.

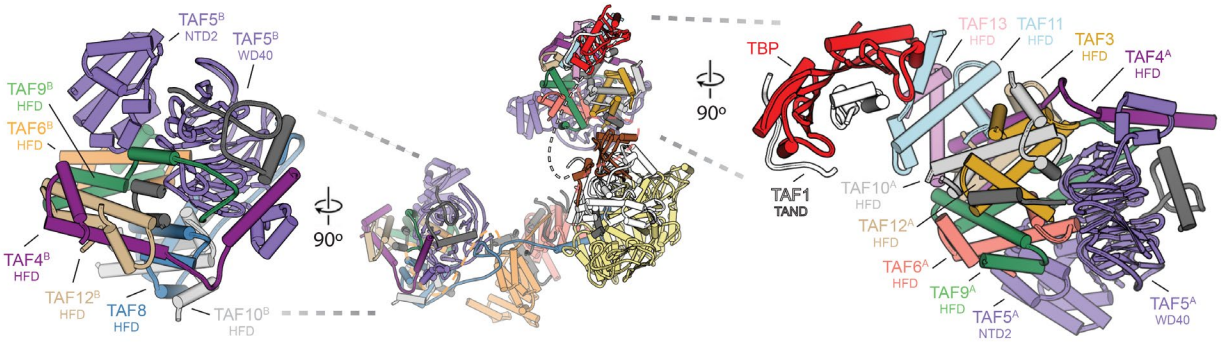

Figure 3.2.2 | Comparison of lobe A and lobe B. Model of TFIID (center) and close-up views of lobe B (left) and lobe A (right).

The two sets of TAFs (-4, -5, -6, -9, -12) shared between lobes A and B act as a base for the assembly of the rest of each lobe. In lobe B, a hexamer of HFDs is formed by the TAF8-TAF10, TAF6-TAF9, and TAF4-TAF12 HFD pairs. In lobe A, the TAF3-TAF10 and TAF11-TAF13 HFD pairs form an octamer-like structure with the TAF6-TAF9 and TAF4-TAF12 HFD. Though the presence of higher-order histone-fold assemblies had been predicted to exist within TFIID, such a structure had not been visualized until now. It has been proposed that these nucleosome core like

structures may be involved in interaction with DNA and promoter binding[36–40]. However, the surfaces of lobes A and B lack the large positively charged patches observed in the nucleosomal histone octamer (Figure 3.2.3). The TAF6-TAF9 HFD pair that was proposed to interact with the downstream DNA is located far from the DNA in the IIDA-SCP complex[18,40]. We instead propose that HFDs serve as a structural scaffold within TFIID.

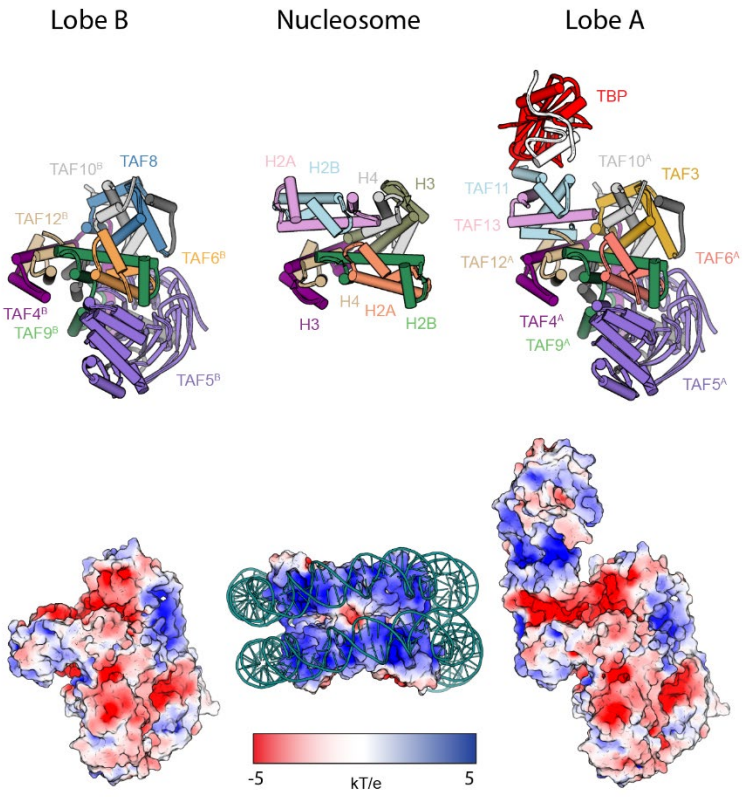

Figure 3.2.3 | Comparison of lobe A and lobe B with the nucleosome core particle. Top: Rendering of the TFIID lobe A and B HFDs aligned to the structure of the histone octamer-containing nucleosome core particle (PDB ID 1AOI). Bottom: Electrostatic surface potentials (generated using ABPS in PYMOL; blue: positive electrostatic potential; white: neutral; red, negative electrostatic potential). The surface of the nucleosome core particle also shows the DNA wrapping around the histone octamer.

The difference in the flexibility of lobes A and B is likely due to the presence of TAF8 in lobe B, which stabilizes its connection with lobe C (Figure 3.2.4). In our model, the highly conserved middle region of TAF8 (residues 130 to 235) snakes through the BC core, interacting extensively with TAF2 and TAF6. Extending from its N-terminal HFD, the TAF6 interacting domain (6iD) of TAF8 forms a bridge between the WD40 of TAF5 in lobe B and the first of the HEAT repeats of TAF6. The long helix of the TAF2-interacting domain (2iD) of TAF8 then bridges the second TAF6 HEAT repeat and TAF2, and then TAF8 folds onto the surface of the TAF2 APD, effectively anchoring TAF2 to the rest of the complex. This network of interactions among TAF8, TAF6, and TAF2 is consistent with previous biochemical studies[34,41].

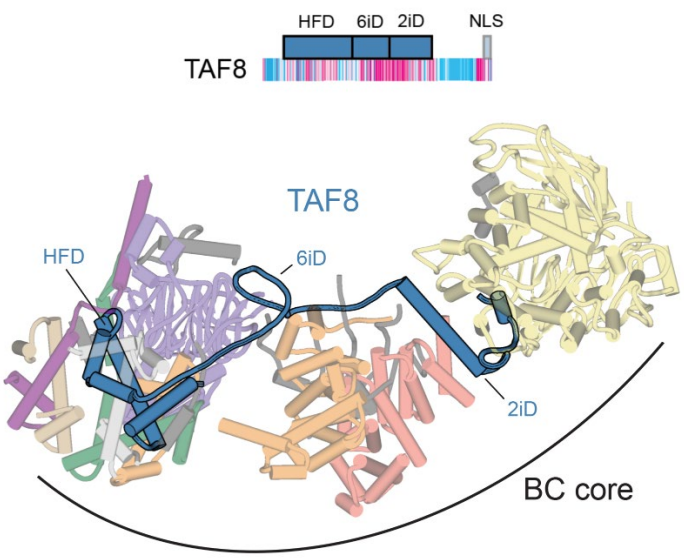

Figure 3.2.4 | TAF8 stabilizes the BC core of TFIID. Domain organization of TAF8, with sequence conservation colored according to ConSurf scores (top). NLS, nuclear localization sequence. Model of the BC core of TFIID with TAF8 highlighted (bottom).

3.3 Lobe B stabilizes upstream promoter binding

Our structural studies indicate that the function of lobe B is to stabilize the upstream DNA and bind TFIIA. Both functions involve the highly conserved C terminus of TAF4 (Figure 3.3.1). The HFD of TAF4, comprising helices α1 and α2, is followed by a large loop and a helix (α3) that interacts with the WD40 of TAF5. Docking of the lobe B structure into the IIDA-SCP map reveals that the highly conserved loop between α3 and a fourth helix in TAF4 (α4) contacts the promoter DNA just downstream of the TATA box. This loop has previously been shown to bind DNA in vitro, and in TAF4$^{-/-}$ human fibroblast cells, stable expression of a TAF4 mutant lacking this loop results in the down-regulation of a subset of genes[42]. From there, α4 continues toward the TBP-TFIIA density and is likely involved in TFIIA recruitment and the stabilization of the TFIIA-TBP-DNA module, in agreement with previous data[43]. The docking of lobe B into the IIDA-SCP map also revealed that the four-helix bundle of TFIIA likely contacts the first helix-turn-helix motif of the TAF12 HFD. Thus, we propose that TAF4 and TAF12 within lobe B act to promote the binding of TBP to the upstream DNA by directly contacting both the DNA and the TFIIA-TBP module (Figure 3.3.2). Therefore, the BC core of TFIID appears to act as a molecular ruler, placing TBP at a defined distance from the downstream promoter elements. This role suggests that maintaining a rigid connection between lobes B and C is important for correctly positioning TBP with respect to the TSS, which in human core promoters are separated by ~30 base pairs (bp)[44,45].

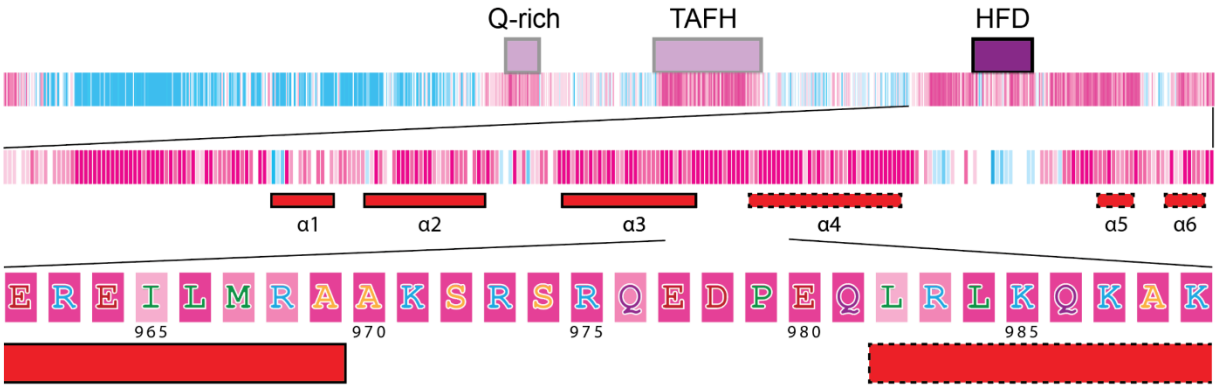

Figure 3.3.1 | Domain map of TAF4. Domain organization and sequence conservation of TAF4 according to ConSurf scores. The first level shows the domain organization of TAF4.The second level zooms in on the C terminus and shows the secondary structure [solid outline corresponds to observed secondary structure and dashed outline to the predicted secondary structure based on PSIPRED results (a4 is not visible in the apo-TFIID structure but becomes ordered upon interaction with the DNA)].The third level shows the amino acid sequence of the loop between helices 3 and 4, which contain several conserved, positively charged residues that could be contacting the DNA.

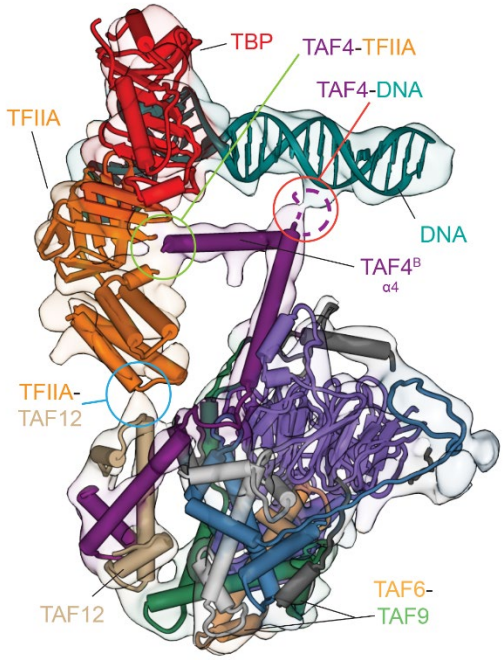

Figure 3.3.2 | Upstream promoter binding stabilized by lobe B. Lobe B of the promoter bound TFIID highlighting the loop between helices 3 and 4 as it contacts the DNA (circled in red), helix 4 continuing toward the TFIIA and TBP (circled in green), and the interaction between the TFIIA and TAF12 (circled in blue).

3.4 Relieving the inhibition of TBP

By combining the mapping of TBP positions through the various states of promoter binding with previous biochemical and structural studies, we propose a model of how TBP within TFIID would transition from being inhibited to being DNA engaged (Figure 3.4.1). We propose that in the canonical and extended states, TBP is bound by the TAND of TAF1 and by TAF11-TAF13 within lobe A, both of which have been found to inhibit TBP from binding DNA[46,47]. The N

terminus of the TAF1 TAND (TAND1) interacts with the DNA-binding cleft of TBP, whereas TAND2 binds the outer surface of TBP where several different TBP-interacting factors are known to interact, including TFIIA[48]. In the scanning state, we propose that DNA displaces TAND1 and interacts with the cleft of TBP, but that the DNA remains in a linear, unbent form, owing to a lack of defined DNA-TBP interaction, in contrast with what is seen for the bent DNA-TBP interaction. In the rearranged state, TFIIA would displace TAND2, releasing the connection between TAF1 and TBP and stabilizing the connection between lobes A and B[48–51]. Finally, in the engaged state, TBP forms a stable complex with bent DNA, which causes the connection between TBP and TAF11 to break and TBP to release from lobe A. This last step of lobe A release is essential for recruitment of TFIIB and for the assembly of the PIC, as it opens the surface on TBP for TFIIB binding[52].

Figure 3.4.1 | Transition of TBP from its inhibited state to promoter bound. From left to right: Models of TBP bound by the inhibitory domains of TAF1 TAND and TAF11, first TAND1 is released for promoter DNA, then TAND2 for TFIIA, TAF11 and TAF13 upon DNA binding and bending and finally TFIIB recruitment.

Though TBP binds the TATA box sequence with the highest affinity of any DNA sequence, it has been observed to be a relatively indiscriminate DNA binder[53,54]. The mechanisms of TBP inhibition within lobe A effectively represent an important role of TFIID as a TBP chaperone, stopping TBP from nonspecifically engaging with DNA outside of gene promoters, and therefore preventing aberrant PIC assembly and erroneous transcription initiation[54]. We propose that, at the same time, the architecture and dynamics of TFIID facilitate the proper loading of TBP at core promoters by progressively releasing those inhibitory interactions with TAFs, and, as explained in Chapter 5, strategically positioning TBP onto the upstream DNA.

Chapter 4

Mechanism of TBP loading by TFIID

4.1 Mechanism of TBP loading

Superposition of the five conformational states of TFIID—canonical, extended, scanning, rearranged, and engaged—illustrates the range of motion TBP experiences with respect to the BC core during the steps leading to full promoter engagement[28,29]. The distance that TBP travels between these states is approximately 130, 40, 30, and 50 Å, respectively, and follows a curved path that directs TBP toward the upstream DNA. Taken together, these structures suggest a stepwise mechanism of TBP loading onto the promoter and the consequent recruitment of the rest of the PIC (Figure 4.1.1). In the first step, TAF1-TAF7 and TAF2 in lobe C bind to downstream DNA. This initial DNA binding facilitates the positioning of the TATA box where it can be reached by TBP as it travels with the mobile lobe A, thus helping the upstream DNA outcompete the inhibitory TAND1 from the cleft of TBP. In the second step, TFIIA displaces TAND2 from TBP and likely stabilizes the upstream DNA through its interaction with lobe B. In this way, the rearranged state constrains the position of lobe A and facilitates TBP binding to the upstream DNA. In the third step, TBP fully engages the promoter DNA, bending it and simultaneously causing a steric clash between the DNA and TAF11 that results in the release of TBP from the rest of lobe A. In the fourth step, TFIIB recognizes the fully engaged TBP-DNA complex and recruits Pol II-TFIIF.

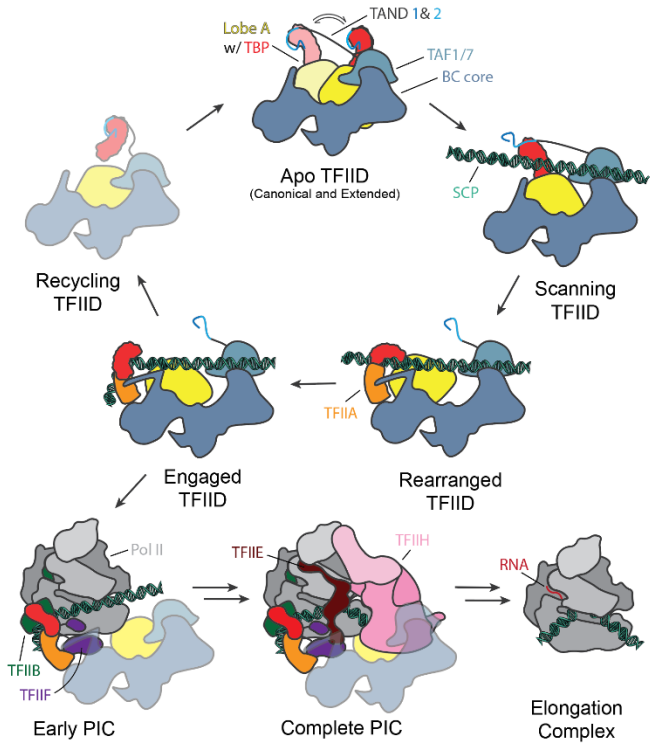

Figure 4.1.1 | Mechanism of promoter binding. Cartoon schematic for the process of TBP loading onto promoter DNA by TFIID, with subsequent PIC recruitment, assembly, and progression to the elongation complex.

4.2 Regulatory roles of chromatin and activators

In vivo TFIID recruitment to the core promoter is aided by gene-specific activators and chromatin marks. Promoters are enriched in certain posttranslational modifications of histones and in histone variants that distinguish them from the rest of the genome[55]. Trimethylation of lysine 4 on histone H3 (H3K4me3) and acetylation of H3 and H4 are especially enriched on the +1 nucleosome (the first nucleosome downstream of the TSS), located ~50 bp downstream of the TSS. TFIID recognizes H3K4me3 through the plant homeodomain (PHD) of TAF3 and the diacetylated H4 via the TAF1 double bromodomain (DBD)[56–62]. A model of the downstream promoter extended with a +1 nucleosome shows how these domains, which our studies indicate are flexibly tethered to the core of TFIID, would be oriented toward the +1 nucleosome in the canonical state of TFIID, suggesting a mechanism of TFIID recruitment by the modified +1 nucleosomes of activated genes (Figure 4.2.1).

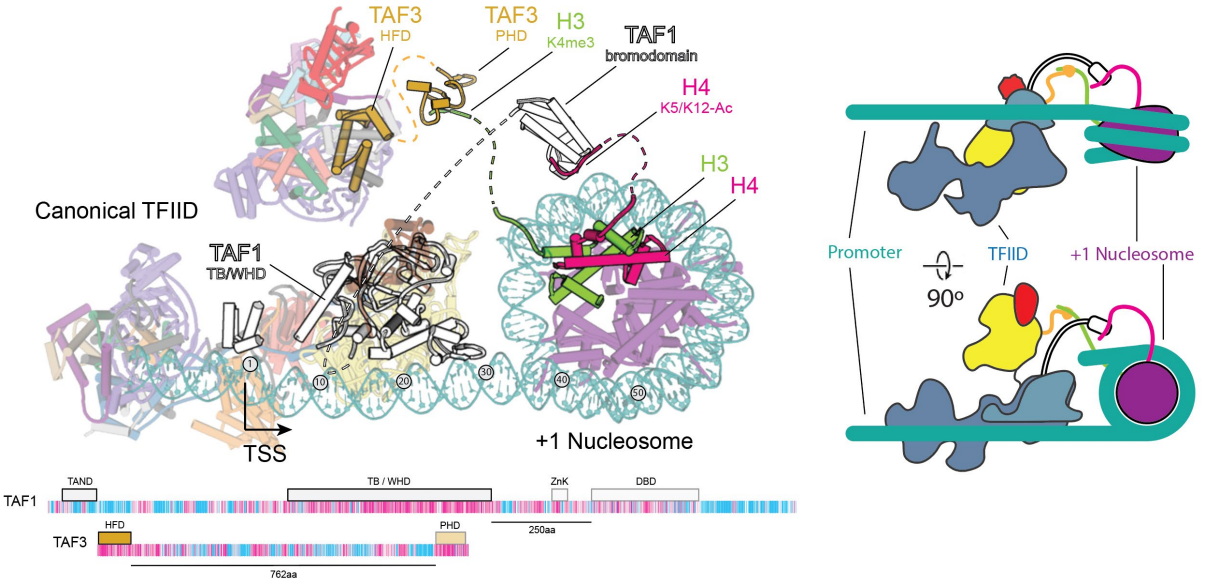

Figure 4.2.1 | Model of TFIID recruitment. Model of TFIID bound to the promoter including a +1 nucleosome. The model is compatible with the binding of flexible histone tails of H3 and H4 to the PHD [PDB ID 2K17] of TAF3 and the bromodomain of BRD2 [PDB ID 2DVR], a homolog of the DBD of TAF1, respectively. Dashed lines indicate the connections between domains contained in the models of TFIID or the nucleosome, with the flexible domains that bridge the two. Domain architecture maps of TAF1 and TAF3 showing the distance between the structured domains modeled within TFIID and the domains that contact chromatin. A cartoon model of TFIID binding to the +1 nucleosome is shown to the right.

Transcriptional activators determine cellular fate by directing the transcription of genes controlling development, differentiation, stimulus response, growth, and maintenance of homeostatic balance[2]. Though many activators have been shown to interact with different TAFs, the strongest evidence has been shown for binding of activators through the conserved glutamine-rich and TAFH domains of TAF4 within its long and flexible N terminus[63–65]. A model generated by extending the upstream DNA in the TFIID rearranged state shows how both

copies of TAF4 are positioned toward the upstream proximal promoter [which is known to remain cleared of nucleosomes and act as a binding site for transcriptional activators[58]] so that they can interact with an activator via their flexible N-terminal domains. This model suggests that transcriptional activators may play a dual role in TFIID recruitment to the promoter, as well as in promoting TBP engagement by stabilizing the rearranged state of TFIID (Figure 4.2.2).

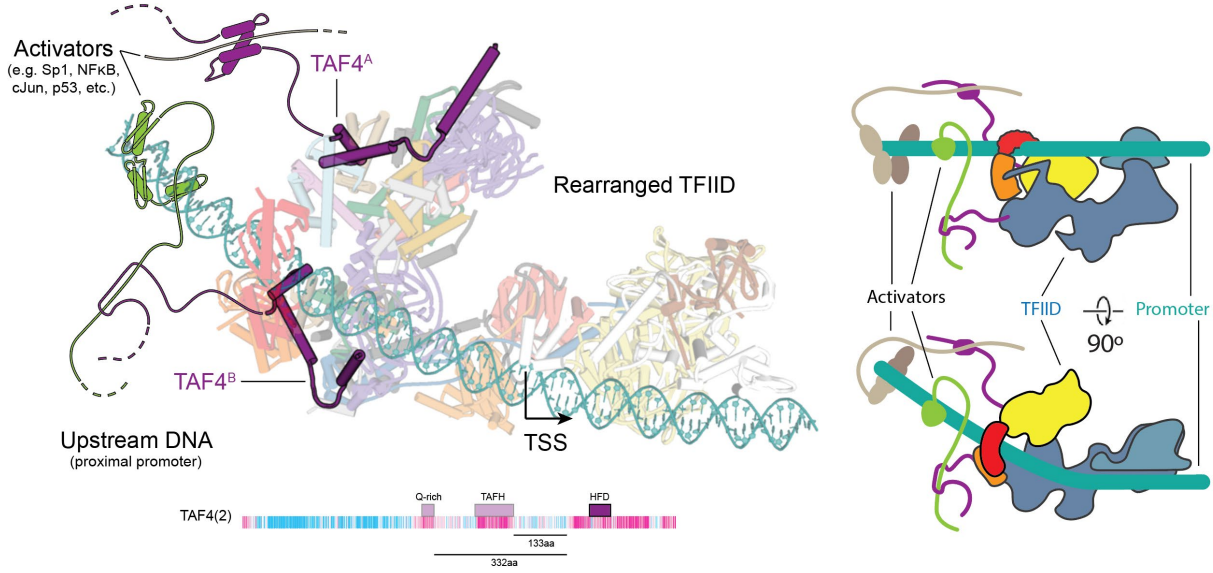

Figure 4.2.2 | Model of TFIID activation. Model of TFIID bound to the core promoter with bound activators at the upstream proximal promoter region. Activators are contacting the N terminus of TAF4 that contains activator interacting regions, like the glutamine-rich and TAFH domains. Domain maps of the highlighted TAFs illustrate the distance between the domains that were part of the TFIID model (solid) and those domains that were not observed (transparent). Distances between the conserved C terminus and the domains that contact activators (TAFH and glutamine-rich) are shown below the domain map. A cartoon model of TFIID binding to activators is shown on the right.

Chapter 5

Implication on the assembly of the PIC

5.1 Interaction of TFIID with the PIC

The structure of the promoter bound TFIID suggests a potential overlap between the contacts that TAF4 makes with the upstream promoter DNA in the IIDA-SCP complex and those established by the TFIIF winged-helix domain within the PIC[8,49]. Additionally, the downstream promoter binding regions of TAF1 and TAF2 were also found to clash with Pol II in the closed PIC complex, and the path of the downstream promoter in the closed PIC is bent compared with the more linear path observed in the IIDA-SCP complex (Figure 5.1.1)[29]. Thus, significant structural rearrangements in TFIID must occur during PIC assembly and transcription initiation, opening the question of whether TFIID can remain promoter bound throughout the transcription initiation cycle.

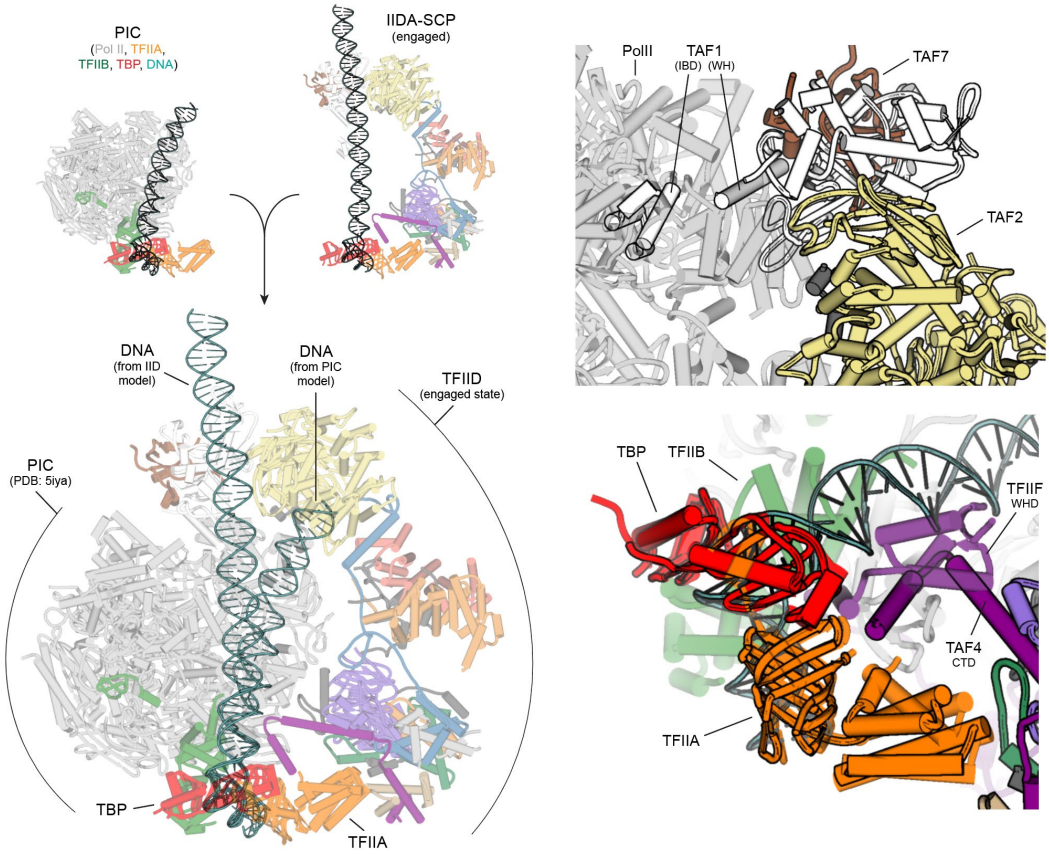

Figure 5.1.1 | Superposition of the PIC and promoter bound TFIID. On the right the models of the human Pol II-PIC (He et al; PDB ID 5IYA (*12*)) and the engaged state of TFIID were superimposed using the shared TBP and TFIIA components as a reference. The DNA (shown with a dark outline) assumes different conformations in the two models, suggesting conformational changes during Pol II loading. On the top right the clashes between TAF1 (IBD and WHD) and Pol II are shown. TFIID is outlined and Pol II is transparent. On the bottom left the clashes between TAF4 (CTD) and TFIIF (WHD) are shown. TFIID is outlined and Pol II is transparent.

5.2 Potential roles of TFIID in PIC assembly

The binding of the TFIIF winged-helix domain in Rap40 and Pol II would displace the TAF4 contact with upstream DNA and the interactions of lobe C with downstream core promoter sequences, respectively. This process could potentially result in the TAFs falling the promoter, unless the interaction between TFIIA and TAF4 was enough to keep TFIID bound or new contacts were to form between TFIID and the PIC at this stage of the assembly (Figure 5.2.1). Although several interactions have been reported between TFIID and other general transcription factors in vitro, it has been shown that upon the addition of Pol II-TFIIB-TFIIF, TFIID remains associated with the promoter only in the presence of activators[21,66–70]. In this potential scenario, TFIID may not remain as part of the growing PIC but could instead bind another TBP to enable formation of a new active complex once the previous complex clears the promoter. Additional experiments will be required to test this model and determine the precise role of TFIID in PIC assembly after TBP loading.

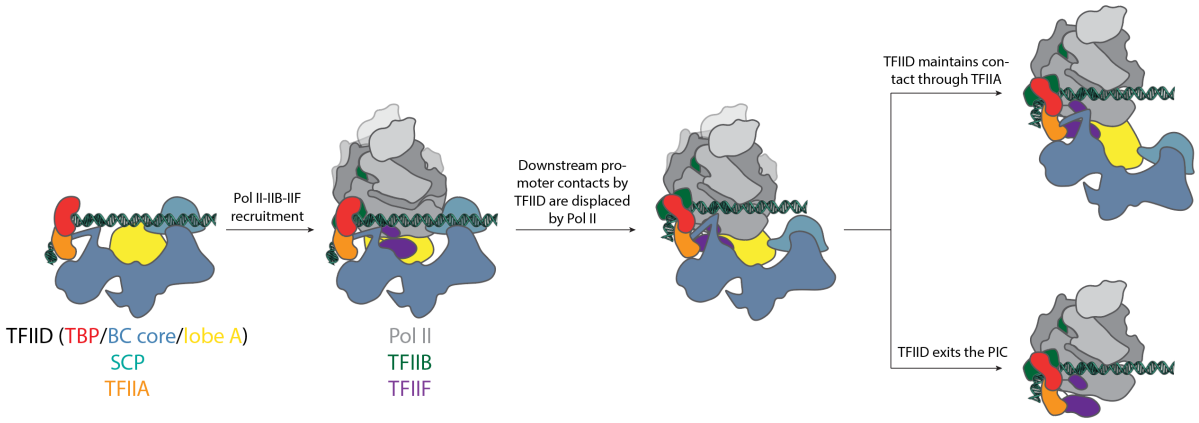

Figure 5.2.1 | Potential models for PIC recruitment. Cartoon model for Pol II loading on to the promoter starting with the engaged state of TFIID. The pathway diverges based on whether TFIID maintains contact with the PIC or not.

Chapter 6

Evolutionary differences in TFIID function

6.1 Role of TFIID in other eukaryotes

It remains unclear whether the mechanism of TBP loading onto the TATA box is fully conserved throughout eukaryotes. While it has been shown that the dramatic flexibility of Lobe A plays an important role in TBP loading for the human complex, and some poorly characterized flexibility is also apparent for yeast complexes, its role in DNA binding has not been considered in the yeast case[28,34,71]. On one hand, the downstream DNA binding residues identified in the human TAF1 and TAF2 are conserved in the yeast subunits but on the other the downstream promoter elements proposed to be recognized by these TAFs have not been identified in yeast. Comparison of the DNA-bound structures of human and yeast TFIID is complicated by the fact that the yeast cryo-EM structure lacks clear density for DNA or for TAF1, TAF2, TBP, or TFIIA, which are all visible in the human promoter-bound TFIID complex and are important elements for DNA binding (Figure 6.2.1)[29,71].

Although the regions responsible for contacting the downstream promoter motifs in human TAF1 and TAF2 appear to be conserved in yeast, downstream promoter elements have not been identified in yeast despite a wealth of genomic data. Thus, it is likely that sequence-specific recognition plays a lesser role in downstream promoter binding in yeast TFIID and that other factors, such as activators and chromatin marks, may play a more substantial role in positioning TFIID.

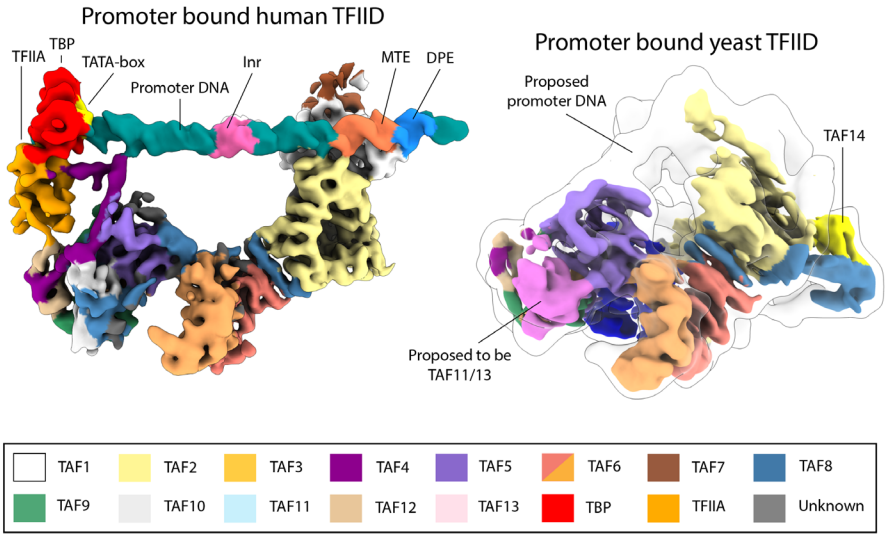

Figure 6.2.1 | Comparison of the yeast and human promoter bound TFIID complex. On the left the cryo-EM reconstruction of human TFIID bound to promoter DNA. On the right the cryo-EM reconstruction of yeast (*Komagataella phaffii*) TFIID bound to promoter DNA. Outlined is the 20Å low pass filtered map to better visualize potential DNA density. Highlighted are the some of the major differences between the human TFIID complex. Maps are colored according to the table shown at the bottom of figure.

6.2 Conservation of the core structure of TFIID

Recent cryo-EM studies of *Komagataella phaffii* TFIID have provided insight into the structure of a yeast TFIID[71]. While the human and yeast structures seem to share a similar architecture for the BC core, there are some clear differences. The BC core of the yeast TFIID appears more compact than its human counterpart, maybe due to possible differences in the spacing or organization of sequence motifs in human and yeast promoters[29,71]. Also, the yeast complex contains an additional structured module that includes part of TAF8 and the yeast specific TAF14. Other structural differences between the two models concern the content of Lobe B. The yeast study suggested that the histone fold pair of TAF11-TAF13 occupies the position that corresponds to TAF8-TAF10 in the human structure[71]. A second difference within Lobe B is the lack of density for the TAF4 helix α3 within the yeast map. While these disparities in Lobe B architecture could reflect true differences between human and yeast TFIID, the apparent anisotropy of the cryo-EM study of yeast TFIID makes it difficult to unambiguously interpret the map.

6.3 Divergency in the regulatory domains of TFIID

Interestingly, the main differences between the yeast and human TFIID complexes do not involve the ordered regions observed in the EM structures, but small domains at the end of unstructured linkers that are too flexible to be visualized, and that include the activator and chromatin binding domains within TFIID (Figure 6.3.1). In humans, the best characterized activator binding motifs are in the N-terminal region of TAF4, while in yeast these appear to be in the N-terminal region of TAF12. Interestingly, TAF4 and TAF12 dimerize, placing their N-terminal domains in roughly the same place with respect to the remainder of the complex. The chromatin binding domains of TFIID, on the other hand, are significantly different between yeast and humans. In human TFIID there are two well characterized histone interacting domains: the TAF3 PHD, which binds H3K4me3 marks, and the TAF1 double-bromo domain, which binds H4 acetyl-lysines[60,61]. Yeast TFIID lacks both domains and instead contains a YEATS domain in the yeast-specific TAF14, which binds H3 acetyl-lysine[72]. This difference likely hints at major differences in how epigenetic marks are being used in humans and yeast, and more generally across eukaryotes. When we extended our analysis of chromatin-binding domains within TAFs to plants, we found that they much more closely resemble yeast, suggesting they represent a more general case for many eukaryotic systems, while metazoans may have evolved a more specialized form of TFIID, maybe to extend the possibilities for gene expression regulation. A more in-depth phylogenetic analysis will be required to better understand how TFIID has changed throughout eukaryotic evolution.

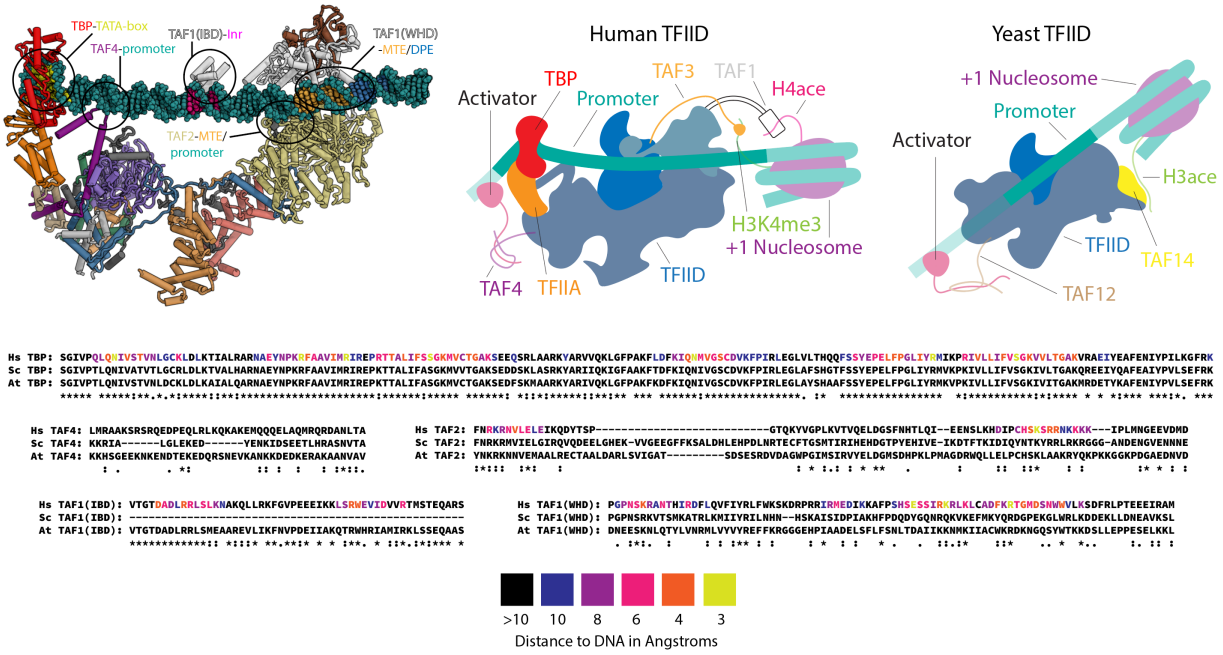

Figure 6.3.1 | Evolutionary comparison of TFIID. Top left, atomic model of human TFIID bound to promoter DNA with subunits colored according to the table at the bottom of Fig 1. Highlighted are some the points of interaction between TFIID and the promoter DNA. Bottom, sequence alignment for the parts of human TFIID that interact with promoter DNA compared to yeast (*Saccharomyces cerevisiae*) and plant (*Arabidopsis thaliana*). Residues are colored based on the distance to the DNA in the model of human TFIID bound to promoter DNA. The TAF4 region that is proximal to the promoter DNA was colored as the sequence register was not able to be determined for that region. Alignment scores for the TAF1 Inr binding domain (IBD) was measured for human and *A. thalana* and leaving out *S. cerevisiae* as is lacks this domain entirely. Top middle, potential points of interaction between human TFIID and chromatin marks on the +1 nucleosome and upstream activators. Top right, potential points of interaction between yeast TFIID and chromatin marks on the +1 nucleosome and upstream activators.

Chapter 7

Graphene oxide coating TEM grids for tilted data acquisition

7.1 Challenges in the sample preparation

Developments in direct electron detector technology and computational image processing tools has made single particle cryo-electron microscopy (cryo-EM) a leading method for the determination of macromolecular structures alongside of X-ray crystallography and NMR[73]. As a single particle technique, cryo-EM has the advantage of being able to determine the structures of large, flexible, and heterogeneous molecules, as each particle can be classified, and individual domains refined separately, all with the need for only small amounts of sample[74,75].

Though cryo-EM opens our ability to look at many new samples, it does come with its own difficulties. A major barrier in obtaining a high-quality reconstruction is preparing a high-quality sample, in which the particles are densely packed, yet homogeneously separated, randomly oriented, and undamaged by the air water interface. To overcome each of these three barriers many methods have been devised and used to varying levels of success.

The easiest way of obtaining a greater number of particles in cryo-EM images, without compromising data quality by lowering the image magnification, is to concentrate the sample. However, if this is not feasible during biochemical sample preparation, the main method of obtaining more particles is to use a support layer over the EM grid that is attractive to the particles and concentrates them by adsorption. The simplest and most popular choice is to use a thin layer of amorphous carbon, but graphene, graphene oxide or a lipid monolayer have also been shown to work[76–78]. Adsorption to a support layer generally allows to increase the number of particles with increasing incubation times. In addition to a simple support layer, affinity grids have been developed with specific interacting groups to capture the molecules in solution[79–83]. Because in the absence of a support layer the molecules may preferentially localized to the surface of the grid, another approach to increase particle numbers in open hole has been to passivate the grid to make it easier for the applied sample to enter the hole[84,85].

The damaging effect of the hydrophobic air-water interface results in the breakdown of many fragile complexes by causing subunits to dissociate and/or proteins to denature[86]. There are several approaches that aim to protect the sample from damage by exposure to the air-water interface, including the use of surfactants such as detergents, stabilizing the protein complexes in buffers containing sugars and/or glycerol or by crosslinking, sequestering the particles away from the air-water interface by binding them to a support layer, or reducing the exposure time to the air water interface by rapid freezing of the sample following blotting[87,88].

Reconstruction of a three-dimensional object from two-dimensional projection images requires multiple views of the complex. While in solution particles are tumbling and in totally random orientations, cryo-EM sample preparation may not preserve this randomicity due to interaction with the air-water interface or a substrate layer. As a result, it is not uncommon that complexes orient in a preferred manner, limiting the number of views present in cryo-EM images. Efforts to overcome this issue include addition of detergents, glow discharging the grid in the

presence of different residual chemical groups, coating the grids with different polymers, or, when all fails, tilting the grid during data collection[30,87–90]. Except for tilting the grid, none of these methods is generally applicable, leading to lengthy screening procedures that may not always prove effective. Unfortunately, tilting has only been shown to be effective for samples prepared using open holes, which may suffer from the previous two problems, and require the use of gold foil grids to reduce the exacerbated beam-induced motion observed for tilted specimens.

7.2 Preparing graphene oxide coated TEM grids

Graphene oxide is a close to ideal support layer for cryo-EM. It produces minimal background in cryo-EM images, is sufficiently hydrophilic to bind biological macromolecules, can be effectively wetted and blots uniformly during sample preparation. However, preparing graphene oxide grids that have single layer that covers every hole has proven challenging with existing methods.

Two main protocols have so far been described to adhere graphene oxide onto the surface of holey carbon grids: the drop cast method and surface assembly method[91,92]. In the drop cast method, a suspension of graphene oxide is directly applied to a glow discharged grid, allowing it to absorb to the surface before being washed with water and dried. In our hands this method leads to minimal coating, making it impractical for automated data collection. The surface assembly method involves forming a monolayer film of graphene oxide on the surface of water, then floating the monolayer onto submerged grids. While this method is effective in coating the grids, the process is time consuming, requires careful testing of the amount of graphene oxide needed to form a monolayer, and tends produce graphene oxide films with high background.

We have developed a modified drop cast method that increases the efficiency of graphene oxide absorption onto the grid's surface by coating the grid with polyethyleneimine (PEI) prior to graphene oxide application (Figure 7.2.1). The surface of commercially available grids is typically hydrophobic, and they have to be glow discharged under a low atmospheric pressure of air or a gas that contain in some part of oxygen in order to render them hydrophilic. During the glow discharge process the residual air becomes ionized, forming ozone that reacts the grid surface and oxidizes it. The oxidized surface is hydrophilic with a slight negative charge. We speculate that the reason why graphene oxide does not effectively bind the grid surface is that both are negatively charged. By coating the glow discharged grid with an additional layer of a positively charged polymer, like PEI, we were able to significantly increase graphene oxide coating. In our protocol, 1.2/1.3 regularly patterned grids with either carbon or gold film are first washed with chloroform to remove any hydrophobic contaminants left on the grid from the manufacturing process. We then glow discharge the grids under low air pressure. These grids are then coated first with PEI and then with graphene oxide. The coating process for both PEI and graphene oxide involves incubating a drop on top of the grid, then blotting away the solution and washing with water before drying the grid. We find that grids prepared in this way have near complete coating, with a limited number of overlapping graphene oxide layers.

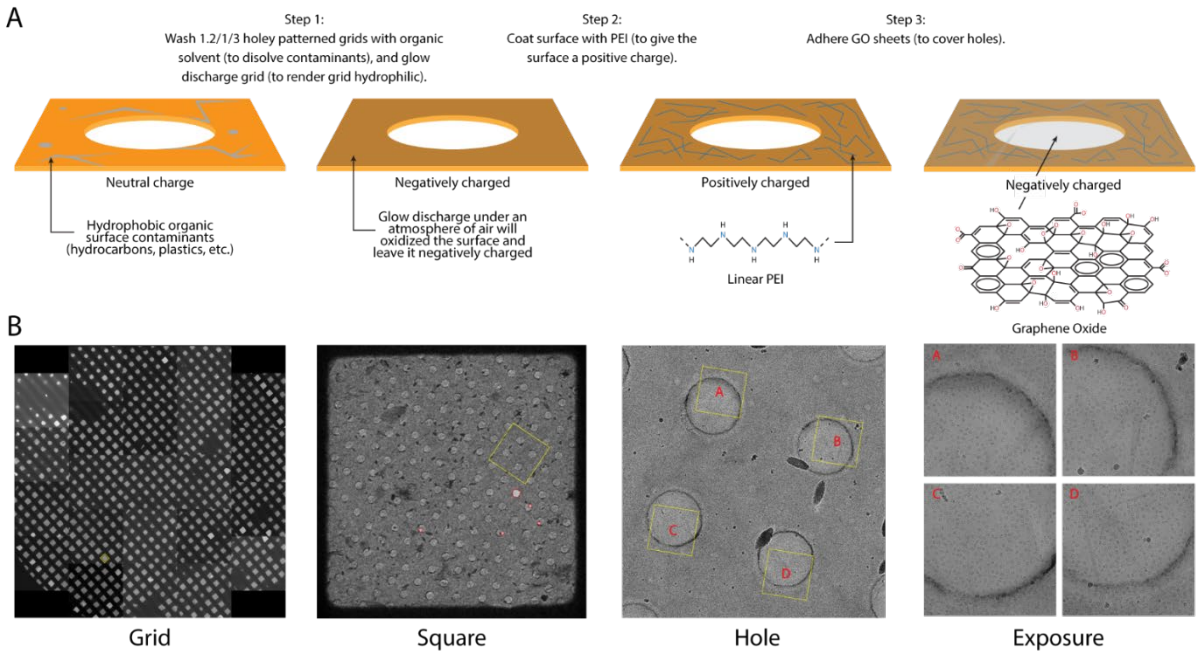

Figure 7.2.1 | Modified graphene oxide coating method. On the top the holey patterned grid preparation and coating work flow. The grids are first washed with organic solvent (i.e. chloroform) and then glow discharged in an atmosphere of air. The grids are then coated with a polycationic polymer (e.i. PEI). Finally, the grids are coated with GO sheets to cover holes. On the bottom an example of a GO coated grid at several magnifications. The yellow squares indicate the region magnified on in the image to the left.

7.3 Tilted data collection with gold foil grids coated with graphene oxide

For many samples preferential orientation is a major barrier in obtaining a three-dimensional reconstruction by cryo-EM. Even when a reconstruction can be generated, the quality of the reconstruction may greatly suffer from anisotropy due to a limited number of views. Of the various approaches used to increase angular distribution for a cryo-EM sample, tilting is the only guaranteed way to produce additional views. However, data collection of tilted specimens has been shown to be effective only when the sample is prepared in open holes using gold foil grids that minimize beam-induced motion. To test the effect of having a support layer over the gold foiled grid, we collected images at 0° and 40° tilts from the differently prepared grids (Figure 7.3.1).

To determine the amount of in-plane (XY) drift, images were collected with a total dose of 40e-/Å^2, fractionated into 40 or 50 frames that were aligned using motioncor2. Twenty images were collected for each grid condition, at both 0˚ and 40˚. The drift between every consecutive pair of frames was then averaged and the standard deviation plotted. From the images collected at 0˚ the XY drift was measured to be approximately 15 Å (12.8 for GO, 15.3 for amorphous carbon [AmC], and 17.0 for open hole [OH]). For the 40˚ tilt images the drift was about double for the GO (35.3 Å) and OH (38.3 Å) grids but was an order of magnitude higher for the AmC (120.7 Å) grid.

To determine the potential cause for the increased drift for the AmC grids, we measured the change in defocus over the exposure for the 0˚ tilt images. Initially we measured the defocus using an individual frame but found the measured fit was low and the change between subsequent frames was large. To increase the signal, we performed rolling averages and found that for images with amorphous carbon 5 frames totaling 4e-/A2 was enough to see a smooth track. For GO images we needed 7 frames totaling 7e-/A2 and for OH images we needed 9 frames totaling 7.2e-/A2. Tracks from twenty images were averaged and plotted and it was found that for GO and OH images there was little to no change in the defocus while the amorphous carbon images increased in defocus by ~100A over a total dose of 40e-/A2 exposure. The large increase in drift for the amorphous carbon coated grids can in part explained by the increase in the defocus seen for these grids. However, given the change in defocus is ~100A and the collected images were tilted 40o the contribution to the XY drift would be only ~60A. The change in defocus also does not explain the increased drift observed in the tilted images for OH and GO images as there is little to not defocus change for these images.

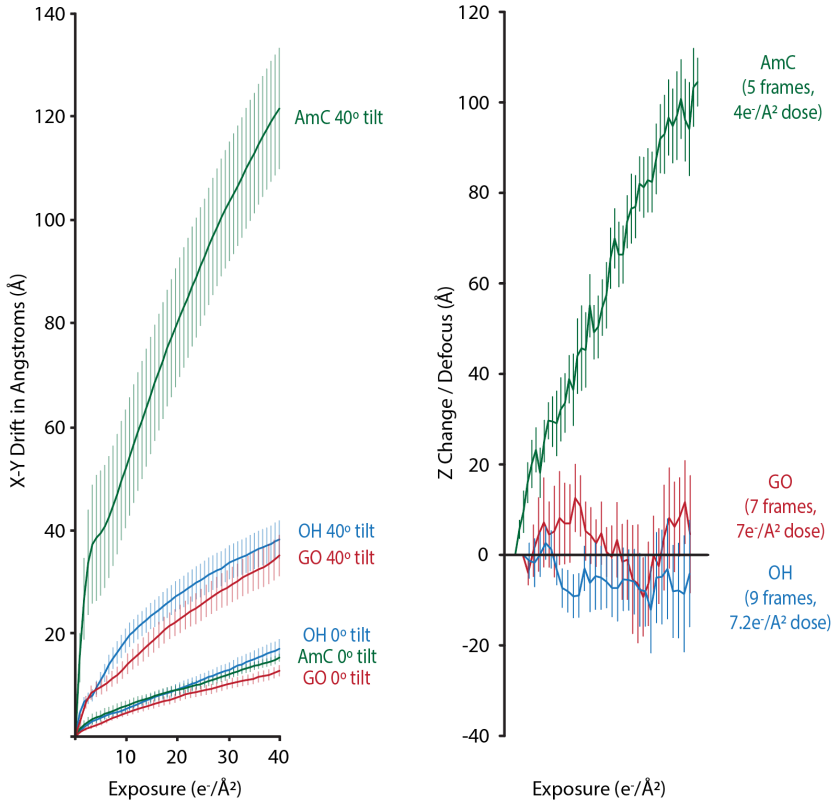

Figure 7.3.1 | Sample drift during titled data collection. On the left graphed are the average drift for image collected on open hole, graphene oxide and amorphous carbon grids at both 0o and 40o. Tracks are averages from 20 images with the vertical lines representing the standard error. On the right the measured change in defocus with exposure. For the collected movies a rolling average of aligned and averaged frames was used to increase CTF signal. The change in defocus from the first sum was then plotted with vertical lines representing the standard error.

Chapter 8

Materials & Methods

8.1 TFIID in-gel sequencing and stoichiometry analysis

TFIID was immuno-purified from HeLa cells as described previously[28]. For gel band analysis, purified TFIID was run on a 4-20% SDS-PAGE gel (Bio-Rad), stained with InstantBlue (Expedeon), imaged using a Gel Doc EZ (Bio-Rad), and analyzed using an Image Lab 3.0 (BioRad). Gel bands were excised, and in-gel digestion was performed using Trypsin Gold (Promega) as is described in ref. [93] to identify protein components. The sample for mass spectrometry was collected using a using a Thermo-Dionex UltiMate3000 RSLCnano liquid chromatography system that was equipped with a C18 column (length: 150 mm, inner diameter: 0.075 mm, particle size: 3 µm, pore size: 100 Å) and a 1-µL sample loop. The LC was connected in-line with an LTQ-Orbitrap-XL mass spectrometer that was equipped with a nano-electrospray ionization source and operated in the positive ion mode (Thermo Fisher Scientific, Waltham, MA). Data acquisition and analysis were performed using Xcalibur (version 2.0.7) and Proteome Discoverer (version 1.3, Thermo) software packages.

8.2 Cryo-EM sample preparation

For cryo-EM sample preparation, TFIID was crosslinked on ice using 0.01% glutaraldhyde for 5 minutes in a buffer containing 20mM HEPES pH 7.9, 0.1 mM EDTA, 5mM MgCl2, 2% glycerol, 1% trehalose, 100mM KCl and 0.01% NP-40. 4µL of sample were then applied to a Cflat CF 2/2 holey carbon gird (Protochips) to which a thin continuous carbon film coated with polyethylenimine (PEI) had been applied to improve preferential orientation bias observed with conventional glow discharged carbon coated grids. The PEI treatment was performed by 4 first glow discharging carbon coated grids using a Cressington 108 Sputter Coater (Cressington Scientific Instruments) for 10 s at 10 mA under 0.5 mbar pressure. 4µL of 1mg/mL PEI (Polyethylenimine HCl MAX Linear MW 40k from Polyscience) buffered with 25mM HEPES pH 7 were applied to the grid for 2 min. The drop was then blotted away, and the grid was washed twice with 4µl of water. The grids were then air dried for at least 15 minutes and used for sample preparation on the same day. For grid preparation, the TFIID sample (see above) was incubated on the grid for 5 minutes at 4°C under 100% humidity in Mark IV Vitrobot (FEI) before blotting with a Whatman #1 for 4 sec at 15N force and then immediately plunge freezing in liquid ethane cooled by liquid nitrogen.

8.3 Cryo-EM data collection

For apo-TFIID, frozen grids were transferred to a 626 Cryo-Transfer Holder (Gatan) and loaded into a Titan Low-base electron microscope (FEI) operating at 300 keV acceleration voltage. Images were recorded on a K2 direct electron detector (Gatan) operating in super-resolution mode at a calibrated magnification of 23,810 x (2.1 Å pixel-1) for the first dataset and in counting mode at a calibrated magnification of 37,879 x (1.32 Å pixel-1) for the second dataset, and using a defocus range of -2 µm to -4 µm in both cases, automated by the Leginon data collection software and monitored using Appion[94,95]. For the first dataset, 30-frame

exposures were taken at 0.6 s per frame, using a dose rate of 9.8 epixel-1 s -1 (1.33 e- Å 2 per frame), corresponding to a total dose of 40 e- Å -2 per micrograph. For the second dataset, 25-frame exposures were taken at 0.372 s per frame, using a dose rate of 7.5 epixel-1 s -1 (1.60 e- Å 2 per frame), corresponding to a total dose of 40 e- Å 2 per micrograph.

8.4 Cryo-EM data processing

For apo TFIID exposure frames were aligned with MotionCor2 software using 8 x 8 patches with 20% overlap for the first dataset and 5 x 5 patches with 20% overlap for the second dataset to correct for local specimen motion[96]. The CTF parameters were estimated using the whole micrograph during the initial data processing stages using Gctf[97]. Particles were picked with Gautomatch (version 0.53, from K. Zhang, MRC-LMB, Cambridge) using gaussian blob templates. All two- and three-dimensional classification and refinement steps were performed within RELION (versions 1.4 and 2.1)[98,99]. The two apo-TFIID datasets were initially processed separately before being combined to for 3D refinement. The initial sets of 996,740 and 1,308,602 particles form datasets 1 and 2, respectively, were subjected to reference-free two-dimensional (2D) classification, and classes that corresponded to ice and junk were removed. This left 681,174 and 820,360 particles for datasets 1 and 2 respectively. This set of particles was subjected to a second round of reference based 2D classification, and only particles that belonged to classes with easily recognizable features were taken forward. The resulting 569,649 and 578,317 particles for datasets 1 and 2, respectively, were then subjected to multi-model 3D classification (using a canonical and an extended state initial reference model) to obtain re-centering parameters. The particles were then subjected to per-particle CTF estimation using Gctf before combining the data from the two datasets. To obtain a high-resolution reconstruction of the BC core of TFIID, a consensus map of all the data was generated focused on the BC core. This was done by using the angular priors form the previous multi-model classification to fix the angular searches for the initial refinement. In this way, the BC cores for the canonical and extended states overlapped and drove the alignment. The refinement was continued using a soft mask around the BC core to focus the refinement, resulting in a reconstruction at 6.1 Å resolution. Masked alignment-free 3D classification was then performed within the BC core region using 5 classes. This resulted in a single good class that had significantly higher resolution estimates than the rest and contained 23.2 % of the particles. The particles form this class were then globally refined and continued with a mask around the BC core, resulting in a reconstruction at 4.5 Å resolution. This focused refinement was then continued using individual masks around lobe B and lobe C, which improved the map for lobe B though the resolution remained at 4.5 Å, and improved the resolution for lobe C to 4.3 Å. Postprocessing and local resolution estimation and filtering were performed using RELION. To classify the various states of lobe A, the combined set of particles was initial split between the canonical and extended state using a multi model 3D classification. Both classes were there further classified using 6 classes for the canonical state and 5 classes for the extended state. One good class for the canonical and extended state had better density for lobe A than the rest. This class was then globally refined and focused refined around lobe A alone. For the canonical state, lobe A refined to 9.75A and the extended state was refined to 10.27A.

8.5 Model building and refinement

We first built a model of the TFIID BC core, which was subsequently used to interpret the density of lobe A, and to generate a complete molecular model of TFIID. Building of the BC core model began with the docking of the available cryo-EM structure of lobe C (5FUR[29]), followed by docking of the crystal structures of the TAF5 NTD2 domain (PDB ID 2NXP[100]) and the HFDs of human TAF8-TAF10 and TAF4-TAF12 (PDB IDs 4WV4[41] and 1H3O[101]). Subsequently, homology models of TAF6-TAF9 HFD (based on PDB ID 1TAF[102]) and the TAF5 WD40 β-propeller domain (based on PDB IDs 2H13[103], 2H9L[104])[105] were generated using MODELLER[106] and placed in the cryo-EM density. The docked models where then manually adjusted to the density in O[107] and COOT[108], and extensions were built to complete the model where continuous density was clearly defined. Continuous density that could be interpreted by placement of a protein chain but could not be unambiguously assigned to a protein subunit of TFIID was interpreted by polyalanine chains (deposition as UNK). The fit of the resulting coordinate model to the cryo-EM map was iteratively manually adjusted and then optimized using the real-space refinement algorithm implemented in PHENIX[109], using Ramachandran, rotamer, Cβ, and secondary structure restraints automatically generated within PHENIX[110]. At an intermediate stage, refinement was performed using 6 macro cycles of global minimization and simulated annealing, the addition of the latter of which resulted in slight improvements of geometry and map fit. This was followed up by correction of small errors and 5 macro cycles of real space refinement without simulated annealing enabled. To avoid overfitting, the amplitude corrected and masked cryo-EM map of the TFIID BC core was filtered according to local resolution, and the spatial frequencies used in the refinement were further globally limited to the nominal 4.5 Å resolution of the cryo-EM map. The b-factors of the resulting coordinate model mirror the local resolution of the cryo-EM, as expected. The refinement statistics of the coordinate model show values typical for the resolution range of our cryo-EM map. The model vs. map FSC curve between the refined coordinates and the cryo-EM map extends to 5.8 Å (according to the FSC = 0.5 criterion ref. [111]), but significant correlation extends to 4.7 Å resolution. This behavior is likely caused by the presence of areas of lower local resolution in the cryo-EM map, which reduces the model vs. map correlation at resolutions beyond 6 Å. Due to the large variation in local resolution and map quality, not all areas of the density could be interpreted at the same level of accuracy. Before coordinate refinement, we therefore removed the side chains from our model in those areas where the register assignment was not supported by the availability of high-resolution crystal structures or evidenced by the features of the density. Register shifts cannot be excluded in these peripheral or lower-resolution regions of the cryo-EM map as well as in the β-sheet regions of TAF2.The coordinate model of lobe B from the BC core refinement was then used to interpret the density of lobe A. First, lobe B and a coordinate model of a nucleosome core histone octamer (PDB ID 1AOI[112]) were fitted into the lobe A density. This fit revealed that three of the four histone fold dimers of the nucleosome core octamer coincided with the coordinates of the histone folds making up lobe B. The fourth histone pair in the nucleosome core octamer was then used for placement of the TAF11-TAF13 histone fold dimer (PDB ID 1BH8[113]). TAF8 from the lobe B model was substituted with the TAF3 histone fold, as both TAFs can be partners of TAF10, and TAF8 had been assigned to be localized to the BC core[114]. Finally, a model of TBP-TAND was generated using (PDB ID 5IYA[8], 4B0A[48], 1TBA[115]) was docked into the tip of Lobe A which done as this region of density would be occupied by TBP in the rearranged state where Lobe A is connected to Lobe B. The TAND region of TAF1 was

included as biochemical evidence indicated that the TAND inhibits TBP form binding DNA[48,50]. These placements were validated by and agree with chemical crosslinking-mass spectrometry data. To generate a complete model of TFIID in the canonical state, the atomic coordinates of the TFIID BC core were combined with the X-ray crystal structure of the human TAF1-TAF7 dimer (PDB IDs 4OY2[116], 4RGW[117]) and the lobe A coordinate model, all of which were docked into the overall TFIID cryo-EM reconstruction.

8.6 Figure creation

Depiction of molecular models were generated using PyMOL (The PyMOL Molecular Graphics System, version 1.8, Schrödinger) and the UCSF Chimera[118] package from the Computer Graphics Laboratory, University of California, San Francisco (supported by National Institutes of Health P41 RR-01081). Movies were generated in Chimera.

References

1. Salzberg, S. L. Open questions: How many genes do we have? *BMC Biology* **16**, 94 (2018).
2. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
3. Reinberg, D., Horikoshi, M. & Roeder, R. G. Factors involved in specific transcription in mammalian RNA polymerase II. *J. Biol. Chem.* **262**, 3322–3330 (1987).
4. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).
5. Cortes, P., Flores, O. & Reinberg, D. Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. *Mol Cell Biol* **12**, 413–421 (1992).
6. Flores, O. *et al.* The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9999–10003 (1991).
7. Holstege, F. C., van der Vliet, P. C. & Timmers, H. T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**, 1666–1677 (1996).
8. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
9. Anthony Weil, P., Luse, D. S., Segall, J. & Roeder, R. G. Selective and accurate initiation of transcription at the ad2 major late promotor in a soluble system dependent on purified rna polymerase ii and dna. *Cell* **18**, 469–484 (1979).
10. Sawadogo, M. & Roeder, R. G. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**, 165–175 (1985).
11. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. & Roeder, R. G. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* **54**, 1033–1042 (1988).
12. Pugh, B. F. & Tjian, R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**, 1935–1945 (1991).
13. Zhou, Q., Lieberman, P. M., Boyer, T. G. & Berk, A. J. Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. *Genes Dev.* **6**, 1964–1974 (1992).
14. Goodrich, J. A., Hoey, T., Thut, C. J., Admon, A. & Tjian, R. Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* **75**, 519–30 (1993).
15. Geisberg, J. V, Chen, J. L. & Ricciardi, R. P. Subregions of the adenovirus E1A transactivation domain target multiple components of the TFIID complex. *Mol. Cell. Biol.* **15**, 6283–90 (1995).
16. Hoey, T. *et al.* Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell* **72**, 247–60 (1993).
17. Verrijzer, C. P., Chen, J. L., Yokomori, K. & Tjian, R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115–1125 (1995).
18. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is

conserved from Drosophila to humans and is recognized by TAF(II)60 of Drosophila. *Genes Dev.* **11**, 3020–3031 (1997).
19. Yatherajam, G., Zhang, L., Kraemer, S. M. & Stargell, L. A. Protein-protein interaction map for yeast TFIID. *Nucleic Acids Res.* **31**, 1252–1260 (2003).
20. Yokomori, K., Admon, A., Goodrich, J. A., Chen, J.-L. & Tjian, R. *Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex*.
21. Ruppert, S. & Tjian, R. Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* **9**, 2747–55 (1995).
22. Chen, J.-L., Attardi, L. D., Verrijzer, C. P., Yokomori, K. & Tjian, R. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**, 93–105 (1994).
23. Guermah, M., Tao, Y. & Roeder, R. G. Positive and negative TAF(II) functions that suggest a dynamic TFIID structure and elicit synergy with traps in activator-induced transcription. *Mol. Cell. Biol.* **21**, 6882–94 (2001).
24. III, F. A., Ladurner, A. G., Inouye, C., Tjian, R. & Nogales, E. Three-Dimensional Structure of the Human TFIID-IIA-IIB Complex. *Science (80-.).* **286**, 2153–2156 (1999).
25. Brand, M., Leurent, C., Mallouh, V., Szlò Tora, L. & Schultz, P. *Three-Dimensional Structures of the TAF II-Containing Complexes TFIID and TFTC*.
26. Grob, P. *et al.* Cryo-Electron Microscopy Studies of Human TFIID: Conformational Breathing in the Integration of Gene Regulatory Cues. *Structure* **14**, 511–520 (2006).
27. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917–922 (2006).
28. Cianfrocco, M. A. *et al.* Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152**, 120–131 (2013).
29. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
30. Patel, A. B. *et al.* Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
31. Stark, H. GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM. *Methods Enzymol.* **481**, 109–126 (2010).
32. Dubochet, J., Ducommun, M., Zollinger, M. & Kellenberger, E. A new preparation method for dark-field electron microscopy of biomacromolecules. *J. Ultrastruct. Res.* **35**, 147–167 (1971).
33. Williams, R. C. *Use of polylysine for adsorption of nucleic acids and enzymes to electron microscope specimen films [RNA polymerase (RNA nucleotidyltransferase)/bacteriophage T7 DNA/poly(dA-dT)/binary complex/negative staini*. **74**, (1977).
34. Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699–702 (2013).
35. Sanders, S. L., Garbett, K. A. & Weil, P. A. Molecular characterization of Saccharomyces cerevisiae TFIID. *Mol. Cell. Biol.* **22**, 6000–13 (2002).
36. Wright, K. J., Marr, M. T. & Tjian, R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl. Acad. Sci.* **103**, 12347–12352 (2006).
37. Hoffmann, A. *et al.* A histone octamer-like structure within TFIID. *Nature* **380**, 356–359

(1996).
38. Selleck, W. *et al.* A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. *Nat. Struct. Biol.* **8**, 695–700 (2001).
39. Leurent, C. *et al.* Mapping histone fold TAFs within yeast TFIID. *EMBO J.* **21**, 3424–33 (2002).
40. Shao, H. *et al.* Core promoter binding by histone-like TAF complexes. *Mol. Cell. Biol.* **25**, 206–19 (2005).
41. Trowitzsch, S. *et al.* Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* **6**, 6011 (2015).
42. Gazit, K. *et al.* TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes. *J. Biol. Chem.* **284**, 26286–96 (2009).
43. Layer, J. H. & Weil, P. A. Direct TFIIA-TFIID protein contacts drive budding yeast ribosomal protein gene transcription. *J. Biol. Chem.* **288**, 23273–23294 (2013).
44. O'Shea-Greenfiel, A. & Smale, S. T. Roles of TATA and Initiator Elements in Determining the Start Site Location and Direction of RNA Polymerase I1 Transcription. **267**, 1391–1402 (1992).
45. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
46. Gupta, K. *et al.* Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *Elife* **6**, e30395 (2017).
47. Liu, D. *et al.* Solution Structure of a TBP–TAFII230 Complex: Protein Mimicry of the Minor Groove Surface of the TATA Box Unwound by TBP. *Cell* **94**, 573–583 (1998).
48. Anandapadamanaban, M. *et al.* High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat. Struct. Mol. Biol.* **20**, 1008–1014 (2013).
49. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
50. Kokubo, T., Swanson, M. J., Nishikawa, J. I., Hinnebusch, A. G. & Nakatani, Y. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol.* **18**, 1003–12 (1998).
51. Bagby, S. *et al.* TFIIA-TAF regulatory interplay: NMR evidence for overlapping binding sites on TBP. *FEBS Lett.* **468**, 149–154 (2000).
52. Nikolov, D. B. *et al.* Crystal structure of a TFIIB–TBP–TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
53. Wong, J. M. & Bateman, E. TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. *Nucleic Acids Res.* **22**, 1890–6 (1994).
54. Coleman, R. A. & Pugh, B. F. Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. *J. Biol. Chem.* (1995). doi:10.1074/jbc.270.23.13850
55. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* **35**, 618–626 (2010).
56. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
57. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional

promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
58. Schones, D. E. *et al.* Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell* **132**, 887–898 (2008).
59. Mavrich, T. N. *et al.* Nucleosome organization in the Drosophila genome. *Nature* **453**, 358–362 (2008).
60. van Ingen, H. *et al.* Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. *Structure* **16**, 1245–1256 (2008).
61. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAF(II)250 double bromodomain module. *Science (80-.).* **288**, 1422–1425 (2000).
62. Umehara, T. *et al.* Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J. Biol. Chem.* **285**, 7610–7618 (2010).
63. Liu, W.-L. *et al.* Structures of three distinct activator-TFIID complexes. *Genes Dev.* **23**, 1510–21 (2009).
64. Hibino, E. *et al.* Identification of heteromolecular binding sites in transcription factors Sp1 and TAF4 using high-resolution nuclear magnetic resonance spectroscopy. *Protein Sci.* **26**, 2280–2290 (2017).
65. Wang, X. *et al.* Conserved region I of human coactivator TAF4 binds to a short hydrophobic motif present in transcriptional regulators. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7839–7844 (2007).
66. Hisatake, K. *et al.* Evolutionary conservation of human TATA-binding-polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8195–9 (1995).
67. Dubrovskayal, V. *et al.* Distinct domains of hTAF1100 are required for functional interaction with transcription factor TFIIFi (RAP30) and incorporation into the TFIID complex. *EMBO J.* **15**, 3702–3712 (1996).
68. Gegonne, A. *et al.* TFIID component TAF7 functionally interacts with both TFIIH and P-TEFb. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5367–72 (2008).
69. Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).
70. Yakovchuk, P., Gilman, B., Goodrich, J. A. & Kugel, J. F. RNA Polymerase II and TAFs Undergo a Slow Isomerization after the Polymerase Is Recruited to Promoter-Bound TFIID. *J. Mol. Biol.* **397**, 57–68 (2010).
71. Kolesnikova, O. *et al.* Molecular structure of promoter-bound yeast TFIID. *Nat. Commun.* **9**, 4666 (2018).
72. Shanle, E. K. *et al.* Association of Taf14 with acetylated histone H3 directs gene transcription and the DNA damage response. *Genes Dev.* **29**, 1795–800 (2015).
73. Kühlbrandt, W. The Resolution Revolution. *Biochemistry* **343**, 1443–1444 (2014).
74. Nogales, E. & Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* **58**, 677–89 (2015).
75. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 (2015).
76. Grassucci, R. A., Taylor, D. J. & Frank, J. Preparation of macromolecular complexes for cryo-electron microscopy. *Nat. Protoc.* **2**, 3239–3246 (2007).
77. Pantelic, R. S., Meyer, J. C., Kaiser, U., Baumeister, W. & Plitzko, J. M. Graphene oxide: A

substrate for optimizing preparations of frozen-hydrated samples. *J. Struct. Biol.* **170**, 152–156 (2010).
78. Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun.* **8**, 629 (2017).
79. Kelly, D. F., Dukovski, D. & Walz, T. Monolayer purification: A rapid method for isolating protein complexes for single-particle electron microscopy. *Proc. Natl. Acad. Sci.* **105**, 4703–4708 (2008).
80. Yu, G., Li, K. & Jiang, W. Antibody-based affinity cryo-EM grid. *Methods* **100**, 16–24 (2016).
81. Benjamin, C. J. *et al.* Selective Capture of Histidine-tagged Proteins from Cell Lysates Using TEM grids Modified with NTA-Graphene Oxide. *Sci. Rep.* **6**, 32500 (2016).
82. Han, B.-G., Watson, Z., Cate, J. H. D. & Glaeser, R. M. Monolayer-crystal streptavidin support films provide an internal standard of cryo-EM image quality. *J. Struct. Biol.* **200**, 307–313 (2017).
83. Wang, F. *et al.* General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. doi:10.1101/657411
84. Meyerson, J. R. *et al.* Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. *Sci. Rep.* **4**, 7084 (2015).
85. Snijder, J. *et al.* Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. *J. Struct. Biol.* **198**, 38–42 (2017).
86. D'Imprima, E. *et al.* Protein denaturation at the air-water interface and how to prevent it. *Elife* **8**, (2019).
87. Noble, A. J. *et al.* Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* **15**, 793–795 (2018).
88. Chen, J., Noble, A. J., Kang, J. Y. & Darst, S. A. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J. Struct. Biol. X* **1**, 100005 (2019).
89. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).
90. Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
91. Bokori-Brown, M. *et al.* Cryo-EM structure of lysenin pore elucidates membrane insertion by an aerolysin family protein. *Nat. Commun.* **7**, 11293 (2016).
92. Palovcak, E. *et al.* A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *J. Struct. Biol.* **204**, 80–84 (2018).
93. Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2007).
94. Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
95. Lander, G. C. *et al.* Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
96. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for

improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
97. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
98. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
99. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**, (2016).
100. Bhattacharya, S., Takada, S. & Jacobson, R. H. Structural analysis and dimerization potential of the human TAF5 subunit of TFIID. *Proc. Natl. Acad. Sci.* **104**, 1189–1194 (2007).
101. Werten, S. *et al.* Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J. Biol. Chem.* **277**, 45502–9 (2002).
102. Xie, X. *et al.* Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**, 316–322 (1996).
103. Couture, J.-F., Collazo, E. & Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat. Struct. Mol. Biol.* **13**, 698–703 (2006).
104. Schuetz, A. *et al.* Structural basis for molecular recognition and presentation of histone H3 By WDR5. *EMBO J.* **25**, 4245–4252 (2006).
105. Malkowska, M., Kokoszynska, K., Rychlewski, L. & Wyrwicz, L. Structural bioinformatics of the general transcription factor TFIID. *Biochimie* **95**, 680–691 (2013).
106. Eswar, N. *et al.* Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinforma.* (2006). doi:10.1002/0471250953.bi0506s15
107. Jones, T. A. & IUCr. Interactive electron-density map interpretation: from *INTER* to *O*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2115–2125 (2004).
108. Emsley, P., Lohkamp, B., Scott, W. G., Cowtan, K. & IUCr. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
109. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).
110. Adams, P. D. *et al. PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
111. Rosenthal, P. B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
112. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
113. Birck, C. *et al.* Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* **94**, 239–249 (1998).
114. Soutoglou, E. *et al.* The nuclear import of TAF10 is regulated by one of its three histone fold domain-containing interaction partners. *Mol. Cell. Biol.* **25**, 4092–104 (2005).
115. Liu, D. *et al.* Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* **94**, 573–83 (1998).
116. Bhattacharya, S. *et al.* Structural and functional insight into TAF1-TAF7, a subcomplex of

transcription factor II D. *PNAS* **111**, 9103–9108 (2014).
117. Wang, H., Curran, E. C., Hinds, T. R., Wang, E. H. & Zheng, N. Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Res.* **24**, 1433–1444 (2014).
118. Pettersen, E. F. *et al.* UCSF Chimera--A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).