UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

A dynamic conformational landscape underlies TFIID's ability to interact with core promoter DNA

Permalink

https://escholarship.org/uc/item/3w56s5fp

Author

Cianfrocco, Michael

Publication Date

2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

A dynamic conformational landscape underlies TFIID's ability to interact with core promoter DNA

by

Michael Cianfrocco

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

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eva Nogales, Chair
Professor James Berger
Professor Robert Glaeser
Professor Michael Levine

Fall 2012

A dynamic conformational landscape underlies TFIID's ability to interact with core promoter DNA

Copyright 2012
by
Michael Cianfrocco

Abstract

A dynamic conformational landscape underlies TFIID's ability to interact with core promoter DNA

by

Michael Cianfrocco

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Eva Nogales, Chair

Proper gene regulation is a problem faced by all organisms, ranging from single-celled bacteria to multi-cellular mammals. In order to respond to the changing environment, thousands of genes across the genome must be controlled in a coordinated manner. The process of evolution has placed these gene regulatory networks under constant selective pressure, which has resulted in the intricate assembly of protein-protein, protein-RNA, and protein-DNA interactions necessary for accurate gene expression. Considering that the mis-regulation of genes results in human diseases ranging from autoimmune disorders to cancers, it has been a major focus of modern molecular biology to understand the molecular mechanisms that determine these complex patterns of gene expression. By advancing our understanding of these processes, we will illuminate both the process of evolution in addition to providing better tools to diagnose and treat human diseases.

In order to study the molecular underpinnings of transcription regulation in multi-cellular organisms, we have focused our investigation on the key human transcription factor known as TFIID. As a 13-14 protein complex, TFIID serves as an important regulatory hub during the process of transcription initiation by simultaneously interacting with distal activators and repressors, promoter DNA elements, and the basal transcription machinery. TFIID integrates these signaling cues to initiate RNAPII loading at specific genes across the genome, thus ensuring the survival of the cell and organism as a whole. Considering that TFIID is conserved throughout eukaryotic life, understanding its ability to license RNAPII transcription stands to provide deep insight into gene regulation across many clades of life.

To address the structural basis for TFIID's ability to communicate with upstream activators and promoter DNA elements, we used single particle electron microscopy to visualize human TFIID's interactions with TFIIA and promoter DNA. First, we discovered that TFIID co-exists in two predominant and distinct structural states differing by a 100Å translocation of TFIID's major sub-domain, lobe A. This result was surprising because this dramatic domain reorganization has been overlooked for the past 10 years. The functionality of this rearrangement was probed by measuring the conformational partitioning of TFIID in

the presence of TFIIA and promoter DNA. The activator TFIIA modulates the transition between these structural states, as the presence of TFIIA and promoter DNA facilitates the formation of a novel rearranged state of TFIID capable of promoter recognition and binding. DNA-labeling and footprinting, together with cryo-EM studies, mapped the locations of the TATA, Inr, MTE, and DPE promoter motifs within the TFIID-TFIIA-DNA structure.

These structural studies have significantly advanced the field of transcription initiation by providing mechanistic and regulatory insight into eukaryotic transcription initiation. The structure of TFIID-TFIIA-SCP represents a conceptual framework for interpreting the past 30 years of biochemical experimentation with purified transcription factors, explaining TFIID's extended footprint (100 bps) on promoter DNA. Additionally, through localization of the Inr, MTE, and DPE promoter motifs, this structure represents an important milestone for understanding these highly utilized promoter DNA elements. Finally, by placing this structure within the context of the extensive rearrangement observed for TFIID, the existence of two structurally and functionally distinct forms of TFIID suggests that the different conformers may serve as specific targets for the action of regulatory factors.

To my dad, a constant source of inspiration.

Contents

List of Figures

Acronyms

AdML	Adenovirus major late promoter
bps	base-pairs
BRE	TFIIB-recognition element
DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
EM	Electron microscopy
GTF	General transcription factor
Inr	Initiator promoter element
kDa	Kilodalton
Mot1	Modifier of transcription 1
MPE-Fe	Methidiumpropyl-EDTA-Fe(II)
mRNA	Messenger RNA
MTE	Motif-ten element
NC2	Negative cofactor 2
RCT	Random conical tilt
RNA	Ribonucleic acid
RNAi	RNA interference
RNAPII	RNA polymerase II
OTR	Orthogonal tilt reconstruction
SA	Streptavidin

SCP	Super core promoter
SNR	Signal to noise ratio
TAF	TBP-associated factor
TBP	TATA-binding protein
TFII-	Transcription factor associated with RNA polymerase II
TFIID	Transcription factor associated with RNA polymerase II, fraction D
TSS	Transcription start site

Acknowledgments

The research presented in this thesis was originally initiated by my inspirational mentor, Eva Nogales. Throughout this work she has served as a constant collaborator, helping to troubleshoot at times of need while also challenging me to propose alternative models to explain our observations. Her efficiency, intensity, and creativity has served as a model for my growth as an independent scientist.

As a complement to her leadership abilities, Eva has put together a fantastic laboratory environment that was essential for this research project. In particular, Patricia Grob taught me everything I needed to know about the electron microscope and data analysis, continuing to teach me new things through the entirety of my Ph.D. Jie Fang has been an amazing biochemist who purified all of the proteins discussed in this thesis, graciously working long hours to achieve extremely pure protein samples. Tom Houweling served as the computer specialist who developed and maintained a computer work environment that maximized the efficiency for everyone in the lab. Throughout automated data collection and acquisition, Gabriel Lander helped to teach me how to use Leginon and Appion while also providing critical feedback to my research. Teresa Tucker has been the life of the lab, keeping in touch with everyone and making sure everything runs smoothly. I have mentored two different undergraduates who helped to keep me focused on my research: Michael Lemus and Sumanjit Mann. Suman spear-headed the biochemical purification of Nanogold, helping me perform key experiments to define the organization of DNA bound to TFIID.

Outside of the lab, I have had the fortunate experience of being a student within the Biophysics Graduate Group. By providing me with a stimulating environment of fellow students from a variety of disciplines, my research has been enhanced through the constant sharing of ideas. The flexibility of the graduate group helped me assemble a Thesis committee that was responsive to my questions while also providing creative feedback to the problems faced during my research. In particular, working with Robert Glaeser on phase-plate cryo-electron microscopy and image simulations has expanding my technical training in ways that I did not anticipate when I joined Eva's lab.

There have been many people whose ideas and technical advice were essential for my research. Weili Liu and Robert Coleman (Albert Einstein College of Medicine) kindly spent many hours with me, discussing biochemical techniques and the implications of my work on TFIID. Robert Tjian also provided critical advice on my research and feedback on our manuscript. Andres Leschziner and Preethi Chandramouli were able to help me implement OTR successfully by providing me with the correct Euler angle conversion routine. The discussions and script sharing with Colin Ophus (NCEM) greatly enhanced my understanding of image formation within the electron microscope while also helping me with image simulations of gold clusters (in preparation).

Despite all of these technical contributions from fellow scientists, perhaps the most important support that I received during my Ph.D has been from my family and friends. Their support has given me the confidence to tackle difficult scientific problems, knowing that I will always have a stable support base to which I can return: Rich, Matt, John, Genessa,

Will, Greg, Andy, Hilary. It goes without saying that my dad has been a constant presence throughout this process, inspiring me to pursue science in a way that will ultimately benefit humanity. My mom and brother have been patient and supportive of my desire to discover more about myself, even though I've been living in California for the past 5 years. Last but not least, my life partner Alice has served as bedrock for me over the last three and a half years, supporting me during the hard-times and celebrating with me during the good times.

Chapter 1

Introduction

The process of evolution and adaptive selection has resulted in cells and organisms that are capable of dynamic interactions with the environment, where information on the surroundings is encoded on short and long time scales. For instance, the physical association of cells with surfaces and other objects requires fast-acting cytoskeletal responses (milliseconds) to maintain cellular integrity. From this information gathered on short time scales, the cell integrates these signals through complex regulatory pathways that amplify or dampening information from the exterior of the cell. These regulatory pathways are mediated through protein, lipid, and nucleic acid interacting factors, as the information is transduced towards decision-making 'centers' within the cell. These 'centers' process the incoming signals to introduce changes in protein and nucleic acid levels, allowing the cell to respond successfully on long time scales (minutes to days) to the changing environment.

As one of the most important areas that controls cellular behavior, the nucleus houses the hereditary genetic material whose expression is modulated in a highly regulated manner. By altering the rate, amount, and synchrony of gene expression across the genome, evolution has selected for gene expression changes that allow cells to respond to diverse environmental stresses in addition to intricate developmental cues. Interestingly, while the number of protein coding genes has remained fairly constant throughout metazoan evolution, the number of regulatory DNA elements has increased dramatically [63]. These and other data suggest that complex cellular choices can be mediated through fine-tuned changes in protein and RNA expression, facilitating species survival.

Given the profound effect that changes in gene expression can have on cellular decisions, understanding the molecular basis for genome-wide regulatory decisions has served as a focal point of modern molecular biology since the discovery of the double helical nature of DNA [115]. To date, most of the proteins involved in regulating the transcription of genes into mRNA sequences have been identified through elegant experimental designs. However, despite this information, a mechanistic understanding of these processes remains limited, and a predicative understanding of that relates protein-nucleic acid interactions to organism phenotype stands as a distant goal. In order to arrive at this deep understanding of cellular decision making, we must have a structural understanding of the proteins involved in order

to appreciate the molecular determinants of gene expression.

1.1 Metazoan gene regulation requires the coordinated assembly of protein complexes on genomic DNA

In order for a cell to respond to specific changes to its environment, DNA sequence-specific proteins serve as the first factors responsible for altering genome-wide changes gene expression. Across the eukaryotic genome there are specific DNA sequences that are recognized by distal-acting enhancers and repressors. These regulatory regions of the genome are bound by proteins that transmit the genetically encoded DNA sequence information into changes in gene expression by affecting RNAPII activity (Figure 1.1). A variety of experimental evidence has demonstrated that these sequence-specific activators and repressors are capable of looping the genomic DNA between the enhancing DNA sequence and the promoter [117]. Through associations between these enhancing regions of the genome and the promoter, the sequence-specific activator can directly affect the loading of RNAPII at the promoter, thus regulating gene expression.

The initiation of transcription by RNAPII requires basal transcription factors known as TFIIA, -IIB, -IID, -IIE, -IIF, and -IIH [107]. These factors assemble onto the core promoters of protein coding genes to form a transcription pre-initiation complex (Figure 1.1) [11, 91]. A sequential recruitment model has been proposed whereby TFIID and Mediator serve as coactivators that facilitate interactions between upstream and promoter proximal factors [15]. The interaction of TFIID with promoter DNA can be further stabilized through a TFIIA-mediated release of the inhibitory N-terminal domain of TAF1 from the concave DNA-binding surface of TBP [2, 27, 66]. This facilitates the interaction of TBP with TATA box DNA, anchoring TFIID onto promoter DNA. The formation of the TFIID-TFIIA-DNA complex is then followed by the binding of TFIIB, RNAPII, TFIIF, TFIIE, and TFIIH to yield the transcriptionally competent pre-initiation complex [107].

The structural transition between promoter recognition and RNAPII recruitment is mediated through a TFIIB-dependent linkage of RNAPII to the TFIID-TFIIA-DNA complex. TFIIB interacts with a stirrup of TBP while also making sequence-specific DNA contacts with TFIIB-responsive elements (BRE) that can be located immediately upstream and downstream of TATA boxes (Figure 1.2) [54, 77]. The binding of TFIIB to the growing pre-initiation complex stabilizes TBP-DNA interactions as RNAPII is recruited through TFIIB's extended N-terminal domain. Through interactions within RNAPII's active site and RNA exit channel [52], TFIIB engages RNAPII in an inhibited state to prepare RNAPII for subsequent promoter melting steps that are mediated by TFIIE, TFIIF, and TFIIH.

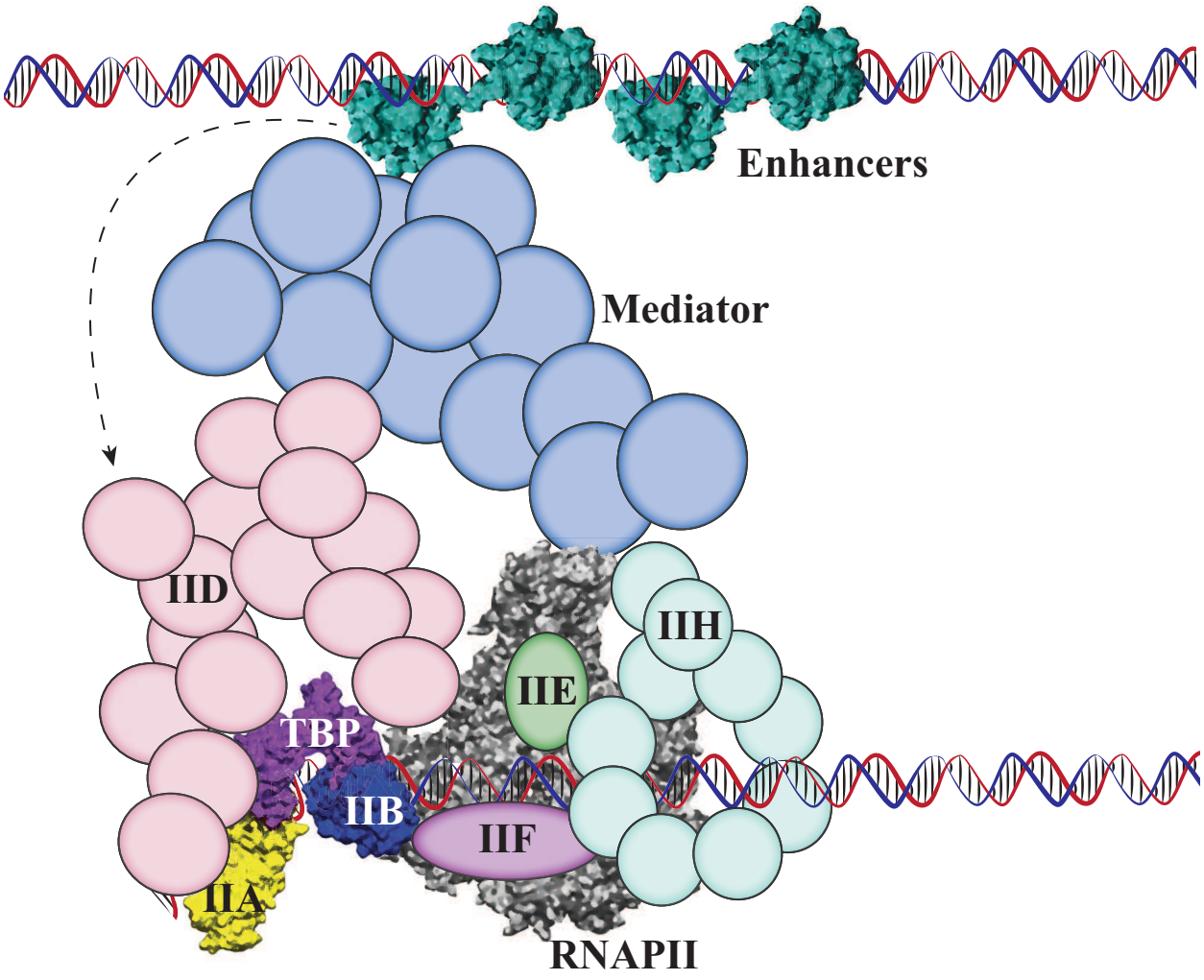


Figure 1.1: Model of eukaryotic transcription initiation. Shown here is a schematic of transcription initiation, where distal-acting enhancers (e.g. activators or repressors) bind to specific DNA sequences to activate or repress transcription activity at a specific gene. These enhancers interact with chromatin remodelling factors (not shown here), Mediator, and TFIID to activate gene expression through long-range protein-protein contacts. Assembling on the core promoter is the pre-initiation complex that comprises TFIID(TBP), -IIA, -IIB, -IIF, -IIE, -IIH, and RNA polymerase II.

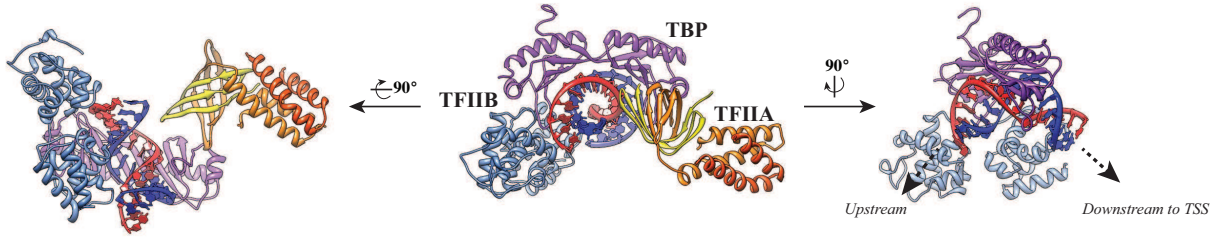


Figure 1.2: TBP-templated assembly of TFIIB and TFIIA on TATA box DNA. Co-crystal structures of TBP-TFIIA-DNA (PDB 1NVP) [5] and TBP-TFIIB-DNA (PDB 1VOL) [77] are shown. Side view (right) of TBP-TFIIB-DNA highlights the TBP-induced kink in TATA box DNA.

1.2 The multi-subunit TFIID complex is essential for regulated transcription initiation

TFIID is a multi-subunit complex that comprises TBP and about 12 to 13 additional proteins known as TAFs [15]. Initial studies with partially purified nuclear extract revealed that the 'D' nuclear extract fraction was necessary for recognition of the TATA box to support *in vitro* transcription by RNAPII [71]. After the identification and cloning TBP [12], Tjian and co-workers performed a series of elegant biochemical experiments that addressed the key differences in activity between endogenously purified TFIID and recombinantly expressed TBP (Figure 1.3) [23, 88]. Utilizing reconstituted *in vitro* RNAPII transcription assays, the activity of TFIID and TBP were tested against a TATA box promoter DNA template with upstream sp1 binding sites (Figure 1.3). The addition of TBP allowed for the specific stimulation of low levels of basal transcription initiation that was further increased by TFIIA, but remained unresponsive to the presence of sp1 (Figure 1.3, Rxns 1 - 4). Surprisingly, unlike recombinant TBP, the endogenously purified TFIID complex was capable of responding to the presence of sp1 by stimulating much high levels of transcription initiation. Furthermore, the addition of TFIIA to TFIID-sp1 resulted in a synergistic coactivation of sp1-stimulated transcription initiation (Figure 1.3, Rxns 5 - 8). These results led to the proposal of the 'coactivator hypothesis' for TFIID [88], where the TAF subunits of TFIID interact with upstream activators to increase the dynamic range of the transcription output.

In addition to mediating contacts with upstream activators, TFIID makes sequence-specific contacts with core promoter DNA elements through the use of TBP and TAFs. While the TATA box is the most evolutionarily conserved core promoter DNA element [46], sequence analysis of metazoan core promoters revealed the presence of additional elements known as Inr [100], MTE [65], and DPE [13] (Figure 1.4). The Inr encompasses the TSS and interacts with TAF1 and TAF2 in a sequence-dependent fashion [17, 114]. Interestingly, a reconstituted TBP-TAF1-TAF2 complex is sufficient to initiate transcription from TATA-Inr containing promoters, suggesting that TFIID may contain a TATA-Inr interacting module that minimally comprises TBP, TAF1, and TAF2.

Working in synergism with the Inr motif, the MTE and DPE motifs are positioned downstream of the TSS (Figure 1.4). Originally identified within *Drosophila* and subsequently characterized within the human system, the MTE and DPE motifs are more commonly used across the human genome than TATA boxes [49]. Through the use of photo cross-linking studies of TFIID with promoter DNA, TAF6 and TAF9 have been identified as the subunits within TFIID that are responsible for interacting with the MTE and DPE motifs [14, 65]. While a crystal structure of the TAF6-TAF9-DNA complex remains elusive, biochemical and crystallographic data indicate that TAF6 and TAF9 heterodimerize through histone-fold domains in a manner that is homologous to histone H3 and H4 [37, 119]. Interestingly, through the use of histone folds, TAF6 and TAF9 can form a higher-ordered complex that comprises TAF6/9/4/12, which is sufficient to interact with DPE-containing promoter DNA [99]. However, even though these subunits of TFIID can assemble into oc-

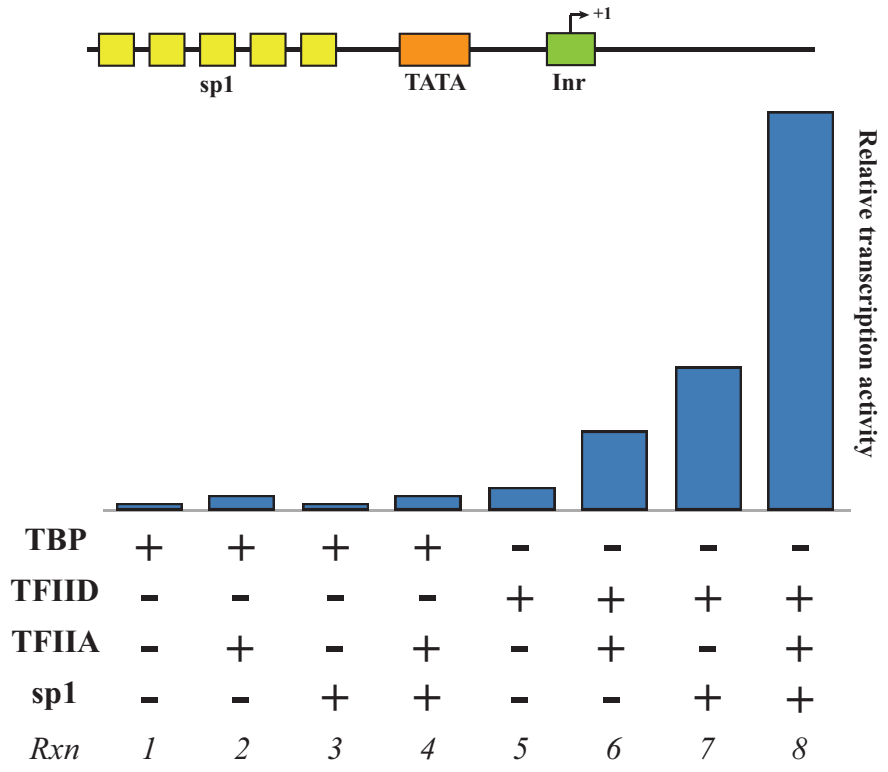


Figure 1.3: TFIID, not TBP, responds to upstream activators *in vitro*. Experimental results adapted from [23, 88]. Recombinantly purified TBP or immunopurified TFIID were used as promoter recognition factors for *in vitro* transcription in the presence or absence of TFIIA and sp1.

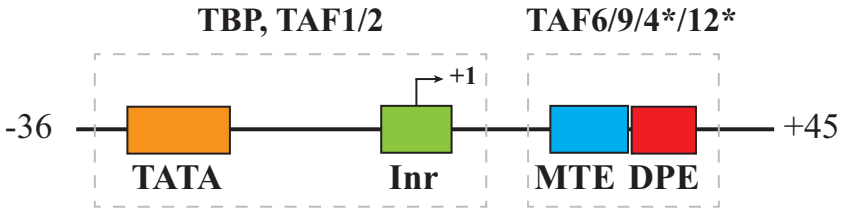


Figure 1.4: Core promoter architecture. Promoter motifs that make sequence-specific contacts with subunits of TFIID are shown relative to the TSS (+1). Dotted lines indicate regions of promoter DNA that interact with indicated sub-complexes of TFIID. Note that for the downstream promoter motifs (MTE/DPE), TAF6 and TAF9 contribute to DNA binding while TAF4 and TAF12 serve as structural support (denoted *).

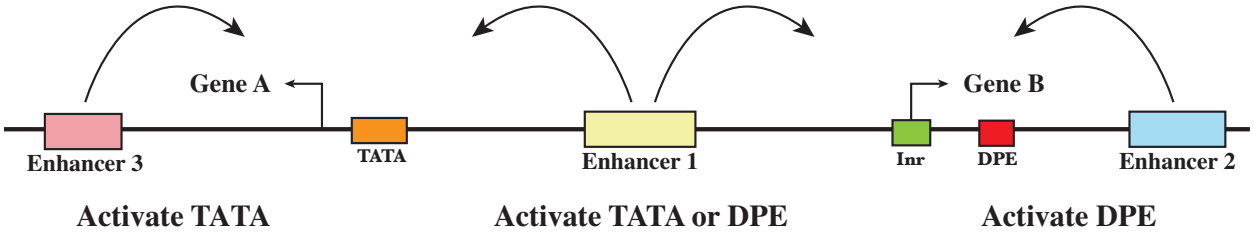


Figure 1.5: Regulation of enhancer-promoter interactions through core promoter elements.

tameric nucleosome-like complexes, it appears that the DNA-binding residues critical for nucleosome-DNA interactions are absent within the histone fold domains TAF6 and TAF9. This results in a TAF6/9/4/12 complex that interacts with promoter DNA in an alternative manner than the well-described nucleosome-DNA system.

These interactions between TFIID and the core promoter elements have been exploited to create a highly active core promoter, termed the 'super core promoter' (SCP). The SCP is capable of high affinity interactions with TFIID through the presence of optimal versions of the TATA, Inr, MTE, and DPE motifs (Figure 1.4) [44]. Through the incorporation of these four promoter motifs, the SCP exhibited the highest affinity of a core promoter for human or *Drosophila* TFIID to date, as measured through *in vitro* transcription and DNase I footprinting [44].

1.3 The core promoter as a regulatory element

The core promoter contributes to the regulatory diversity seen with metazoan genomes, playing an active role in gene regulation [46]. Initial evidence for promoter-enhancer interactions came from studies within the *Drosophila* Hox gene cluster [80]. Careful comparison of enhancer and promoter architecture (e.g. presence or absence of TATA or DPE motifs) allowed the authors to identify two regulatory strategies [80]. One strategy involved a single enhancer that indiscriminately activates genes with TATA or DPE motifs, whereas an alternative strategy made use of a TATA-specific enhancer, where the presence of a TATA box was sufficient to confer activation by the TATA-specific enhancer (Figure 1.5).

These findings have been extended through further studies in *Drosophila* that identified a DPE-specific activator. Inspection of the *Drosophila* Hox gene cluster revealed that many genes contained DPE-containing promoters [45]. After further study, Caudal was identified as the transcription activator responsible for regulating Hox gene expression in a DPE-specific manner. Additional experiments showed that Caudal specifically activated DPE- but not TATA-containing promoters, confirming its specificity. Given Caudal's role as a master regulator of body-plan development, these data suggest that core promoter DNA sequences contribute to the regulatory diversity observed across the metazoan genome (Figure 1.5).

Core promoter-specific enhancers suggest that the architecture of promoter motifs may

be communicated to the enhancers through changes in the structure or activity of the GTFs. Specifically, since the TATA box Inr, MTE, and DPE motifs are recognized by subunits within TFIID, it has been proposed that changes in the architecture of core promoters may impart specific structures on TFIID [80]. Consistent with this model, RNAi knock-down of specific subunits within TFIID and other basal transcription factors within *Drosophila* revealed the presence of TATA and DPE activating factors [39]. These studies revealed that, in addition to positively regulating TATA-containing promoters, TBP could inhibit DPE-dependent transcription. Conversely, two activators for DPE-dependent transcription - NC2 and Mot1 - inhibited transcription from TATA-containing promoters. Therefore, changes in core promoter architecture can impart specific requirements for the assembly of the transcription machinery on promoter DNA.

1.4 Electron microscopy studies of TFIID

Despite the importance of TFIID as a coordinator of transcription initiation, high-resolution structural information has been restricted to crystal structures of a small number of subunits and domains within TFIID [4, 41, 48, 50, 66, 116, 119]. The size and scarcity of TFIID, typically purified from endogenous sources, has restricted structural studies to methods requiring microgram quantities of sample. Single-particle EM has proven to be an indispensable tool for the structural characterization of large multi-subunit complexes, even when sample is available in minute amounts. This technique also has the potential to characterize the structural dynamics of large protein complexes [60].

EM structural studies of TFIID have yielded low-resolution structures (20 to 30 Å) of yeast and human TFIID (Figure 1.6) [1, 8, 24, 31, 61, 62, 68, 84, 83]. A number of these studies suggested the role of conformational flexibility in promoter binding by TFIID, where large sub-domains of TFIID appear to adopt multiple conformational states. Given the low resolution of these structures, the underlying conformational flexibility of TFIID likely limits the resolution due to the computational sorting necessary to describe the intermediate structural states. Recently, several groups have reported single particle EM structures of purified endogenous yeast TFIID bound to promoter DNA. These studies examined the binding of TFIID from *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* to promoters that contain both TATA and Inr sequence elements [24, 83]. Computational methods were used to sort TFIID into distinct conformational states, and additional densities, which were attributed to TATA box DNA, were localized to the surfaces of their structures.

In addition to limiting the resolution of the structures, the low SNR of the images likely limits the accuracy of particle alignments. Cryo-EM analysis of human TFIID suggested the presence of multiple conformational states based upon a 3D variance-based supervised classification strategy (Figure 1.6) [31]. The resulting 2D averages and 3D models likely reflected these conformational differences, although since these data were collected at 200 kV on film under low dose conditions, there may not be a sufficient SNR to assess the validity

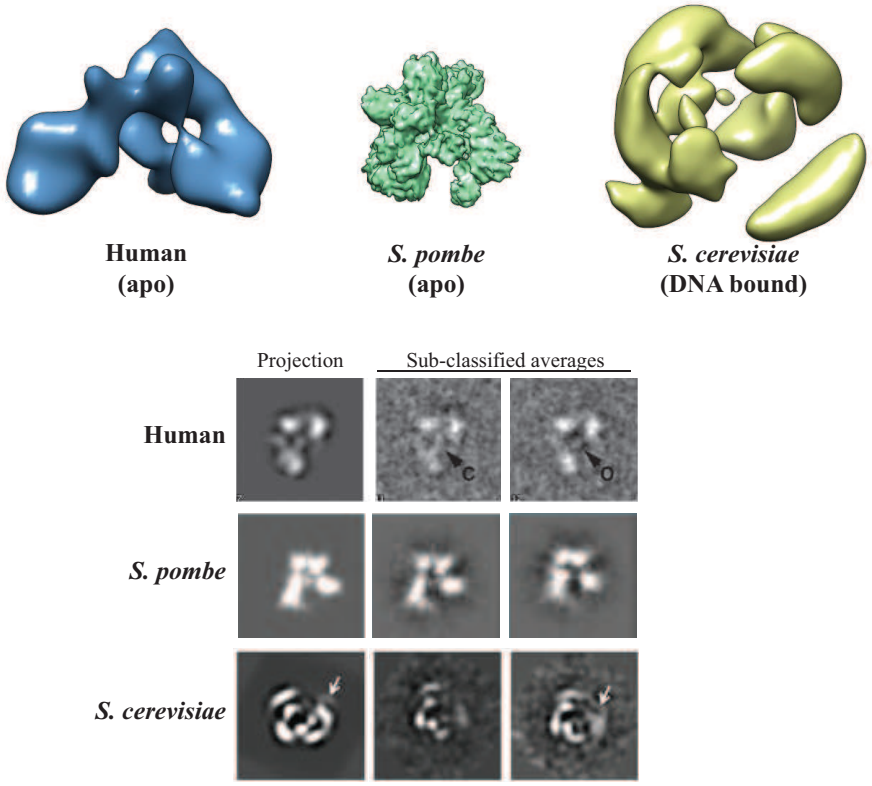


Figure 1.6: Comparison of previously obtained TFIID structures. Shown are 3D models for human (EMDB 1194) [31], *S. pombe* (EMDB 5134) [24], and *S. cerevisiae* (EMDB 5176) [83]. Note that the *S. cerevisiae* structure was obtained from a sample of TFIID-TFIIA-Rap1-DNA. For each 3D model, there are corresponding projections shown alongside sub-classified averages.

of these structures. Importantly, given these limitations, the authors do not over interpret their structures.

On the other hand, the yeast TFIID structures [24, 83] appear to be affected by computational errors that led the authors to over-interpret the resulting 3D models. Obtaining two structural states of *S. pombe* TFIID at sub-nanometer resolution should constitute an important discovery for the field [24]. However, given that there are not obvious alpha-helical densities within this map, the resolution has likely been over-estimated. This incorrect assessment of the resolution is likely due to an over-alignment of the low SNR particles, where successive rounds of alignment with low SNR particles can impose model bias and an over-estimation of map resolution [102]. This suggests that the conclusions drawn from the *S. pombe* structure may not be based on real structural changes within TFIID.

While the *S. cerevisiae* structural analysis of TFIID-TFIIA-Rap1-DNA did not over-estimate the resolution, the final 3D reconstructions exhibit non-physical characteristics indicative of a computational error in the 3D alignment scheme. Throughout the structural work presented by the authors [83, 84], the 3D models exhibit a hard-edge, reminiscent of an 'onion shell' (Figure 1.6). This structural feature of the 3D models is probably due to a 3D mask that has a smaller radius that the radius of the TFIID particles. Analysis of human TFIID in this incorrect fashion has resulted in similar structural features (data not shown). Imposing this incorrect mask during the 3D refinement may also have affected the particle alignments, resulting in a rotationally averaged 3D model. This is seen clearly in the 2D projections of the *S. cerevisiae* DNA bound structure (Figure 1.6) where outer regions of the structure appear to be blurred. These considerations indicate that the structural conclusions proposed by the authors regarding TFIID-promoter interactions may not be accurate, due to incorrect particle alignment strategies.

1.5 Research rationale

In the beginning of my graduate career, the first goal of my project was to pursue a structure of human TFIID bound to promoter DNA. Given that previous work in the lab had identified structural transitions existing within the human TFIID sample [31], we hypothesized that conformational selection may be important for promoter binding. However, after a number of years of failed attempts to determine the 3D structure of TFIID-TFIIA-SCP, we were surprised to discover that the DNA binding configuration for TFIID is within a reorganized structural state. This structural reorganization was then characterized using extensive 2D image analysis, in addition to obtaining an *ab initio* 3D reconstruction. These studies verified that TFIID can move its lobe A (approximately 300 kDa in size) by 100Å across its central channel in a dynamic equilibrium, providing an explanation for years of unsuccessful experiments.

After revealing the presence of two predominant conformational states of TFIID, we discovered that the novel 'rearranged' conformation of TFIID corresponds to a high affinity

DNA binding state. By using gold labelling in combination with multi-model refinements, we defined the organization of the rearranged conformation of TFIID bound to promoter DNA. To provide biochemical evidence for the model of DNA binding, the protein-DNA interactions were probed through DNA footprinting experiments with DNase I and MPE-Fe. By mapping the digestion patterns of wild-type and mutant SCP sequences onto the DNA-bound structure, we were able to show that the rearranged conformation is the DNA binding conformation for promoters of varying architecture. Our data suggest a model in which the distinct conformations of TFIID may serve as targets through which regulatory factors recruit TFIID to specific types of core promoters, facilitating the net-stabilization of TFIID bound to promoter DNA within the rearranged state.

Chapter 2

Human TFIID binds promoter DNA in a reorganized structural state

2.1 Detailed analysis of TFIID reveals dramatic flexibility of lobe A

Lobe A is flexibly attached to a stable core of TFIID that comprises lobes B and C

In previous cryo-EM analyses of human TFIID, the use of 3D variance suggested that conformational flexibility is an intrinsic property of TFIID [31]. Due to the potential effect that this movement could have on TFIID function, we investigated this property more thoroughly by extensive 2D image analysis of negatively stained TFIID samples (Figure 2.1A and Figure 2.2A). The resulting 2D averages showed a variety of different TFIID orientations, as seen previously [31]. However, upon careful analysis, we noticed that a number of the class averages that initially looked like different views of TFIID, actually corresponded to the same orientation (Figure 2.2B). When comparing these averages side-by-side, it becomes clear that there is a rigid sub-structure within TFIID that is shared between all of the class averages. Surprisingly, a flexibly attached lobe is connected to the rigid sub-structure in a variety of conformations (Figure 2.2B). Thus, these class averages indicate that TFIID may exhibit a greater degree of flexibility than previously appreciated.

Comparison of these class averages with a 3D model indicated that the flexible domain is lobe A. The class averages were compared to a model containing all domains of TFIID (Figure 2.3A & C) and a model corresponding only to the BC core (Figure 2.3B & D). In contrast to the apo-TFIID model (Figure 2.3A), the BC core model aligned all of the averages within the same orientation, confirming that these class averages correspond to distinct structural states. Thus, lobe A is the flexible domain that differs in position relative to the rigid sub-substructure that comprises lobes B and C, termed 'BC core.'

To confirm that the conformational flexibility seen for lobe A was not an artifact due to

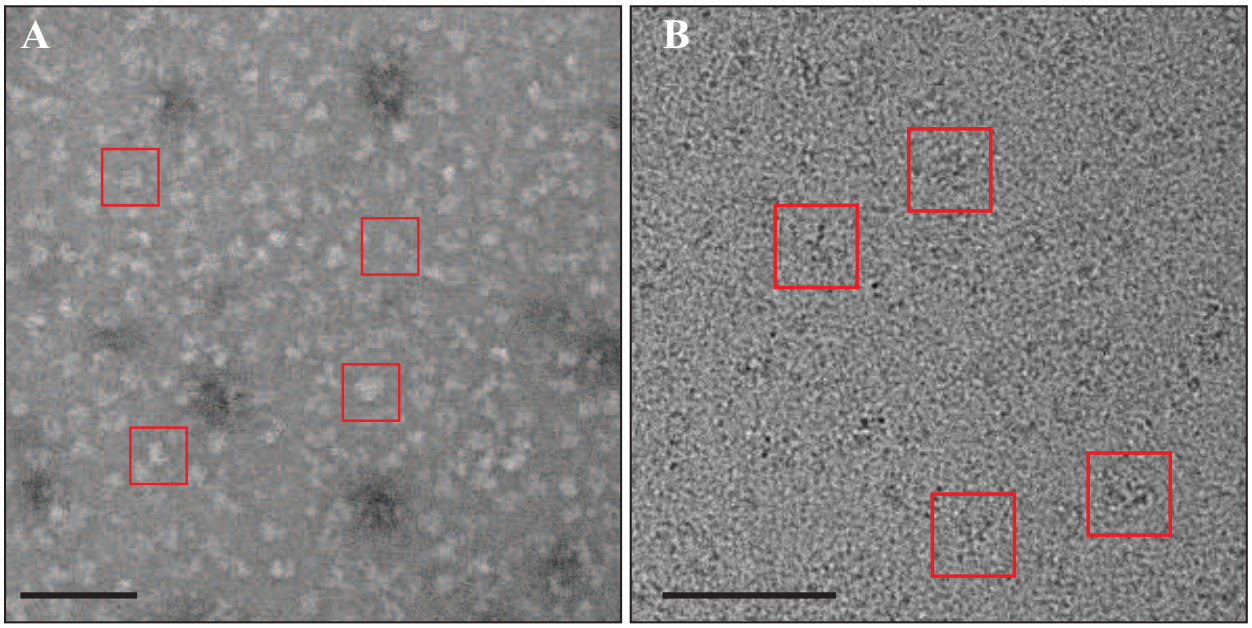


Figure 2.1: Micrographs of negatively stained and vitrified TFIID samples. (A) Negatively stained sample of TFIID in uranyl formate. (B) Cryo-EM micrograph of TFIID. Representative particles are shown in red boxes. Scale bars are 100 nm.

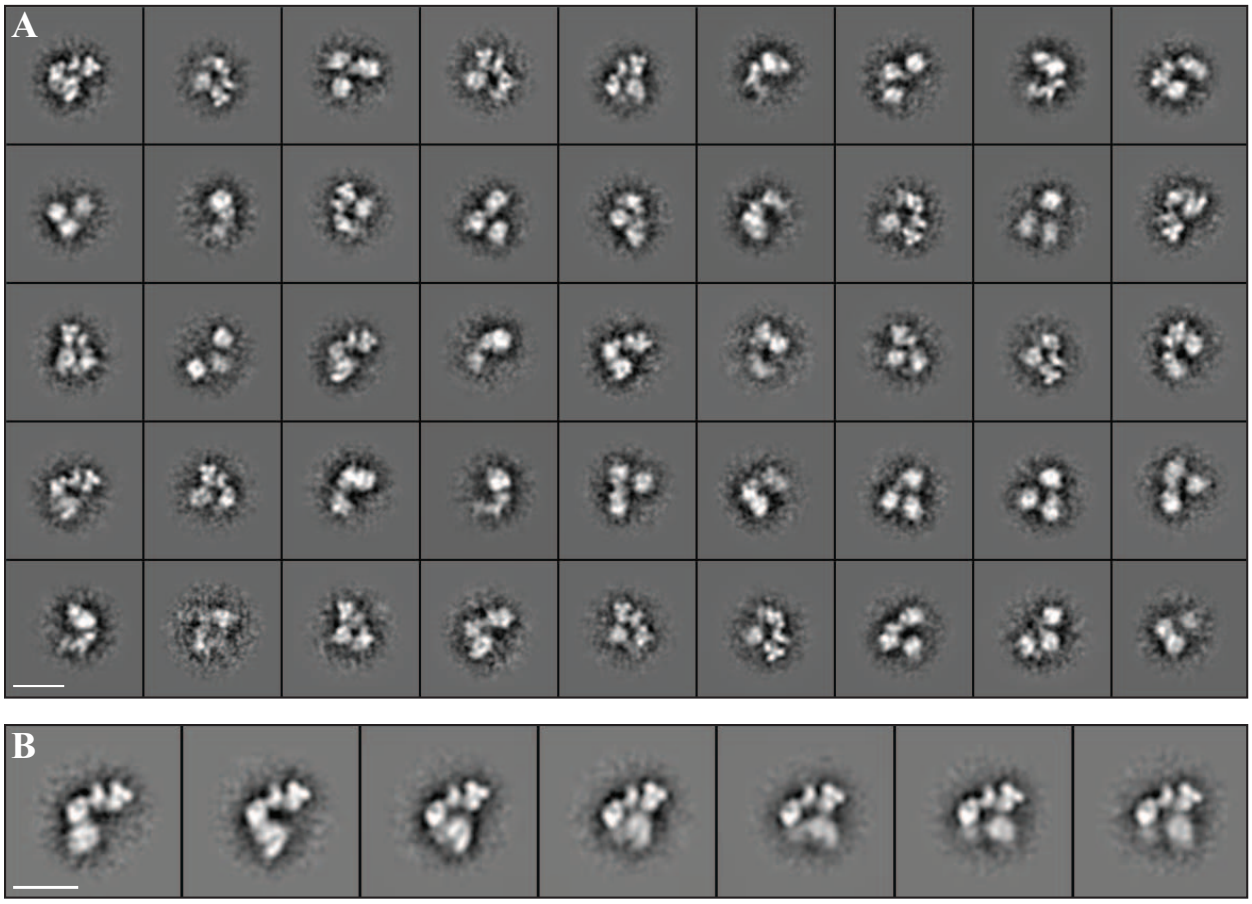


Figure 2.2: 2D reference-free class averages of negatively stained TFIID. (A) Representative class averages obtained during analysis of all particles within dataset. (B) Selected averages from (A) that correspond to a similar particle orientation to each other. Scale bars are 200Å.

negative staining procedures, cryo-EM data of TFIID were analyzed in an identical manner. Indeed, after performing 2D reference-free alignment on the particles, similar conformational states were observed within the cryo-EM class averages, where lobe A is flexibly attached to a stable BC core (Figure 2.4). Thus, lobe A undergoes a 100Å structural reorganization as it changes connectivity from lobe C to lobe B. Given that the structural state where lobe A contacts lobe C corresponds to the previously determined 3D structure of TFIID [1, 31], we will refer to this state as the 'canonical' state. When lobe A contacts lobe B, TFIID forms a rearranged horseshoe structure, hence we will refer to the newly discovered structure as the 'rearranged' state (Figure 2.4).

2.2 Focused classification & positional measurements of lobe A describes TFIID's conformational landscape

After initially characterizing the extent of lobe A's conformational variability, we wanted to quantify the relative occupancy of each structural state to arrive at a more complete description of TFIID's conformational landscape. To this end, we concentrated on a 'standard' view, where the three main lobes of TFIID, lobes A, B and C, were clearly separated. After performing 2D reference-free alignment, all particles corresponding to this standard view were extracted and re-aligned to a reference of the BC core only (Figure 2.5A). The average of these re-aligned particles showed that lobe A was 'blurred-out' due to the wide range of conformational states that it adopts (Figure 2.5A). Sub-classification within a 2D mask that excluded the stable BC core revealed that lobe A adopts a wide range of positions (Figures 2.5B). However, unlike the previously obtained averages (Figure 2.2B), this analysis provides a direct measure of the occupancy for each structural state.

To quantify the structural plasticity exhibited by TFIID, the position of lobe A along the BC core was measured within each sub-classified average. The measurements were normalized so that values greater than 0.70 correspond to particles resembling the canonical state, in which lobe A is at its closest position to lobe C (Figure 2.5D). Values less than 0.50, on the other hand, indicate that the particles are in a rearranged state, wherein lobe A is proximal to lobe B (Figure 2.5D). This analysis revealed that the position of lobe A can be described by a bi-modal distribution with peaks centered at 0.40 and 0.80 (Figure 2.5E and Figure 2.6A). Surprisingly, this showed that approximately 50% of the TFIID particles are found in the rearranged state. These distinct structural states of TFIID were observed in datasets from both cryo-EM and negative stain data, demonstrating that the sample preparation did not alter the results (Figure 2.5E). Hence, there are two predominant and structurally distinct states of TFIID that differ by the reorganization of lobe A within the complex.

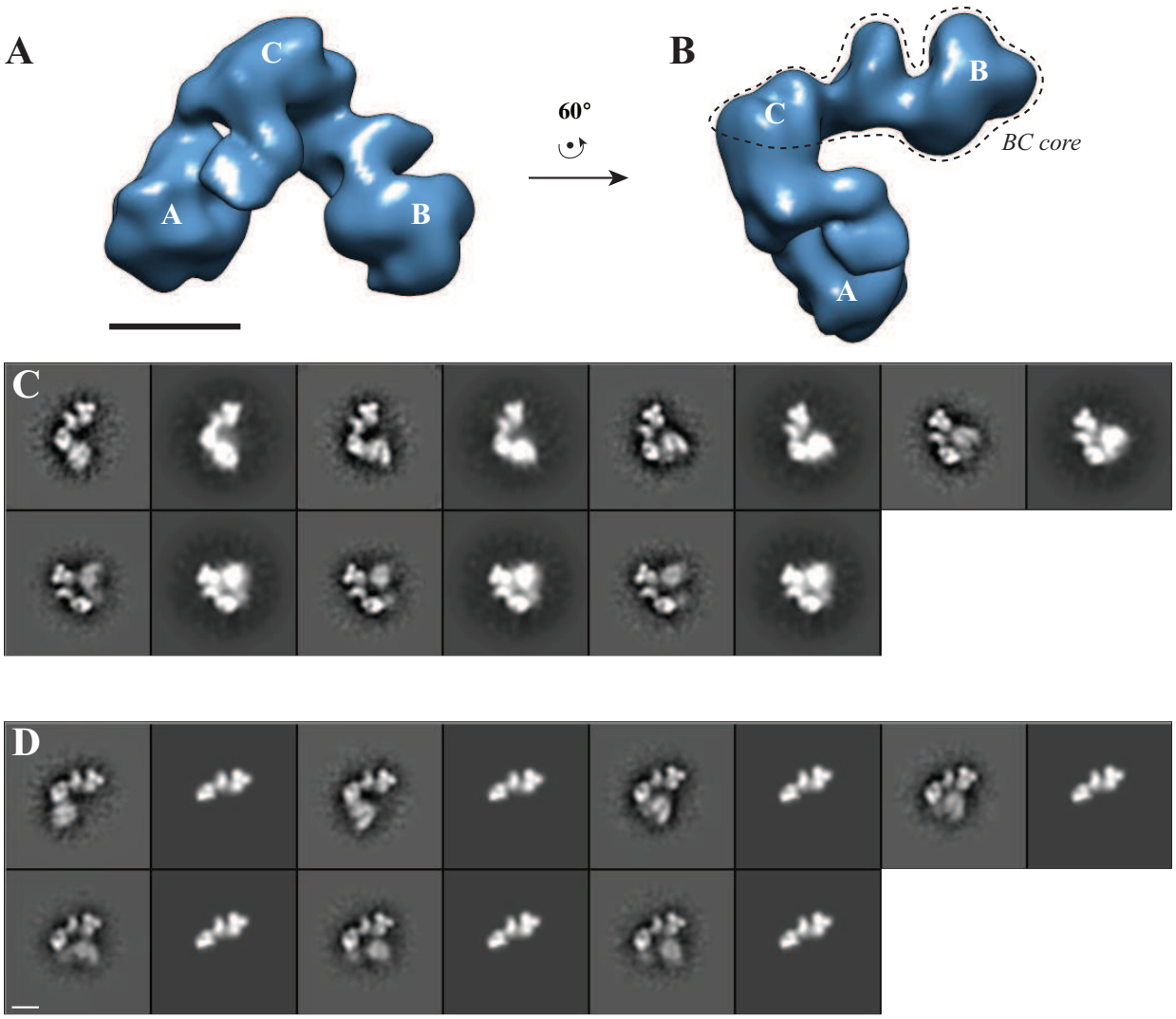


Figure 2.3: Lobe A exists in a range of positions relative to a stable BC core within TFIID. (A & B) Negative stain reconstruction of TFIID from previous work [31]. Lobe positions are indicated. (C) Class averages from Figure 2.2B aligned to re-projections of model in (A). (D) Class averages from Figure 2.2B aligned to re-projections of TFIID's BC core. The BC core is indicated in (A) by the dotted line. Scale bars are 100Å .

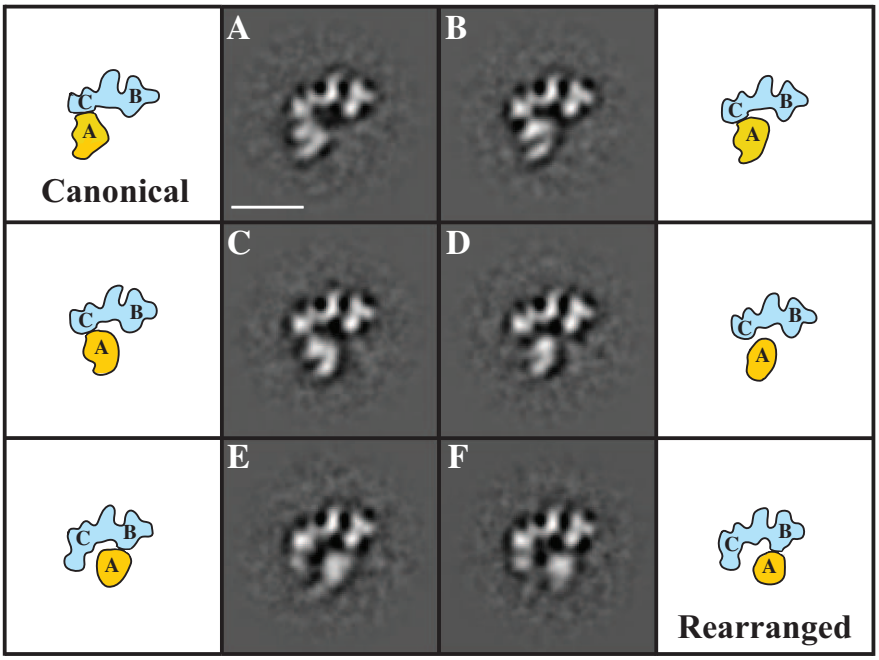


Figure 2.4: Cryo-EM class averages of TFIID confirm lobe A's flexibility. Selected 2D reference-free class averages of TFIID highlight large movement of A relative to a stable BC core. 'Canonical' and 'rearranged' conformations states are indicated. The BC core is colored in blue while lobe A is colored in yellow. Scale bar is 200Å .

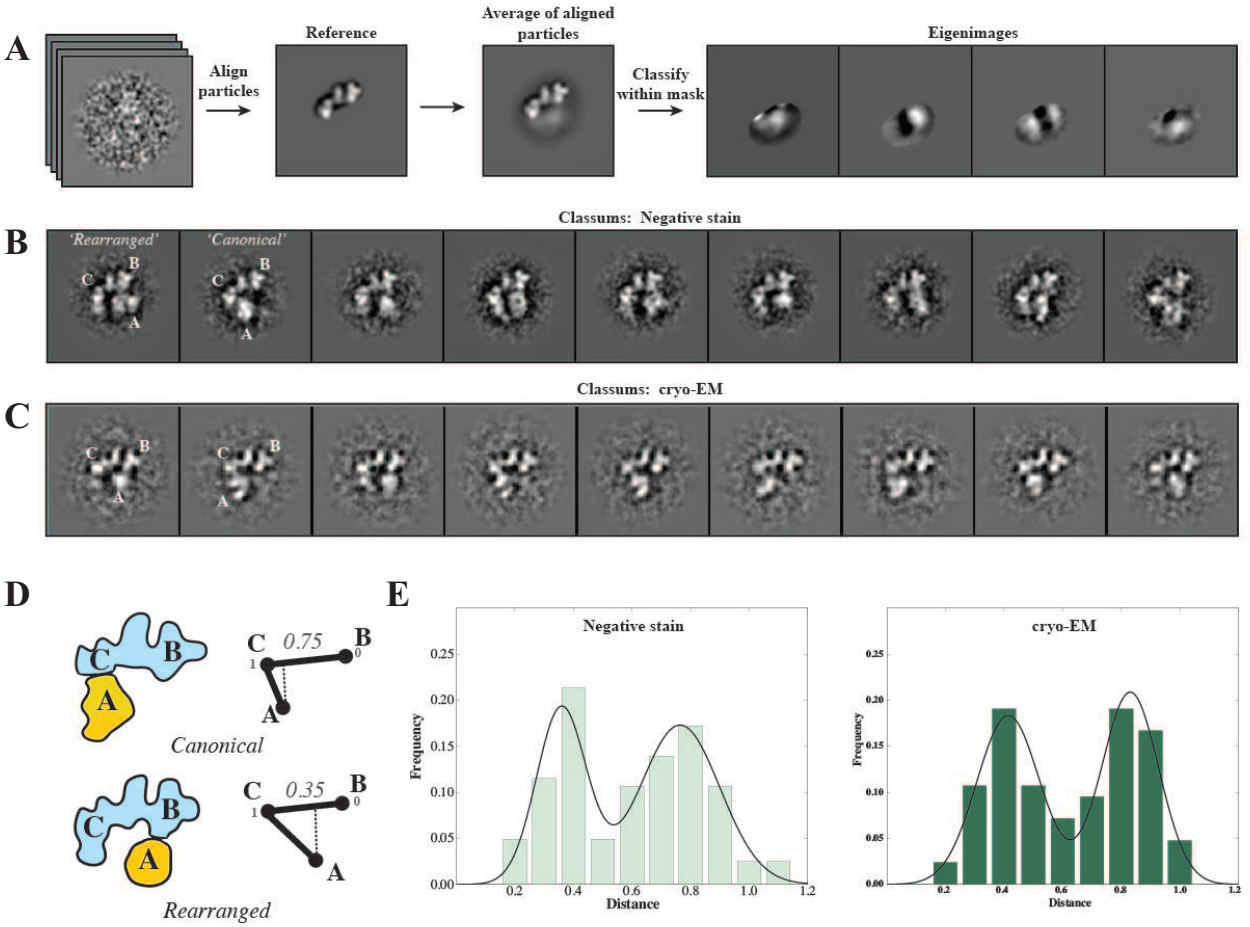


Figure 2.5: Focused classification and positional measurements of lobe A within TFIID. (A) Strategy for alignment & classification. Particles were selected based upon membership within the characteristic 'standard view' of TFIID. These selected particles were then aligned to the rigid BC core and classified within a mask centered on lobe A. (B) Resulting sub-classified classums from negatively stained TFIID particles after hierarchical ascendant classification. (C) Resulting sub-classified classums from cryo-EM particles of TFIID after hierarchical ascendant classification. (D) Schematic for measuring lobe A position within sub-classified classums. The position of lobe A was projected onto the B-C axis to provide a measurement for lobe A's localization within TFIID. (E) Histogram of lobe A positions for negatively stained (left) and vitrified TFIID (right).

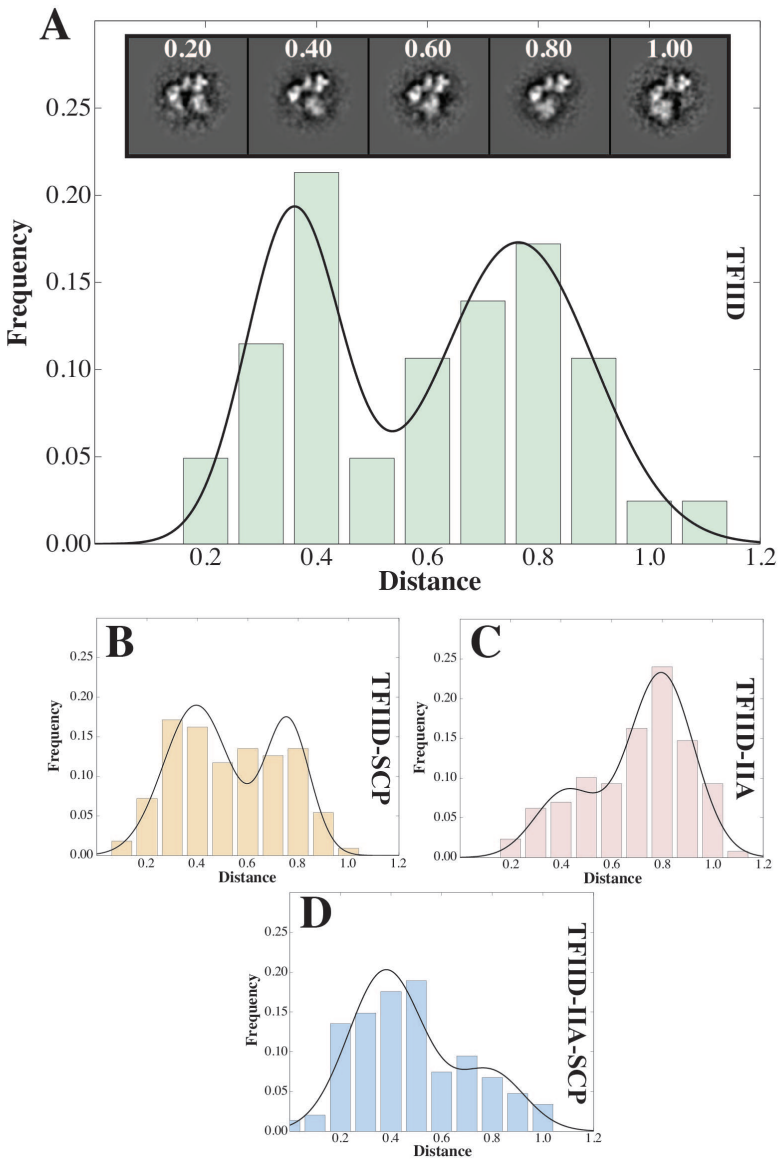


Figure 2.6: The conformational landscape of TFIID changes in response to TFIIA and SCP DNA. Distribution of lobe A positions relative to the stable BC core for TFIID (A), TFIID-SCP (B), TFIID-TFIIA (C), and TFIID-TFIIA-SCP samples (D). Inset within (A): Class averages corresponding to specific lobe A measurements from the TFIID sample.

2.3 TFIIA and SCP DNA modulate the position of lobe A within TFIID

The combined presence of TFIIA and SCP DNA stabilize the rearranged state

To examine whether the binding of TFIIA and promoter DNA is linked to the conformational states of TFIID, we analyzed TFIID in the presence of excess SCP DNA and/or the cofactor TFIIA. The SCP contains the TATA box, Inr, MTE, and DPE core promoter motifs (Figure 1.4). When TFIID was incubated with SCP DNA and negatively stained samples were visualized and analyzed as described above, lobe A showed a similar distribution of positions relative to TFIID (Figure 2.6B) ($p = 0.0353$, Wilcoxon signed-rank test). This result suggested either that TFIID does not interact with SCP under the conditions used, or that DNA binding does not alter the conformational partitioning of lobe A within the accuracy of our measurements.

We then tested whether the presence of TFIIA affects the properties of TFIID. When TFIID was incubated with TFIIA in the absence of DNA, the distribution of lobe A positions was shifted to the peak centered at 0.80. It thus appears that TFIIA stabilizes TFIID in the canonical state (Figure 2.6C). On the other hand, in the presence of SCP DNA, TFIIA promotes the formation of the rearranged state of TFIID, with a strong increase in the peak of lobe A positions centered at 0.40 (Figure 2.6D). The distribution of lobe A positions within the context of TFIID-TFIIA (Figure 2.6C) was significantly different from that seen with TFIID-TFIIA-SCP (Figure 2.6D) ($p = 1.5 \times 10^{-}13$, Wilcoxon signed-rank test). Since TFIID is purified using established protocols [67, 68], any underlying biochemical heterogeneity will be constant between these experiments, suggesting that these conformational changes are due to the presence of TFIIA and SCP DNA. These results indicate that TFIIA serves the dual function of maintaining TFIID in the canonical state in the absence of DNA and promoting the formation of the rearranged state in the presence of promoter DNA.

Cryo-EM analysis verified the presence of DNA within the rearranged state of TFIID

To characterize the structure of this novel rearranged state of TFIID and its binding to DNA, we carried out cryo-EM visualization of frozen-hydrated samples. Analysis of 2D reference-free averages from cryo-EM data of the TFIID-TFIIA-SCP ternary complex showed the presence of both the rearranged and canonical states (Figures 2.7A & B, leftmost panels). Importantly, the class averages corresponding to the rearranged state revealed extra density that appeared to be DNA extending across the central channel of TFIID (Figures 2.7A, left panel). This additional density was confirmed to be DNA by examining cryo-EM 2D reference-free class averages collected for TFIID, TFIID-SCP, and TFIID-TFIIA. 2D

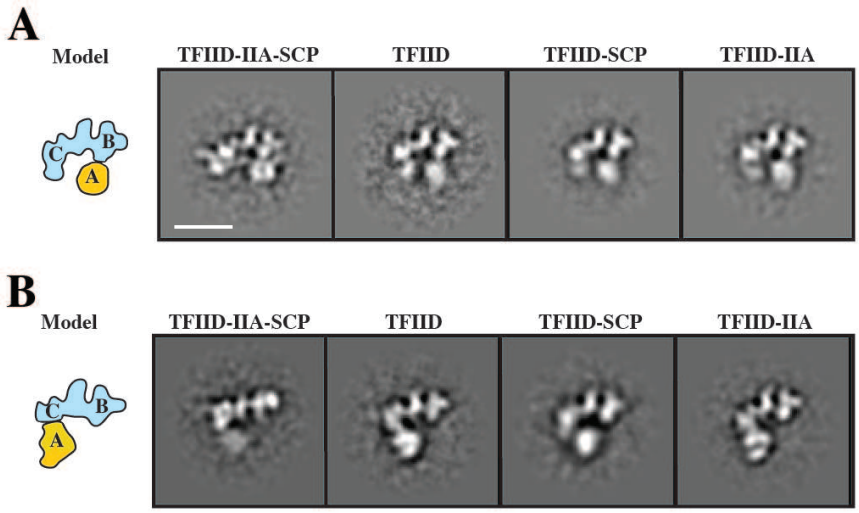


Figure 2.7: Cryo-EM analysis confirms that the rearranged conformation of TFIID interacts with SCP promoter DNA in the presence of TFIIA. 2D reference-free class averages obtained for the rearranged conformation (A) and the canonical conformation (B) in the presence of the indicated factors. Note the presence of linear density within the rearranged conformation for TFIID-TFIIA-SCP. Scale bar is 200Å .

reference-free class averages of TFIID without the combined presence of TFIIA and SCP showed no extra density within the rearranged state (Figure 2.7A). These observations suggest that TFIIA is required for efficient binding of TFIID to SCP and that the predominant form of TFIID that is bound to DNA is the novel rearranged state.

2.4 Discussion

Even though conformational flexibility was originally observed for lobe A within frozen-hydrated samples of TFIID [31], the large magnitude of this newly observed flexibility for lobe A within human TFIID went unappreciated for the past 10 years [1, 31, 67, 68]. Furthermore, re-analysis of all datasets of TFIID available reveal the presence of the rearranged conformation, in addition to the continuum of conformational states adopted by lobe A (data not shown). This indicates that the rearranged conformation has been consistently overlooked (e.g. [67]: Supplemental Figure S4). Given the subtle differences observed between projections of the canonical state when it was aligned to 2D class averages of the rearranged state (Figure 2.3C), it is not surprising that this slight difference was ignored.

The pendulum nature of the lobe A's movement relative to the stable BC core suggests that lobe A maintains a connection to the BC core within all conformational states. The nature of this linkage remains unknown and, given that there is not visible extra density connecting lobe A to the BC core, we postulate that the linking domain is an unstruc-

tured polypeptide chain. Furthermore, given the lack of information on subunit organization within TFIID, it is difficult to speculate on the composition of lobe A or the BC core at this time (see Chapters 3 & 4 for model of TFIID composition). Given the large distance (>100 Å)that lobe A moves relative to the BC core, understanding the molecular basis for modulating lobe A's position may have important implications for regulating TFIID activity.

Chapter 3

Cryo-EM analysis of TFIID-TFIIA-SCP defines the 3D organization of DNA and TFIIA bound to TFIID

3.1 Implementation of OTR to generate a model of the rearranged state

Principles of *ab initio* 3D reconstruction strategies

The refinement and 3D reconstruction of individual particles in single particle EM requires an initial model to be input by the user. Unfortunately, given that the single particles are inherently noisy due to low-dose image collection strategies, 3D reconstructions from single particle datasets can be sensitive to the initial model originally provided. Furthermore, given the 'model bias' problem observed for aligning individual noisy particles with a reference projection [102], generating an initial model for the refinement of protein structures represents a non-trivial step in single particle EM.

There are three commonly used approaches for calculating *ab initio* 3D models for use as initial models for 3D refinements of single particles: angular reconstitution [112], RCT [90], and OTR [59]. Each method relies on calculating 2D reference-free class averages through an iterative process of classification and alignment. From these averages, angular reconstitution relies upon the shared 'common line' in Fourier space between different 2D views of the same object. The 'common line' is the direct result from the central section theorem of image formation within transmission EM [22]. By calculating the 'common lines' between 2D reference-free class averages, an *ab initio* 3D reconstruction can be obtained for the particles analyzed. While this technique provides a powerful tool for calculating 3D models of both symmetric and asymmetric structures, it is limited by the assumption that the 2D

reference-free averages represent different views of the *same* object. Therefore, conformational or biochemical heterogeneity within the sample can result in 2D class averages that are *not* from the same object and, when these averages are combined during angular reconstitution, the resulting 3D model will not be an accurate representation of the underlying protein structure.

As an alternative methodology to angular reconstitution, RCT relies on collecting particle tilt pairs at a defined angle in order to calculate *ab initio* 3D reconstructions [90] (Figure 3.1). Since each particle collected for RCT has a corresponding tilt mate at known tilt angles (60 - 65°), generating 2D class averages for the untilted particles results in grouping tilt mates from different views together. And, since the angles between all of the particle tilt mates are known, 3D reconstructions can be calculated for each class average. This approach for calculating *ab initio* 3D reconstructions does not make any assumptions about the sample homogeneity (unlike angular reconstitution, see above), instead relying on the geometry of particle tilt pairs to provide structural information on 2D class averages. The only limitation with this technique is that the maximum tilt angle achievable in the EM is 60 - 65° due to the physical nature of the sample holder and support grid. This results in the famous 'missing-cone problem,' where there is a large cone-shaped area in reciprocal space that does not contain any structural information and can lead to artifacts within 3D reconstructions [25].

These limitations of RCT led to the recent development of OTR as an *ab initio* methodology that does not suffer from the 'missing-cone problem' (Figure 3.1) [58, 59]. The conceptual framework of OTR is similar to RCT: particle tilt pairs are collected in the EM through direct manipulation of the stage angle, but, instead of collecting images at 0°and 60°as in RCT, OTR images are collected at +/- 45°. This technique has the ability to fill in completely all of reciprocal space for a given 2D class average (Figure 3.1). The full description of 3D structural information for OTR class volumes abolishes the requirement for sub-volume averaging techniques, a method normally used for merging RCT class volumes to remove the 'missing-cone problem.' While there are robust techniques for accurate merging of RCT class volumes [97], combining class volumes of pseudo-symmetric volumes (like TFIID, see below) may result in loss of 3D structural information of distinct structural states.

3D reconstruction and refinement of rearranged conformation

In order to calculate an *ab initio* 3D model for the rearranged conformation, the above points were considered in deciding to implement OTR in determining the 3D structure of TFIID-TFIIA-SCP (Figures 3.2 & 3.3). Since OTR preserves three-dimensional features of individual class volumes without suffering from the 'missing-cone' problem of RCT [18], individual class volumes calculated from 2D reference-free class averages of TFIID-TFIIA-SCP (Figure 3.3A) were refined against untilted negative stain data for TFIID-TFIIA-SCP. These refined models were then quantitatively compared to the characteristic class averages of TFIID in six distinct conformations (Figure 3.3C & D) and models were excluded based

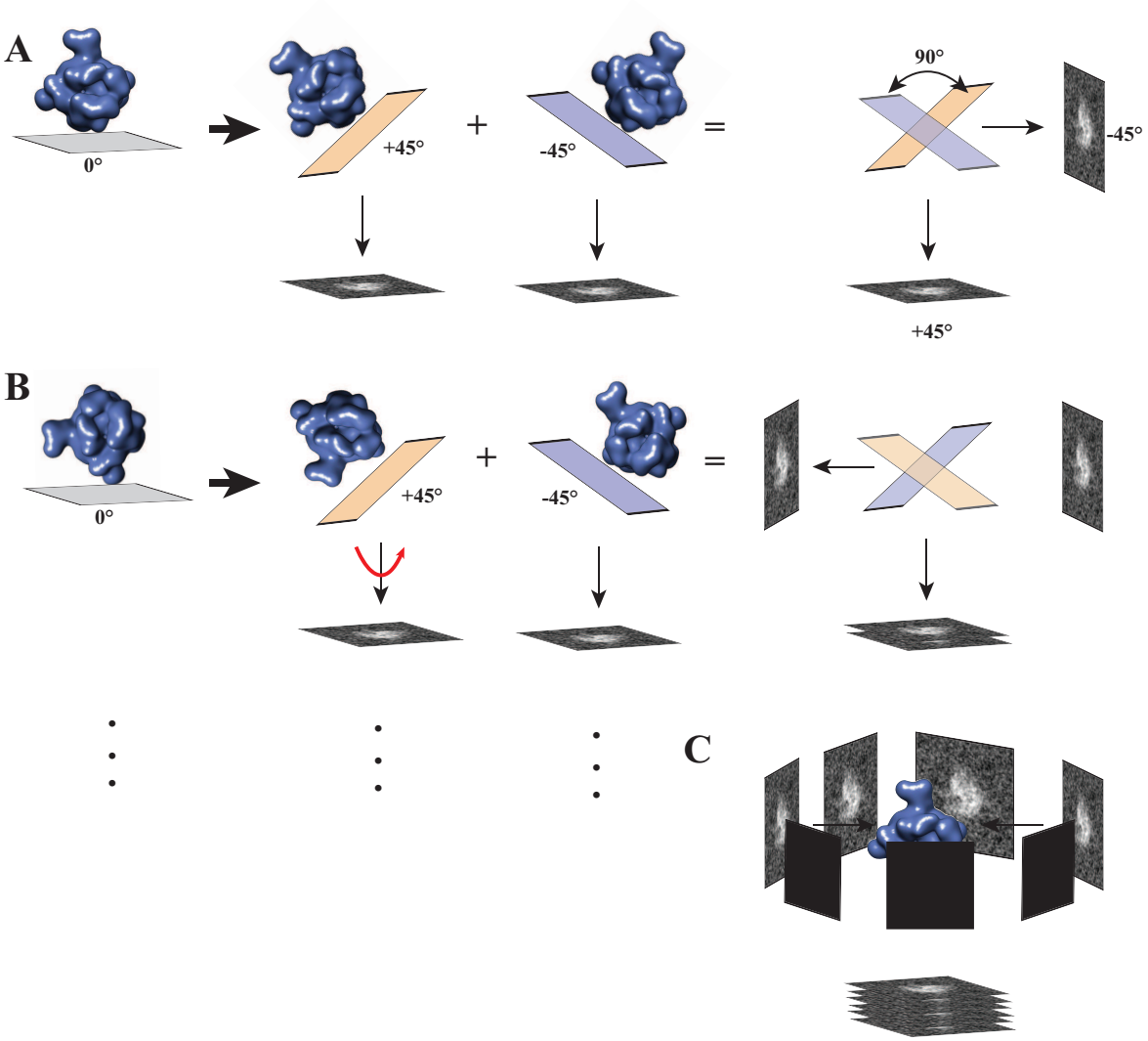


Figure 3.1: Principles of OTR and RCT. (A) First, the stage is tilted by +45° in the microscope before taking the first exposure. After recording the exposure, the stage is then tilted to -45° and a second exposure is taken. Superposition of a tilt pair reveals that the images are separated by 90° . (B) Another particle tilt pair is collected and, when averaged together with the first image through in-plane rotation of the +45° image, the particles further fill-in the 3D structure of the class average. (C) After averaging many particles that have been aligned through in-plane rotations, 3D information is built up for a given class average until a 3D reconstruction can be performed. Note that RCT is identical to the description of OTR here with the exception that images are collected at 0°and 60°.

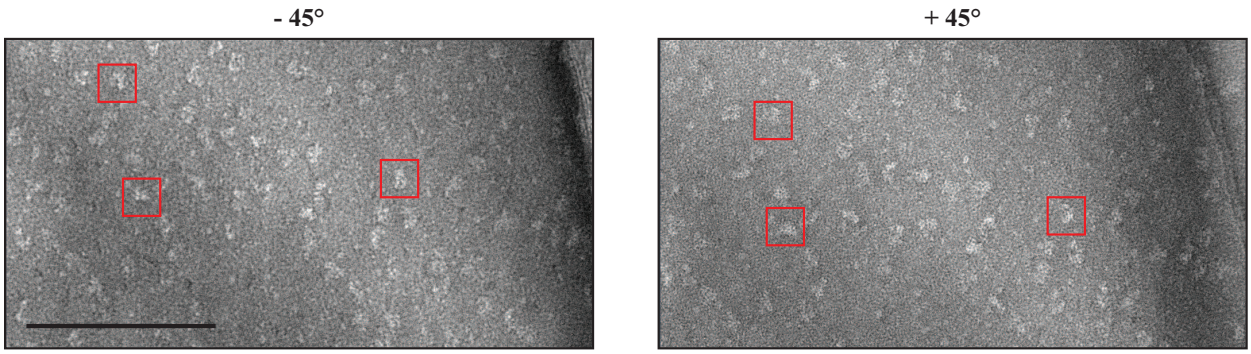


Figure 3.2: OTR tilt pair micrographs for TFIID-TFIIA-SCP. Scale bar is 200 nm.

upon Fourier Ring Correlation (FRC) [96] values below 60Å (Figure 3.3D). A majority (8/10) of the refined, validated models were found to be in the rearranged state (Figure 3.3E, c - g), whereas a minority (2/10) were in the canonical conformation (Figure 3.3E, a).

Comparison of the resulting refined 3D models corresponding to the canonical and rearranged conformations revealed the three-dimensional path of lobe A's rearrangement (Figure 3.4). As expected from analysis of 2D reference-free averages of TFIID (Figure 2.2B), lobe A's movement from the canonical conformation, where it contacts lobe C, to the rearranged conformation, contacting lobe B, mostly involves a translational component on the BC core (Figure 3.4A & B). Additionally, the BC core appears to be nearly identical between the two structures, indicating that a majority of the structural changes exhibited by TFIID occur through the repositioning of lobe A (Figure 3.4A & B, blue). While it is difficult to orient lobe A between the two conformations, lobe A likely rotates since the connecting density from A to C in the canonical state may rotate to contact lobe B in the rearranged conformation (Figure 3.4A & B, arrows). Unfortunately, despite the structural information gathered from negative stain on TFIID-TFIIA-SCP, nucleic acids are not preserved within the negative stain. This required cryo-EM sample preparation and analysis in order to visualize the path of DNA through TFIID-TFIIA-SCP.

3.2 Comparison of TFIID-TFIIA-SCP cryo-EM structures in the canonical and rearranged conformations

Previous 2D image analysis indicated that both TFIID and the ternary TFIID-TFIIA-SCP samples existed in a distribution of conformational states (Figure 2.6A & D), requiring the use of a multi-model 3D refinement strategy. A total of 35,000 cryo-EM particle images from the purified TFIID-TFIIA-SCP complex were sorted using multi-reference projection matching with two distinct reference structures obtained from negatively stained samples that represent the canonical and rearranged conformations. After implementing a cross-correlation cut-off that excluded 25% of the particles, the reconstruction of the rearranged state (comprising 60% of the remaining particles) was refined to a resolution of 32Å (Figure 3.5C). A prominent feature of the rearranged structure is the presence of density (Figure 3.5C, green), which we attribute to DNA, extending over lobe C and across the central channel of TFIID towards the region connecting lobes A and B.

A 3D model for the canonical state was refined simultaneously from the TFIID-TFIIA-SCP samples to a similar resolution and included the remaining 40% of the selected particles (Figure 3.5A). Importantly, the canonical state does not show the presence of any apparent DNA density, but is otherwise similar in overall features to the previously reported cryo-EM [31] and negative stain [68] structures of human TFIID. Comparison of the coexisting canonical and rearranged structures confirms the presence of a common BC core (blue, Figure 3.5A

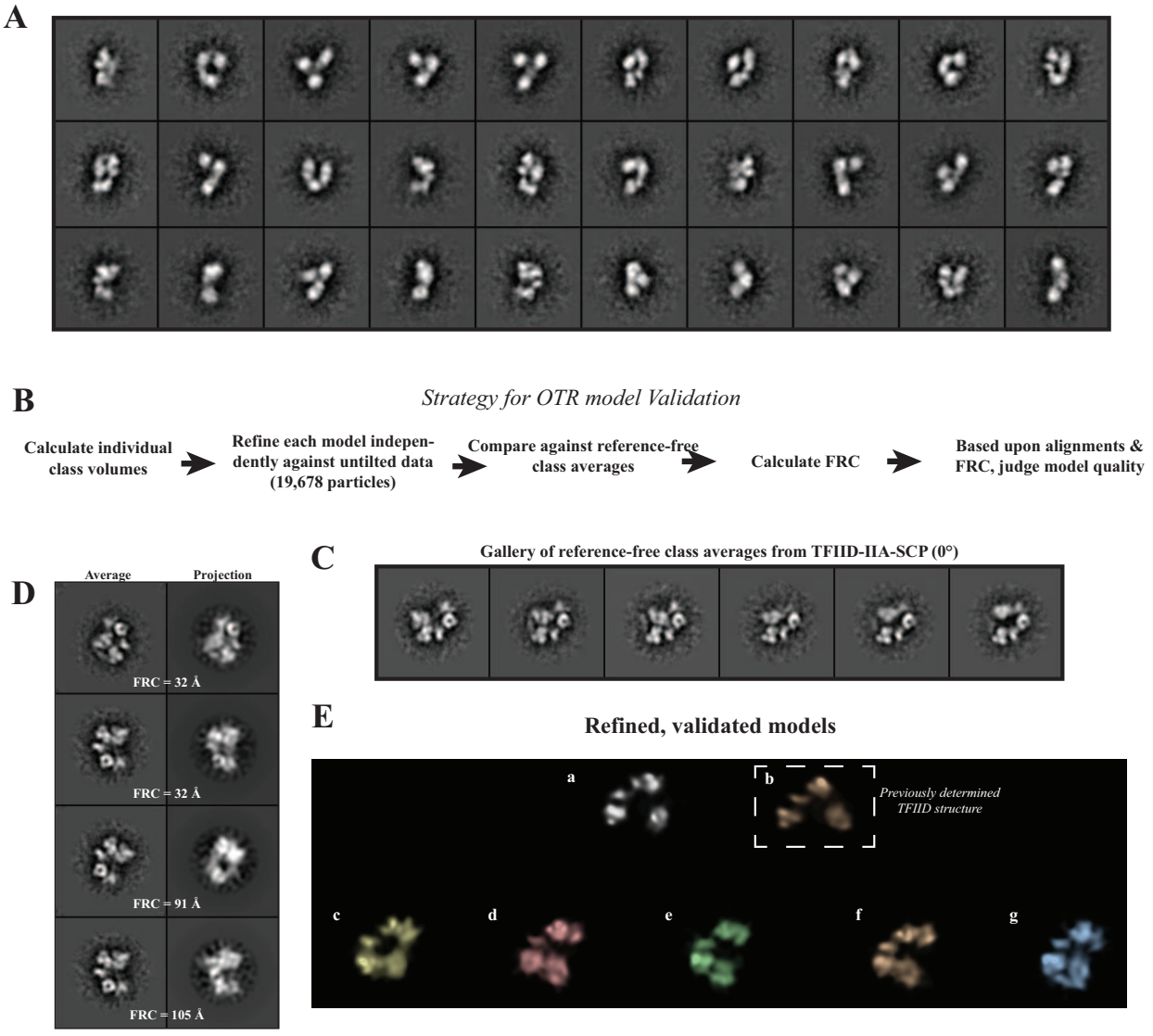


Figure 3.3: OTR implementation and model validation for both canonical and rearranged reconstructions of TFIID-TFIIA-SCP. (A) 2D reference-free class averages for -45° tilt data. (B) Strategy for validating OTR models. (C) Untilted TFIID-TFIIA-SCP 2D reference-free class averages that were used to assess model quality for refined OTR class volumes. (D) Highest & lowest resolution models shown as the top two and bottom two rows, respectively. (E) 3D model projections for refined & validated OTR models. Only 5 out of 8 rearranged models (c - g) and 1 out of 2 (a) canonical models are shown for space considerations. For comparison, the corresponding projection for a previously determined TFIID negative stain structure of the canonical state is shown (b) [31].

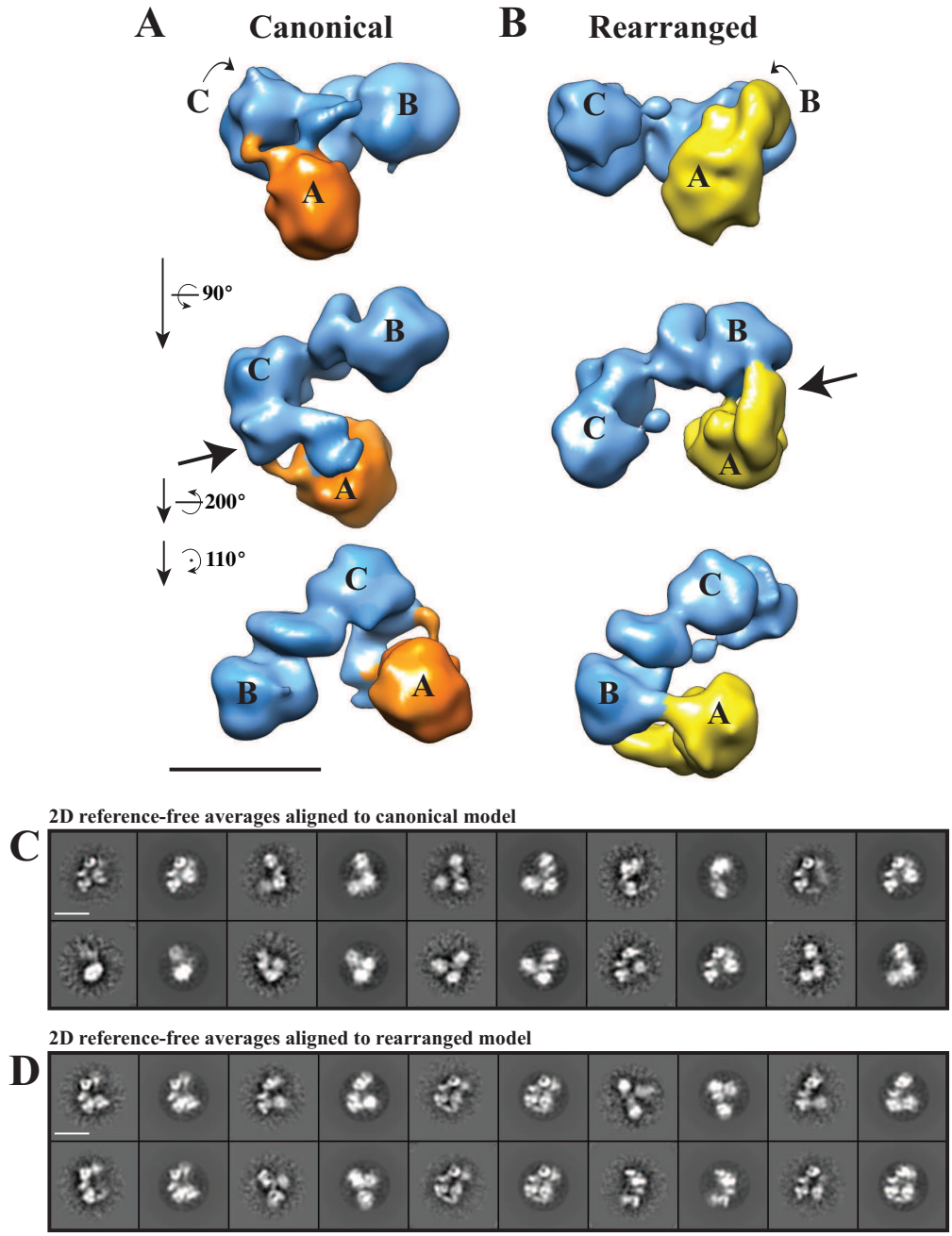


Figure 3.4: 3D reconstructions of TFIID-TFIIA-SCP in canonical and rearranged conformations from negatively stained particles. Canonical (A) and rearranged (B) reconstructions at 30Å and 32Å, respectively. The stable BC core is colored in blue in both structures while lobe A is colored orange within the canonical conformation and yellow in the rearranged conformation. 2D reference-free averages were aligned to the models in a multi-model manner, where the best matching averages for the canonical (C) and rearranged (D) conformation are shown. Scale bars are 200Å .

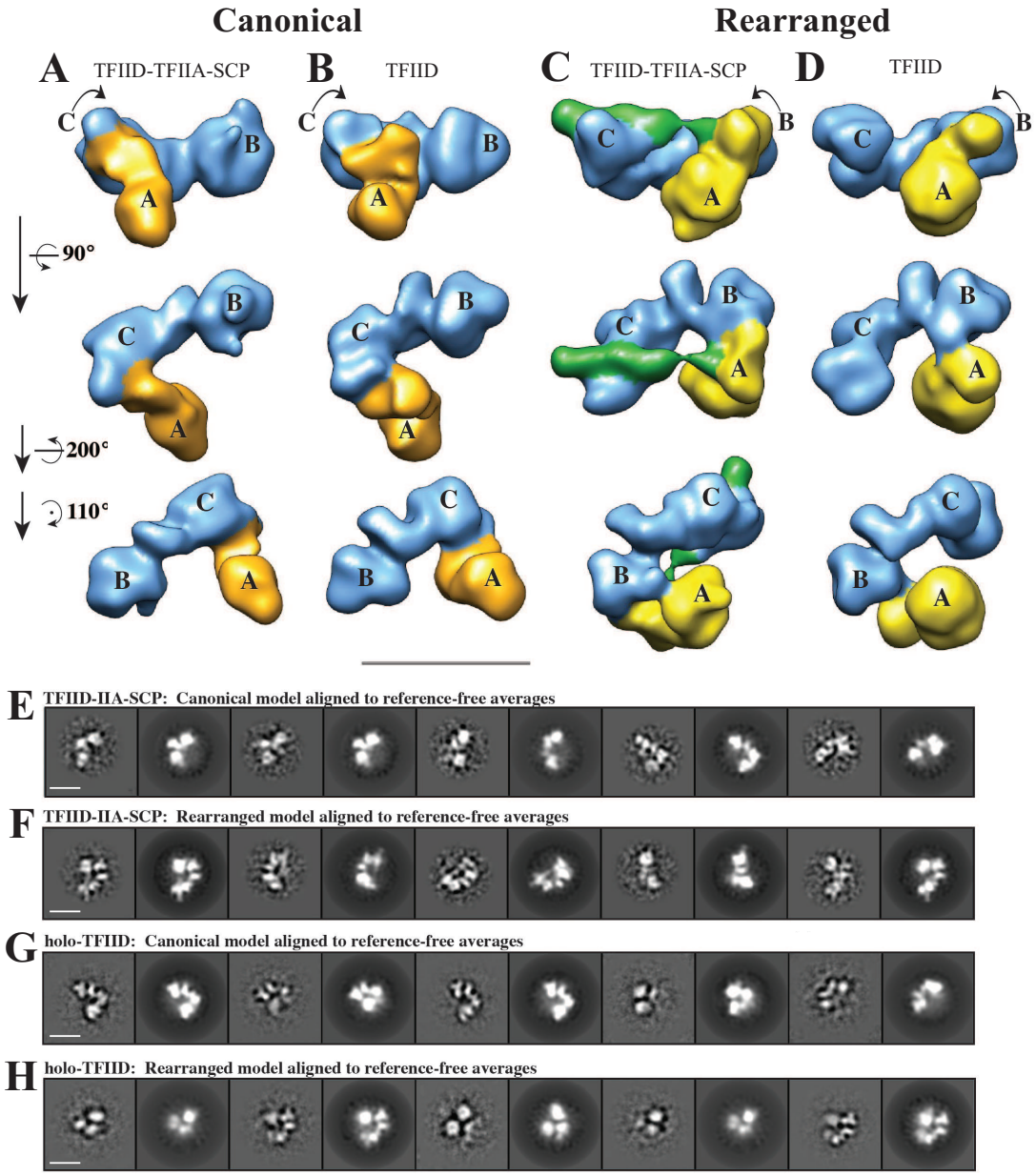


Figure 3.5: TFIIA-mediated binding of SCP DNA to the rearranged state of TFIID. 3D reconstructions of the canonical conformation for TFIID-TFIIA-SCP at 32Å (A) and TFIID at 35Å (B). 3D reconstructions of the rearranged conformation for TFIID-TFIIA-SCP at 32Å (C) and TFIID at 35Å (D), where density in (C) attributed to DNA is shown in green. The stable BC core is colored blue for each model. Lobe A is colored orange for (A) and (B) and yellow for (C) and (D). 2D reference-free averages for TFIID-TFIIA-SCP (E & F) and TFIID (G & H) are aligned to canonical and rearranged conformations, respectively. Scale bars are 200Å .

& C). However, the position of lobe A is dramatically shifted from one side of the core to the other, where lobe A within the canonical state is within close proximity of the DNA binding site of lobe C (orange, Figure 3.5A; yellow, Figure 3.5C). The movement of lobe A from the canonical state to the rearranged state involves a translational component along the BC core axis, as lobe A changes its apparent connectivity from lobe C (canonical) to lobe B (rearranged). These two reconstructions from the TFIID-TFIIA-SCP sample suggest that TFIID exhibits high affinity DNA interactions only when it adopts the rearranged conformation, considering the lack of any apparent DNA density for the canonical conformation.

The assignment of DNA as the narrow, linear density present only in the rearranged state of the TFIID-TFIIA-SCP sample was further supported by comparison of the rearranged cryo-EM structure from TFIID alone. Using the same multi-model approach described above, we obtained 3D cryo-EM reconstructions of the two alternative states of TFIID (Figure 3.5B & D). Comparison of the rearranged conformation (50% of selected particles) of TFIID with the rearranged structure of TFIID-TFIIA-SCP shows that there is strong density, which we ascribe to DNA, that is uniquely in the ternary complex and extends across the TFIID channel from lobe C to lobe A (Figure 3.5C & D). The structures of the canonical conformation of TFIID observed with TFIID alone and with the TFIID-TFIIA-SCP sample appear to be similar (Figure 3.5A & B), which suggests that the particles in the canonical conformation in the TFIID-TFIIA-SCP sample either do not bind DNA or lack stably positioned SCP DNA.

Orientation of DNA within TFIID-TFIIA-SCP

To examine the orientation of the DNA bound to the rearranged TFIID-TFIIA-SCP complex, we performed a DNA extension experiment. Cryo-EM data were collected from a sample of TFIID, TFIIA, and SCP DNA with a 30 bp 5' extension to position -66 (termed '-66') upstream of the TATA box (Figure 3.6). Comparison of cryo-EM 3D reconstructions for the rearranged state of TFIID-TFIIA-SCP(-66) and TFIID-TFIIA-SCP revealed significant extra density extending out of lobe A, which was the strongest difference between the two cryo-EM reconstructions at $\sigma = 4$ (Figure 3.6, arrowheads). The dimensions of this additional density are consistent with the length of the DNA extension. Beyond identifying the upstream DNA sequence at position -66, this additional DNA density extending from lobe A also provided a marker for the position of the TATA box, which is 36 bp from the end of this extended DNA, and thus within lobe A of the rearranged state.

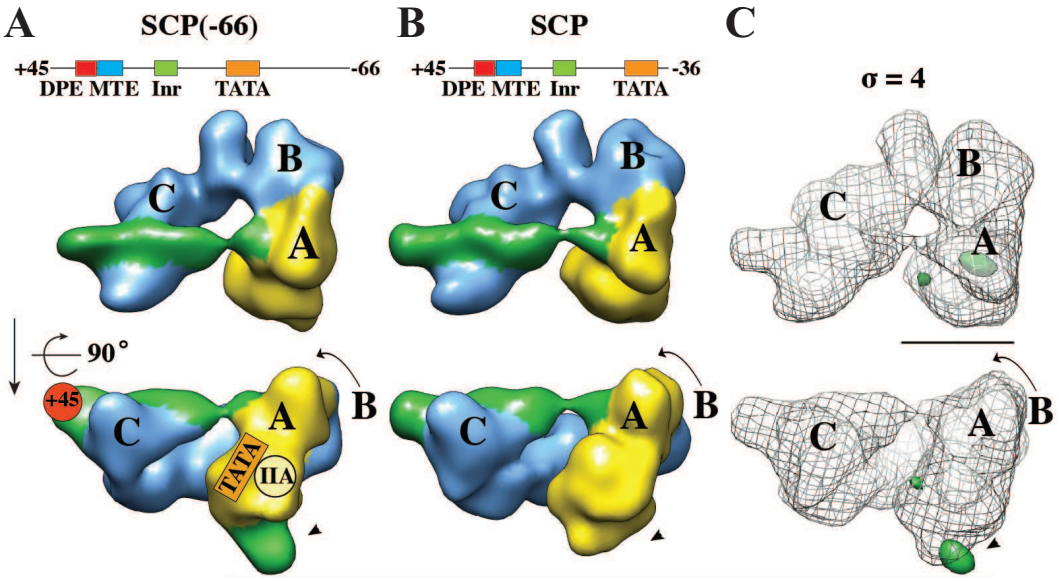


Figure 3.6: Extending SCP DNA upstream by 30 bps reveals upstream DNA path exiting lobe A. Cryo-EM reconstruction of TFIID-TFIIA-SCP(-66) at 35 Å (A) aligned with the cryo-EM reconstruction of TFIID-TFIIA-SCP (B) and their difference map (C). DNA density is colored in green, whereas lobe A and the BC core are colored in yellow and blue, respectively. Summary of Nanogold labeling from Figure 3.16 is shown in (A). Scale bar is 100Å .

3.3 Validation of the rearranged conformation using the free-hand test

As discussed in Chapter 1, the published literature of TFIID structures has provided a number of different 3D models for TFIID alone [24, 61, 62, 83] and the structure that TFIID adopts on promoter DNA [24, 84] (Figure 1.6). Therefore, in order to ensure the validity of the 3D models presented here, the free-hand test was performed on cryo-EM data of TFIID-TFIIA-SCP. Originally proposed by Rosenthal & Henderson [92], the free-hand test simultaneously provides an un-biased assessment of 3D map handedness and quality of particle alignments (Figure 3.7) [3, 35, 57, 92], and thus constitutes an external validation of 3D maps not unlike the R_{free} value in x-ray crystallography [9].

Test study on 26S proteasome

Using the strategy outlined (Figure 3.7) and computer software kindly provided by John Rubinstein (University of Toronto), I implemented the free-hand test on a well-behaved sample, the 26S proteasome from *S. cerevisiae*. Previous research in the lab found that the 26S proteasome achieved the highest resolution to date for a non-helically symmetric molecule within the Nogales lab (6 - 10Å) [56]. After Gabriel Lander kindly collected cryo-EM tilt pairs, I manually extracted the particles (Figure 3.8A) and aligned them using established laboratory 3D refinement parameters (see Materials & Methods). The free-hand test results showed a narrow clustering of particle alignments at 20° (Figure 3.8C). Satisfyingly, this is the same tilt angle applied to the sample holder in the EM, indicating that the model of the 26S proteasome has the correct handedness. This is a trivial result considering that the high resolution of the 26S proteasome model allowed unambiguous docking of crystal structure components, confirming the handedness [56]. However, this sample still served as an important positive control for this new test, where the 26S proteasome data provided insight into the accuracy of particle alignments needed to achieve sub-nanometer resolution of single particles in cryo-EM.

Free-hand test of the rearranged & canonical conformations

After validating the implementation of the free-hand test on the 26S proteasome, the free-hand test was performed on the rearranged and canonical cryo-EM models of TFIID-TFIIA-SCP. Particle tilt pairs (Figure 3.9A) were manually selected, extracted, and subjected to identical alignment parameters that were originally used for the 3D refinement and reconstruction of the TFIID-TFIIA-SCP data (see Materials & Methods). The free-hand test results show that the 3D model for the rearranged conformation has the corrected handedness (Figure 3.9B). Furthermore, these data show that, overall, the particle alignment

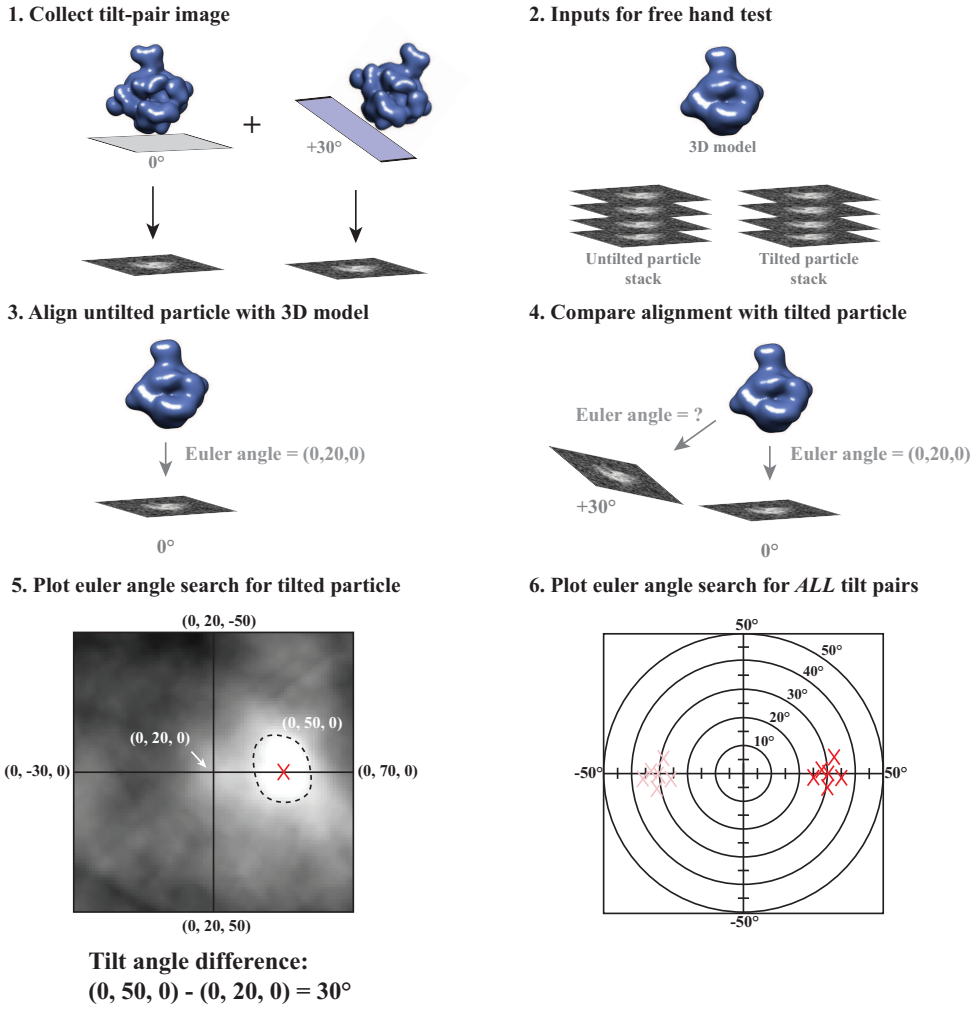


Figure 3.7: Unbiased assessment of model handedness and particle alignments using the free-hand test. (1) Tilt image pairs are collected in the microscope at a small tilt angle, typically between 10 - 30°. (2) Necessary input files for the free hand test: 3D model (handedness to be tested), untilted stack, and tilted particle stack. (3) Align model to an untilted particle. In the example, the Euler angle assigned to this particle is (0, 20, 0). (4) According to this starting angle (e.g. (0, 20, 0)), Euler angles are searched systematically within a given angular range. (5) Graphical depiction of resulting cross-correlation plot from systematic Euler angle search. The middle of the plot is the Euler angle for the untilted particle, (0, 20, 0). Each Euler angle combination is tested and the cross-correlation is scored and recorded. Note the intense peak centered at (0, 50, 0). Accurate alignment of the untilted particle results in a angular difference between the untilted and tilted particles to be the same as that applied within the microscope (e.g. +30°). (6) This analysis is then extended for all particle tilt pairs and the peak value is recorded. Accurate alignment of the particles results in a cluster of peaks around +30, the angle applied in the microscope for this example. Note that a model with opposite handedness would have a peak at -30°.

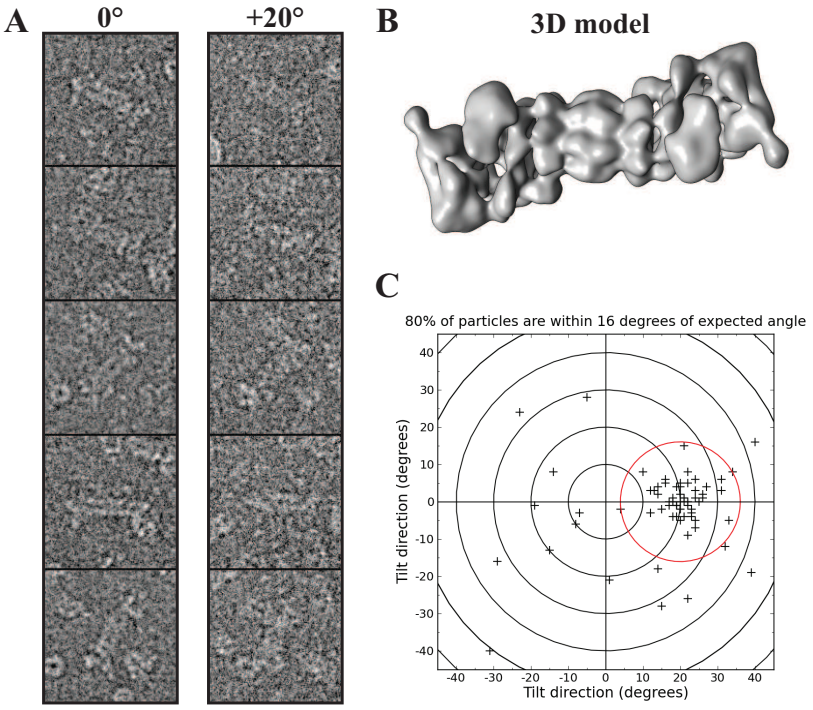


Figure 3.8: Free-hand test on 26S yeast proteasome. (A) Selected particle tilt-pairs. (B) Refined 3D cryo-EM model for 26S proteasome used for free-hand test. (C) Free-hand plot for 86 particle tilt pairs.

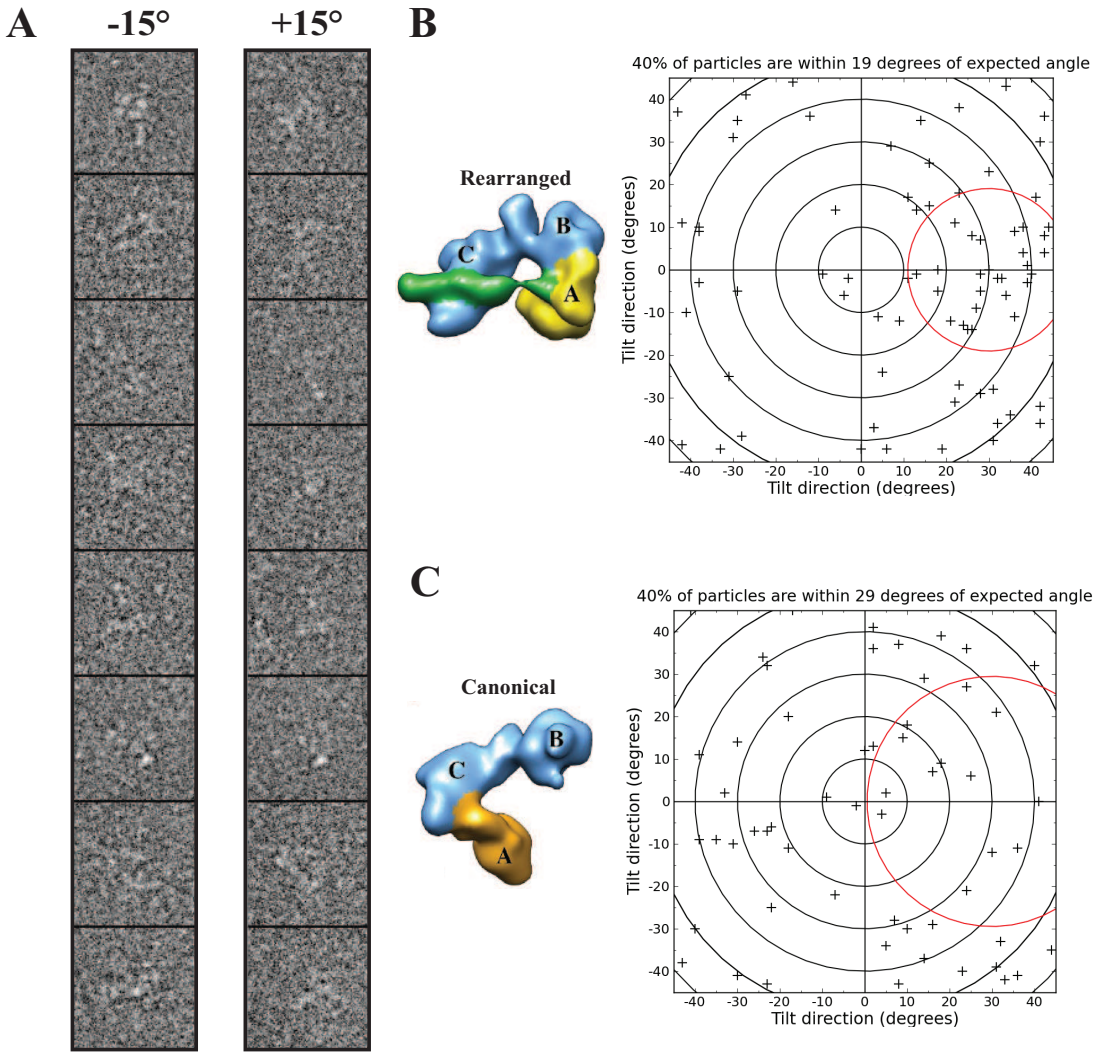


Figure 3.9: Free-hand test on TFIID-TFIIA-TFIIB-SCP. (A) Particle tilt pairs were collected at -15˚and +15˚on a sample of TFIID-TFIIA-TFIIB-SCP (see Chapter 4 for more information). Tilt pair plot for particles belonging to the rearranged conformation (B) and canonical conformation (C).

accuracy is correct, but, when compared to the free-hand test for the 26S proteasome, the more diffuse clustering of the rearranged particles provide insight into the limited resolution of the rearranged 3D model (32Å) when compared to the 26S proteasome (6 - 10Å) (Figure 3.9B).

The free-hand test for the corresponding canonical conformation revealed that the particle alignments were almost 'random' when compared with the rearranged model particle alignments (Figure 3.9C). This result is in contrast to the previously observed agreement between projections of the canonical conformation and 2D reference-free class averages (Figure 3.5E). Considering that there is an interdependence of model quality, particle quality, and 3D alignment routine that is measured by the free-hand test [3, 35, 57], it is difficult to understand the poor particle alignments for the canonical conformation. However, considering that the predominant conformation within the TFIID-TFIIA-SCP sample is the newly identified rearranged conformation (Figure 2.6D), the free-hand test results for the rearranged conformation provide an objective validation of the model handedness and particle alignment accuracy. This approach for validating 3D models can be extended to the other TFIID structures [24, 61, 62, 83, 84] so that the field can arrive at a consensus on TFIID's structure.

3.4 Nanogold as a tool to localize ligands bound to TFIID

The free-hand test verified the accuracy of the 3D rearranged model from TFIID-TFIIA-SCP, allowing further experiments to be performed in order to define the stereochemistry of this ternary complex. In order to visualize the DNA within the rearranged state for labeling, all structural experiments were performed on plunge frozen samples of TFIID-TFIIA-SCP. Due to the low SNR of cryo-EM images, traditional protein-based tags (e.g. antibody, streptavidin, or MBP) may be difficult to localize accurately. Therefore, in order to overcome this problem, a Nanogold-based tagging strategy was developed to label the SCP DNA and TFIIA within TFIID-TFIIA-SCP.

Theoretical considerations for imaging Nanogold within vitreous ice

Gold nanoclusters have been an important tool for EM image analysis for the past 20 years [32]. From immunolocalization in cell sectioning experiments to labeled-ligands within single particle cryo-EM, the high contrast of gold relative to protein and nucleic acids makes it an ideal labeling reagent. Previous work that labeled specific sites within single particles using Nanogold showed that in order to take advantage of the high contrast of gold low defocus images needed to be collected [6, 10, 74]. While these low defocus images were appropriate to localize the gold nanoclusters, the contrast of protein decreases dramatically at these defocal

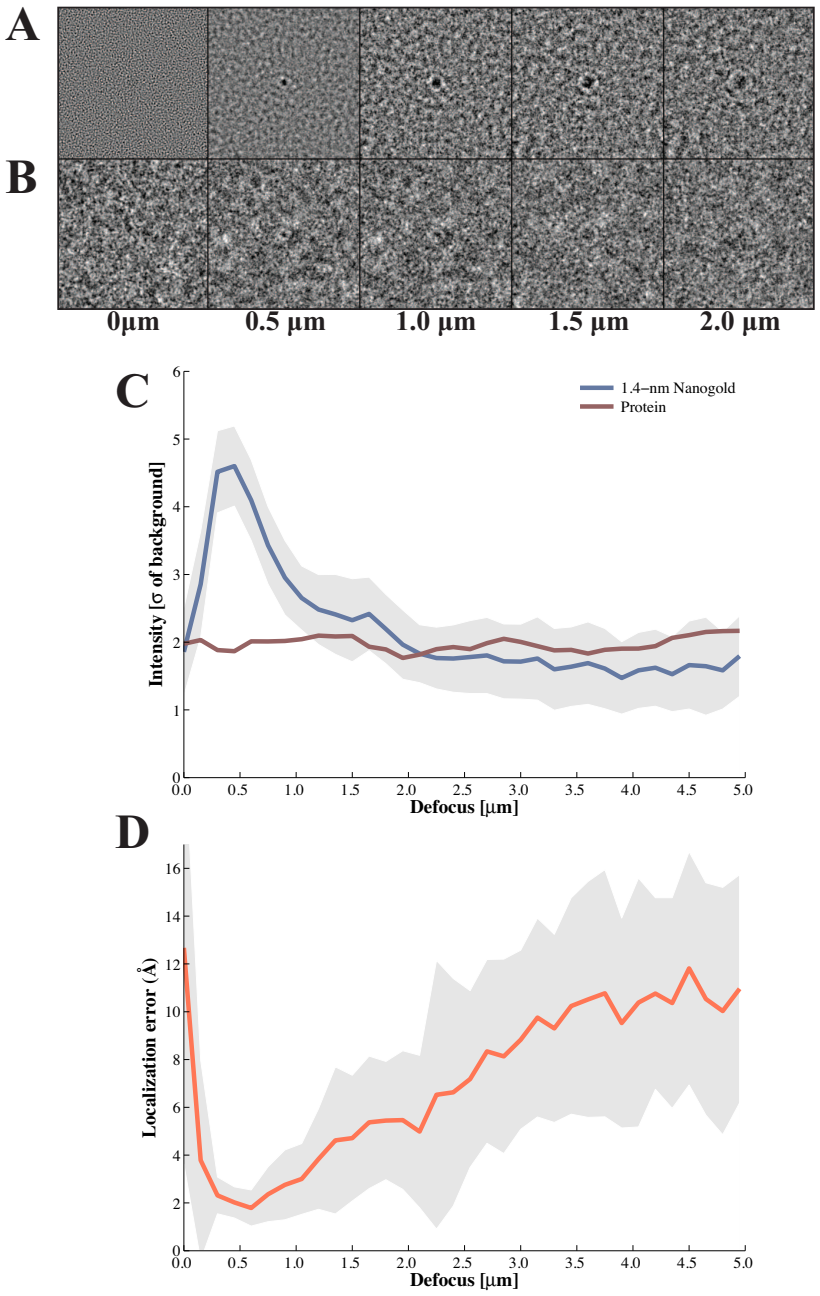


Figure 3.10: Determination of optimal defocus for 1.4-nm Nanogold. 2D Gaussians were fit to the simulated images of a 70 atom gold nanocluster embedded in water (A). For the peak finding shown here, noise was added to the simulated images to model accurately cryo-EM image characteristics (B). (C) Intensity of a 70-atom gold nanocluster (blue) is plotted against a range of defocal values. The standard deviation of the gold nanocluster intensity is shown in grey. For comparison, the intensity of a protein similar in size (amylase, PDB 1PPI) is shown (red). (D) Distance between peaks found in simulated gold images and the actual position of the gold nanocluster. Standard deviation of localization is shown in grey.

values. Therefore, analysis of gold labeled single particles requires the use of 'focal pairs' where a high defocus image is collected to visualize the protein density while the low defocus image is used to localized the gold nanocluster.

While there appears to be a wide-spread acknowledgement that gold nanoclusters display increased contrast at low defocus values compared to the high defocal values needed to image proteins in vitreous ice, there has not been a rigorous theoretical and experimental validation of 'best use' practices with single particles labeled with Nanogold. Therefore, before utilizing Nanogold as a labeling tool, we wanted to perform theoretical image simulation analysis of gold contrast across a range of defocuses.

To provide a theoretical assessment of the intensity of Nanogold in the EM, we performed multislice image simulations of gold nanoclusters embedded in water at 120 keV [28, 33, 51]. We used the multislice approximation of image formation in the EM [28], where the transmission of the incident wave is modeled as Fresnel propagation through a series of thin specimen 'slices.' Multislice imaging of biological macromolecules has been used successfully on previous projects within the lab [33], allowing existing computer programs to be manipulated in order to study gold nanoclusters. A 70 atom gold nanocrystal, comparable in size to 1.4-nm Nanogold, was constructed and embedded into a 373Å x 373Å box of water molecules. Image simulations of this model at varying defocuses showed changes in the apparent size and intensity of the gold nanocluster (Figure 3.10A). In order to simulate cryo-EM images accurately, Poisson noise was added to the images to simulate a does of 25 $e^-/Å^2$ (Figure 3.10A) and peak intensities were recorded over defocuses ranging from 0.0 to 5.0 μm defocus (Figure 3.10B). Relative to the simulated intensity for amylase, a 55 kDa particle, the gold nanocluster showed an intense peak at 0.45 μm defocus, reaching values almost 5σ over the average pixel intensity (Figure 3.10C). Considering the signal of this peak, the simulation clearly demonstrates that the optimal defocus for a 70 atom gold nanocluster is 0.45 μm.

The significant changes in gold nanocluster intensity suggested that the accuracy of gold nanocluster localization would be affected by the defocus. To test this, the distance between the actual location of the gold nanocluster within the model and the measured peak position in the image was recorded and the results plotted against defocus (Figure 3.10D). The results show a minimum at 0.45 μm defocus, demonstrating that the intense gold nanocluster peak at this defocus provides near-perfect localization, within 1 - 2Å of the actual location. Furthermore, at very low and high defocus values, there is a substantial increase in the mis-localization of the measured gold location. This indicates that images collected at 0.45 μm defocus of a 70 atom gold nanocluster (1.4-nm Nanogold) will provide the strongest signal and highest accuracy in gold localization ability.

The image simulations were an important tool because they reproduce accurately the diffracted beam of electrons through the simulated specimen. This approach is distinct from the 'typical' approach used by many EM users, where the optimal defocus is estimated by maximizing the intensity of the first CTF-peak in the power spectrum. For instance, using the latter approach, the predicted optimal defocus for a 70-atom gold cluster (1.4 nm) 0.15 μm, a defocus value that does not exhibit high gold signal signal in the simulations (Fig-

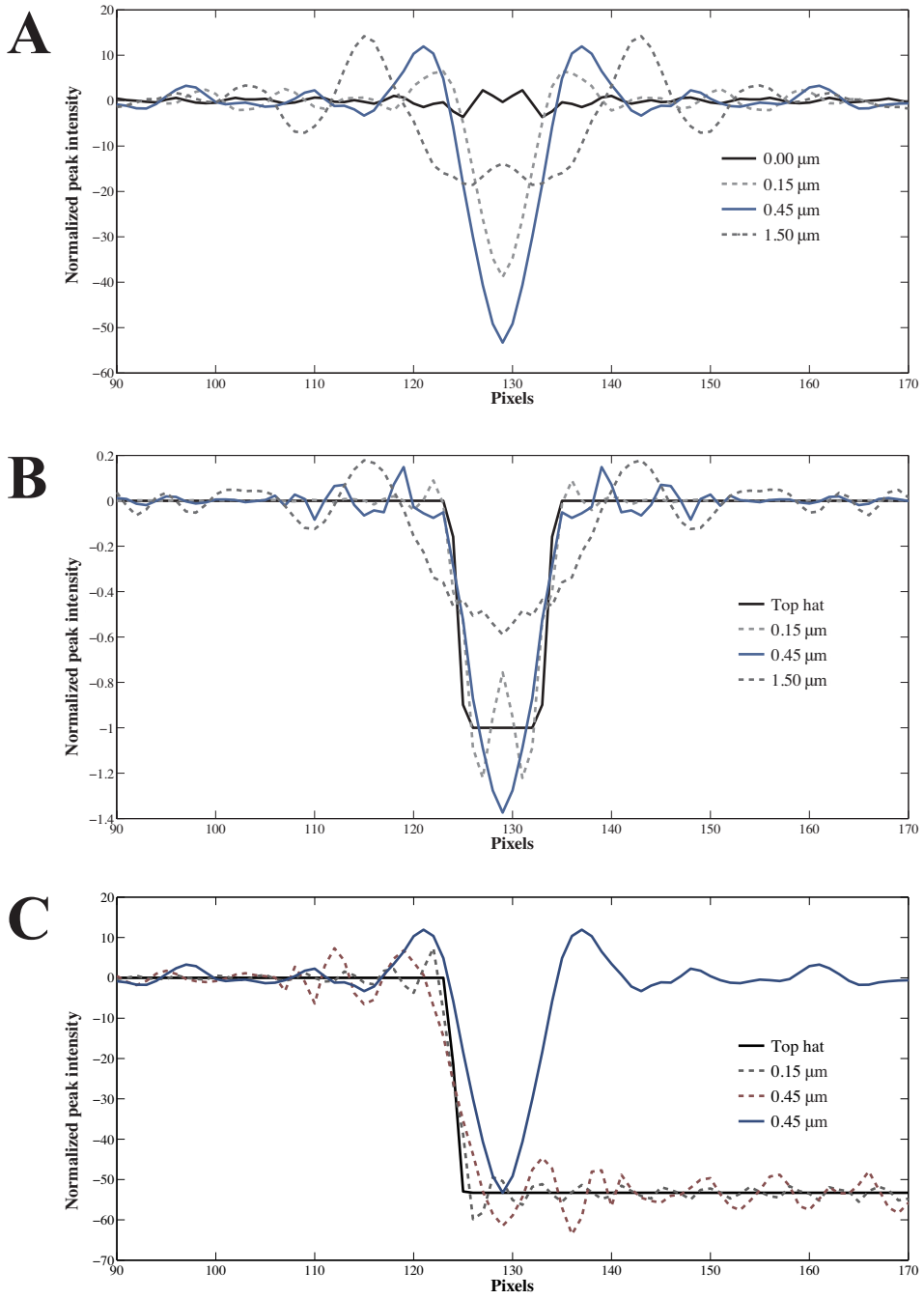


Figure 3.11: Fresnel fringe-induced amplification of gold nanocluster intensity. (A) Cross-sections through simulated gold images from Figure 3.10A at 0, 0.15, 0.45 and 1.5 μm defocus. (B) Line traces from CTF-applied to a top hat function with a 9 pixel diameter. (C) Line traces from CTF-applied to a step function ('edge') centered on the edge of the gold nanocluster. Plots are shown alongside 0.45 μm gold line trace from (A).

ure 3.10C). This observed difference betewen the image simulations and *ab initio* predication suggest that the observed gold contrast is likely modulated by fringe effects of the diffracted wave traveling through the spherical gold nanocluster. Fresnel fringes are obvious features introduced when spherical objects are visualized by defocus images from a transmission EM, where the fringes modulate the image intensity at the object's edge. To visualize the Fringe effects introduced by the CTF, line traces were calculated through the noise-free image simulations of gold nanoclusters and plotted (Figure 3.11A). The traces show the characteristic effect of increased intensity at the edge of the gold nanocluster that is accompanied by a deep valley corresponding to the center of the gold particle.

Since the size of the Fresnel fringes change in a defocus-dependent manner, we hypothesized that the optimal defocus observed for a 70 atom gold nanocluster was the result of constructive Fresnel fringe interference. This interference would increase the contrast of the gold nanocluster only if the Fresnel fringes were equivalent in size to the size of the gold nanocluster (14Å). To test this hypothesis, we constructed a synthetic amplitude-based model image containing a circle with a diameter of 14Å, where pixel values within the circle had a normalized intensity of -1 and all other pixels in the image had a normalized intensity of 0. Line traces through this synthetic amplitude image were equivalent to a top hat function in one dimension (Figure 3.11B, black line). After imaging this synthetic model image across a range of defocus values, the resulting line traces showed that 0.45 μm defocus exhibited the highest intensity peak (Figure 3.11B, blue line). Interestingly, the peak intensity from the 0.45 μm line trace showed an intensity of -1.37, indicating that the signal was amplified 37% by the CTF, from -1.00 to -1.37. These data provide further evidence that the optimal defocus observed for a 70 atom gold nanocluster is the result of constructive interference between the Fresnel fridges induced by the edge of the gold nanocluster.

While the interactions between the top hat function and CTF showed that the Fresnel fringes were sufficient to enhance the signal of a 14Å circle, it was not clear that the fringe size was the same as the gold nanocluster peak-width. To directly compare the Fresnel fringe size and observed gold peak-width at 0.45 μm, a step function was utilized instead of a top hat function as the synthetic amplitude-based model image (Figure 3.11C, black line). This allowed the visualization of the fringe size and location relative to the observed gold nanocluster peak width. Compared to 0.15 μm, the Fresnel fringe from 0.45 μm defocus showed a peak width concomitant in size and location with the observed gold nanocluster (Figure 3.11C). In contrast, the Fresnel fringe from 0.15 μm showed a smaller fringe size and higher frequency of oscillation, preventing the CTF-mediate amplification of signal. Therefore, these image simulation data provide an objective metric for collecting cryo-EM images of Nanogold: images of Nanogold must be at 0.45 μm in order to take advantage of a CTF-mediated amplification of the Nanogold signal at 120 keV.

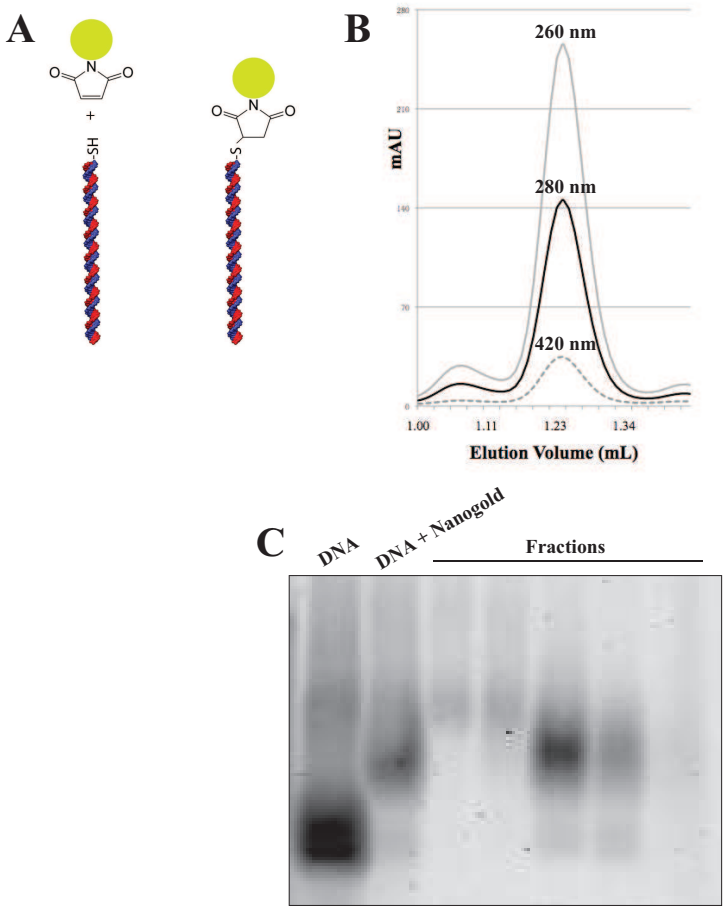


Figure 3.12: Covalent labeling of DNA using 1.4-nm maleiomide-Nanogold. (A) Strategy for labeling promoter DNA with Nanogold. Terminal thiol moieties are incorporated into DNA to provide specific gold labeling location. (B) Chromatogram of DNA-Nanogold monitored at 260, 280, and 420 nm. (C) 2% agarose gel of DNA-Nanogold labeling reaction.

Covalent and non-covalent labeling of SCP DNA

In order to label DNA with Nanogold, two strategies were developed that involved either covalent or non-covalent linkages of the Nanogold to the DNA sample. Covalent labeling of the DNA involved the incorporation of a thiol moiety at the 5' or 3'-termini of the promoter DNA (Figure 3.12A). After incubation of 1.4-nm monomaleiomide-Nanogold with thiol-containing SCP DNA, the sample was purified through size exclusion chromatography (Figure 3.12B & C). Strong absorbance of the sample at 420 nm indicated that the DNA was labeled with Nanogold, where the labeling efficiency was estimated to be 70%. The covalent conjugation of Nanogold to the SCP DNA caused a shift in molecular weight when visualized using gel electrophoresis (Figure 3.12C). This higher molecular weight complex was coincident with the strong aborbance at 420 nm on the chromatogram and was used for cryo-EM sample preparation.

Compared to the covalent labeling protocol for DNA, non-covalent approaches proved much more challenging due to aggregation of the Nanogold reagent. Biotin-labeled DNA was utilized for these experiments in an attempt to assemble a streptavidin(SA)-Nanogold-DNA ternary complex(Figure 3.13A). Unlike maleiomide-Nanogold, which was shipped in lyophilized form, SA-Nanogold was shipped resuspended in solution since the SA-Nanogold complex could not be frozen. When freshly ordered SA-gold was incubated with biotinylated-DNA, a clean band shift was observed indicating that the streptavidin bound in a stoichiometric manner (Figure 3.13B, lanes 1 & 2). This assembled complex proved to be very labile because, after 72 hrs, the complex was degraded and both the labeled DNA and unlabeled DNA species precipitated over time (Figure 3.13B, lanes 2 & 3). Furthermore, when this reaction was performed 1 month after ordering SA-Nanogold, higher order DNA-SA-gold complexes were visualized on the gel (Figure 3.13B, lane 4). This suggested that the SA-Nanogold was precipitating over time because there was less 'freely available' SA-Nanogold for binding to the DNA. This resulted in the formation of higher order DNA-SA-Nanogold complexes because there was no longer a 10-fold excess of SA-Nanogold to the DNA. These data indicate that non-covalent strategies for labeling DNA were less robust than the covalent methods and, accordingly, all Nanogold-DNA labeling experiments were performed with covalent Nanogold-DNA complexes.

Non-covalent labeling of TFIIA

The presence of multiple cysteine residues within TFIIA, one of which that makes contacts with TBP [5], excluded the possibility of pursuing a covalent labeling approach with maleiomide-Nanogold. Therefore, in order to label TFIIA with Nanogold, a non-covalent strategy was implemented that utilized the 5x-histidine tag on the gamma subunit of TFIIA to bind 1.8nm-Ni^{2+}-Nanogold (Figure 3.14A). After screening a variety of conditions, the appropriate buffer and reaction time was determined to yield a labeled TFIIA-Nanogold complex. When comparing the migration of TFIIA alone vs. TFIIA-Nanogold using size

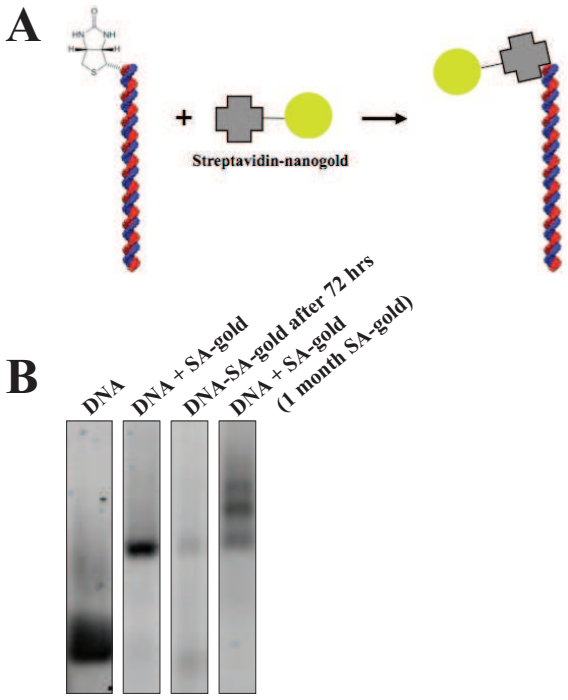


Figure 3.13: Attempted non-covalent labeling of DNA using streptavidin-Nanogold. (A) Strategy for labeling promoter DNA with streptavidin-Nanogold. Terminal biotin moieties are incorporated into DNA to provide specific gold labeling location. (B) 2% agarose gels of DNA-SA-Nanogold complexes.

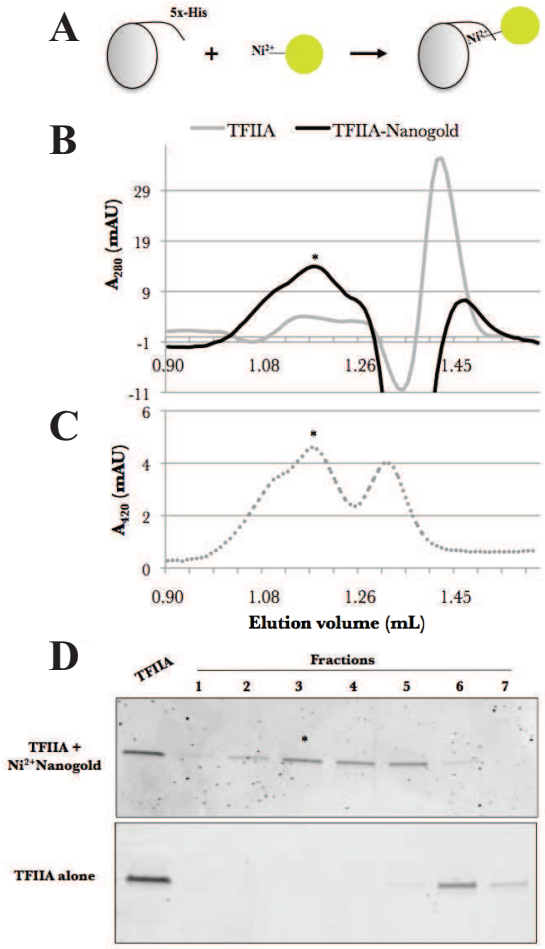


Figure 3.14: Non-covalent labeling of TFIIA using Ni^{2+}-Nanogold. (A) Strategy for labeling TFIIA with Ni^{2+}-Nanogold. TFIIA contains an engineered His(5x) tag in addition to a naturally occurring polyhistidine sequence. Chromatogram of TFIIA-Ni^{2+}-Nanogold purification monitored at 280 & 260 nm (B) and 420 nm (C). (D) SDS-PAGE gel analysis of fractions from gel filtration column. The presence of Ni^{2+}-Nanogold shifts the TFIIA peak dramatically into two separate peaks. The asterisk denotes the fraction used for cryo-EM grid preparation with TFIID and DNA. For clarity, only one of the two subunits of TFIIA is shown on the gel that corresponds to the fusion subunit α-β.

exclusion chromatography, there is a distinct shift in the size exclusion profile to a larger molecular weight (Figure 3.14B & D). This new peak is accompanied by the presence of strong absorbance at 420 nm (Figure 3.14C), thus indicating that Nanogold is present within this fraction. While non-covalent labeling TFIIA with 1.8nm-Ni^{2+}-Nanogold proved to be more successful than SA-Nanogold labeling of DNA, this approach still required much more optimization than the covalent DNA labeling with Nanogold, providing additional evidence that covalent labeling strategies should be pursued if possible.

Strategy for cryo-EM data collection & analysis

As indicated by the image simulations of gold nano-clusters (Figure 3.10), cryo-EM image processing of single particles labeled with Nanogold requires collecting defocal image pairs at high and low defocus (Figure 3.15A). This strategy will utilize the high defocus particles to guide the 2D reference-free alignment and classification, where the resulting alignment parameters will be transferred to the low defocus particles to investigate the localization of Nanogold. In order to extract the maximum amount of information from the particle focal pairs, the individual particle images were thresholded to highlight the location of Nanogold within the low defocus particle images. Both theoretical and experimental measurements of Nanogold in low defocus images revealed that the Nanogold intensity was routinely $> 4\sigma$ above the average image intensity. Therefore, to highlight the gold location within images, high defocus and low defocus particles were thresholded at $\sigma = 4$, where pixels $> 4\sigma$ are set to an intensity of 1 while pixels $< 4\sigma$ are set to an intensity of 0 (Figure 3.15B). These resulting particle images will be averaged in an identical manner to the high defocus particles in order to reveal the location of Nanogold within labeled TFIID-TFIIA-SCP complexes.

3.5 Defining the stereochemistry of the rearranged TFIID-TFIIA-SCP complex

To provide further evidence for the organization of DNA within TFIID-TFIIA-SCP, we examined the orientation of promoter DNA on TFIID by covalent labeling of the SCP DNA with 1.4-nm maleimido-Nanogold at either +45 or -36 (TATA box) on the SCP promoter sequence. This strategy takes advantage of the high contrast of Nanogold relative to that of protein and DNA at low defocus, providing a strong signal for the localization of gold particles relative to protein and DNA density. Analysis of the 2D reference-free class averages for high defocus particles of TFIID-TFIIA-SCP(+45 gold) and subsequently applying the alignments to the low defocus particles revealed that the downstream region of the SCP is bound by lobe C (Figure 3.16A). This is indicated by the narrow clustering signal from the low defocus thresholded particles (Figure 3.16A, right). When this localization is mapped

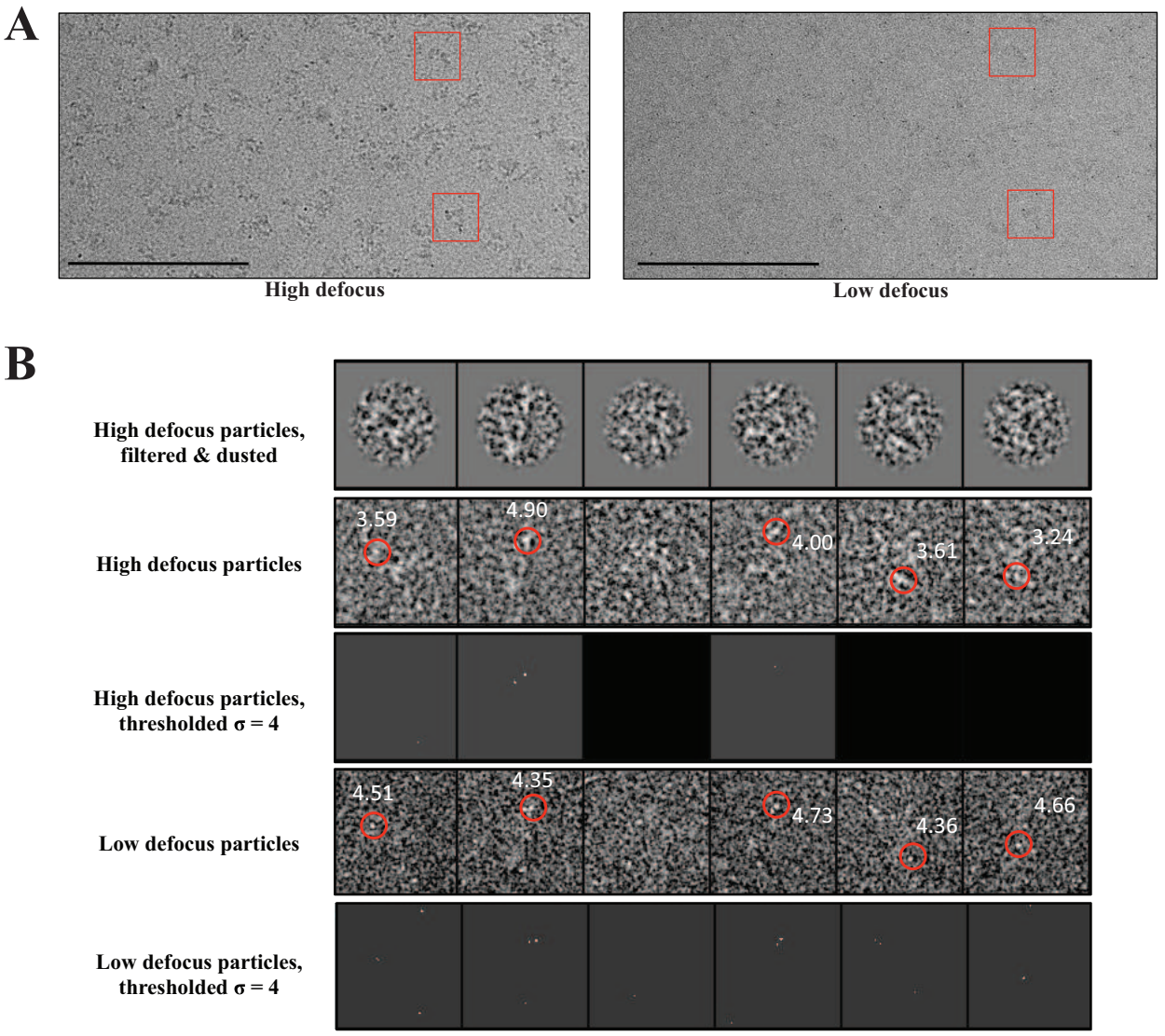


Figure 3.15: Data processing strategy for single particles labeled with Nanogold in vitreous ice. (A) Micrographs of TFIID-TFIIA-SCP(gold) at high defocus (left) and low defocus (right). Scale bar is 200 nm. (B) Individual particles from high defocus and low defocus micrographs. 2D reference-free alignments were performed on high defocus images that were filtered and processed to remove high intensity pixels, known as 'dusting' (top row). These alignments will be transferred to the low defocus particles. Consistent with image simulations of gold nano-clusters, Nanogold intensity is greater in the low defocus particle than the corresponding high defocus particle. This is indicated by red circles with peak intensity value.

onto the 3D structure of the rearranged conformation, the localization of +45 gold is consistent with the DNA extension experiment because the TFIID-TFIIA-SCP(+45 gold) average labels the DNA on a different face of the TFIID structure (Figure 3.6A). Moreover, the length of DNA between the +45 gold label and lobe C (Figure 3.6A) is 15 bps, the same distance between the DPE and +45 site on the SCP.

To provide an orthogonal labeling strategy for the TATA box, cryo-EM data were collected on a sample of TFIID-TFIIA-SCP(TATA gold) where the TATA box was covalently labeled with Nanogold (Figure 3.16B). When the high defocus alignments were applied to the low defocus thresholded particles, there was a cluster of signal within lobe A at a location consistent with DNA extension experiments. Thus, these SCP DNA labeling experiments, which are summarized in Figure 3.6A, have defined the organization of TFIID bound to promoter DNA, providing the first structural insight into promoter binding by human TFIID.

TFIIA localizes to lobe A in both the canonical and rearranged conformations of TFIID

We next investigated the location of TFIIA within the rearranged TFIID-TFIIA-SCP complex by analysis of TFIIA that was labeled with 1.8-nm Ni^{2+}-NTA-Nanogold. Upon 2D reference-free analysis of the high defocus cryo-EM particles, the determined alignment parameters were applied to the low defocus focal mate and class averages were calculated. Nanogold densities identified the TFIIA binding site within lobe A of the rearranged state (Figure 3.16C), adjacent to the TATA Nanogold location (Figure 3.16B). Importantly, analysis of 2D reference-free class averages corresponding to the canonical state also demonstrated that TFIIA localized to lobe A in its alternate configuration relative to the BC core (Figure 3.16D). These findings are in agreement with the well-characterized direct interaction of TBP and TFIIA, and further demonstrate, in the context of TFIID, that TFIIA interacts with the complex at a location close to the TATA box-binding site. These findings therefore suggest that both TBP and TFIIA are localized within lobe A during the structural transition between canonical and rearranged states of TFIID.

Model of promoter DNA path through TFIID-TFIIA-SCP

The results presented within this chapter have structurally characterized and validated the three-dimensional structure of TFIID bound to TFIIA and promoter DNA within the rearranged conformation. This process involved calculating an *ab initio* 3D reconstruction of the rearranged conformation (Figure 3.3) that was simultaneously refined in a multi-model fashion against cryo-EM data to provide both canonical and rearranged reconstructions of TFIID-TFIIA-SCP (Figure 3.5). Given the dramatic nature of the structural rearrangement, we wanted to provide an objective measure of the accuracy of the rearranged 3D model bound to promoter DNA. Therefore, we implemented the free-hand test to assess the

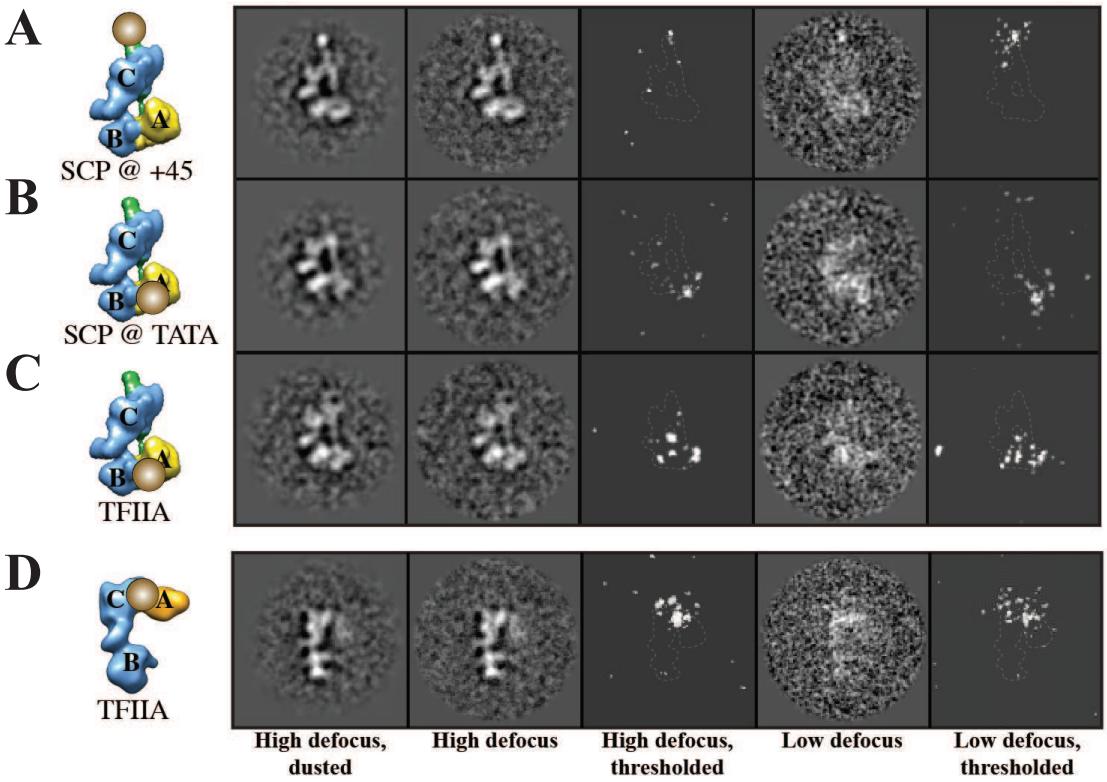


Figure 3.16: Organization of promoter DNA and TFIIA within the TFIID-TFIIA-SCP complex. 2D reference-free class averages for the rearranged TFIID-TFIIA-SCP complex with Nanogold labels on SCP DNA at +45 (A), SCP DNA at TATA (B), and TFIIA (C). (D) 2D reference-free class average for the canonical state of TFIID-TFIIA-SCP containing Nanogold labeled on TFIIA. 3D models are shown alongside high defocus averages (density threshold at $\sigma = 3.5$) with a gold sphere marking the localization of Nanogold for each experiment.

model handedness and quality of particle alignments. This test confirmed that the handedness of the rearranged model was correct, in addition to revealing that the model was able to align > 80% of the tilted particles into the overall correct orientation (Figure 3.9B).

After confirming the accuracy of the rearranged model of TFIID-TFIIA-SCP, labeling experiments were used to define the orientation of DNA and TFIIA on TFIID. By integrating these labeling data within the structure of the DNA-bound rearranged state, we have developed a model to describe the DNA path through the modular sub-domains of the TFIID-TFIIA-SCP complex (Figure 3.17). Gold labeling of the SCP DNA at the +45 position indicated that the density extending off of lobe C is the downstream promoter DNA (Figure 3.16A). It thus appears that the MTE and DPE promoter elements are bound by lobe C (Figure 3.17A), which suggests that lobe C contains subunits TAF6 and TAF9 [14, 106]. Further support for this conclusion came through antibody localization of α-TAF6 antibody to lobe C (Appendix A, Figure A.2). Given the linear nature of the DNA contacts with lobe C, we can conclude that the histone-fold containing subunits TAF6 and TAF9 do not interact with DNA in a nucleosome-like manner. This is consistent with recent biochemical studies that reconstituted a TAF6/9/4/12 tetrameric complex that was capable of interacting with DPE-containing promoter DNA. Interestingly, DNA-binding was only disrupted when loops extending away from the histone fold domains of TAF6 and TAF9 were truncated, but not when mutations were introduced within the core histone fold domains [99]. Thus, the histone-fold containing subunits TAF6 and TAF9 are located within lobe C and interact with promoter DNA in a linear manner that is inconsistent with nucleosome-like distortions of DNA.

Across the central channel from lobe C, lobe A interacts with SCP DNA extending from the Inr to the TATA box (Figure 3.17A). Previous work demonstrated that this region of the core promoter interacts with a TAF1, TAF2, and TBP sub-complex that is also able to direct transcription from TATA-Inr promoters [17]. Given the modularity of lobe A within the context of TFIID, it is likely that lobe A contains TAF1, TAF2, and TBP and serves as a modular domain of TFIID that is capable of interactions with promoter DNA upstream of the TSS.

This proposed composition of lobe A is also consistent with studies addressing the integrity of the TFIID complex *in vivo* [118]. Through systematic RNAi-mediated knock-down of TBP and TAFs within Drosophila S2 cells, the authors defined a stable sub-complex of TFIID that was nucleated by TAF4. Moreover, RNAi knock-down of TAF1, TAF2, or TBP did not affect the integrity of the TFIID complex, suggesting that these subunits were located on the periphery of TFIID. We believe that these data provide independent support for the conclusion that lobe A comprises TAF1, TAF2, and TBP, existing as a modular domain of TFIID. Furthermore, this study identified the composition of the stable TFIID sub-complex to be TAF4, -5, -6, -9 and -12. We propose that this stable sub-complex corresponds to the BC core identified in this study, providing a framework for further studies addressing the contributions of the BC core to TFIID function.

To model the DNA path through lobe A, two bends were incorporated to accommodate the 120° angle that TFIID imposes on the downstream and upstream DNA regions

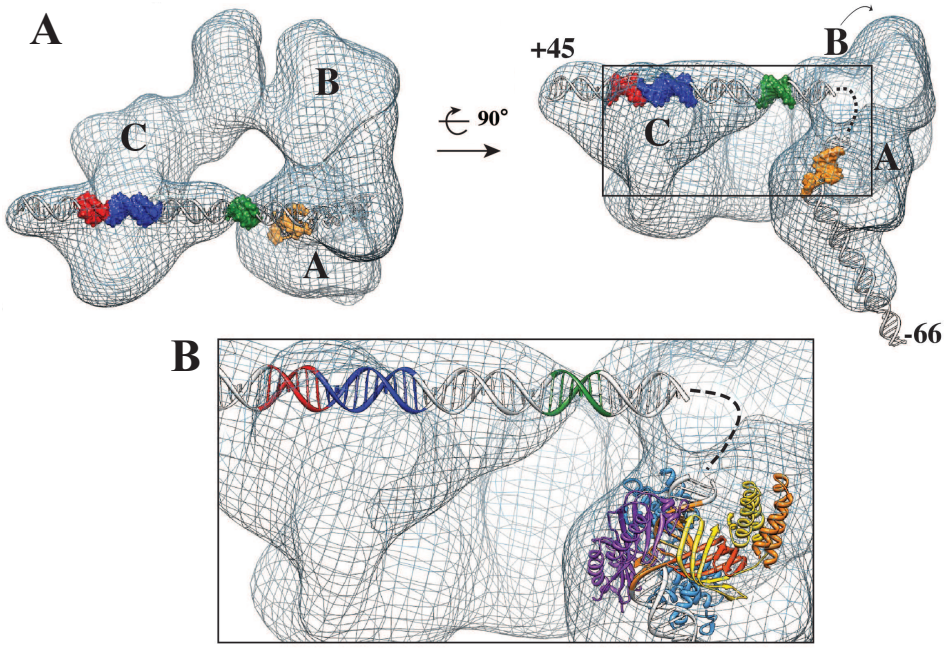


Figure 3.17: Structural description of the rearranged TFIID-TFIIA-SCP complex relative to the TSS. (A) Promoter DNA for SCP(-66) docked into the TFIID-TFIIA-SCP(-66) map (shown in mesh). DNA model corresponds to sequences from -66 to +45. (B) The proposed location of the crystal structures of TBP-TFIIA-TFIIB on TATA box DNA is in close proximity to Inr (PDB accession codes 1VOL and 1NVP). The unresolved DNA path between Inr and TATA is indicated by a dotted line.

(Figure 3.17A). Initial attempts to trace the DNA path through the TFIID-TFIIA-SCP structure using a single bend at the TATA box failed to generate a model that was compatible with the experimentally obtained positions of downstream and upstream DNA locations (data not shown). Given the angle of upstream DNA and the location of the -66 position, we modeled the location of TBP and TATA box DNA to be 36 bp (12.2 nm) downstream from the -66 position observed in the TFIID-TFIIA-SCP(-66) structure. After incorporating this bend at -31/30, a second gradual bend is proposed to form along the DNA between the TATA box and Inr. We suggest that this DNA deformation between the TATA box and Inr could play a role in positioning TBP and TFIIB in close proximity to the TSS for efficient loading of RNAPII.

Given that the above structural analyses did not provide a direct probe for the location of the TATA box within lobe A, we collected more cryo-EM data for TFIID-TFIIA-SCP(TATA gold) to try to localize extra density corresponding to the Nanogold within the high defocus particles. This study was motivated by the strong additional density that appeared at the +45 location on the SCP within the high defocus class average for TFIID-TFIIA-SCP(+45 gold) (Figure 3.16A). After collecting and analyzing a large dataset for TFIID-TFIIA-SCP(TATA gold), a strong additional density appeared on lobe A within the high defocus averages of rearranged state bound (Figure 3.18A & B, left). When 2D difference maps were calculated between these reference-free class averages and TFIID-TFIIA-SCP class averages that did not contain Nanogold, a strong difference peak ($> 4\sigma$) was observed. The strength and size of this difference peak indicates that the additional density seen within class averages of TFIID-TFIIA-SCP(TATA gold) corresponds to Nanogold. This finding is in close agreement with the cluster of peaks seen in the thresholded low defocus particles (Figure 3.16B).

Since two different orientations were found to contain this additional density, alignment of projections from the rearranged 3D model from TFIID-TFIIA-SCP allowed back-projection of the difference density (Figure 3.18C - E). Superimposing the back-projection of the two related, but distinct, orientations of the difference densities provided a precise location of the TATA box within the rearranged TFIID-TFIIA-SCP(-66) structure. It is important to consider that the Nanogold was covalently linked to the SCP sequence at -36, placing the Nanogold 15Å upstream of the TATA box location (-31/30). Furthermore, the Nanogold was tethered to the DNA through a six carbon linker (9Å). Thus, there is approximately 25Å between the TATA box and the Nanogold label. Modeling the size and location of Nanogold within the previously proposed model of TFIID-TFIIA-SCP(-66) shows that the modeled location of the TATA box is within 20Å of the experimentally determined TATA box location (Figure 3.18F). Given this new localization data, we should be able to provide a more accurate model of the DNA path through lobe A. However, it is difficult to utilize this new labeling information while also using the characteristic kink of DNA within TBP-TATA [48, 50] and the extension of upstream DNA exiting lobe A seen within the structure of TFIID-TFIIA-SCP(-66).

Precise modeling of the promoter DNA path and topology through TFIID will require determining a structure of TFIID-TFIIA-SCP at a higher resolution (15Å) in order to visu-

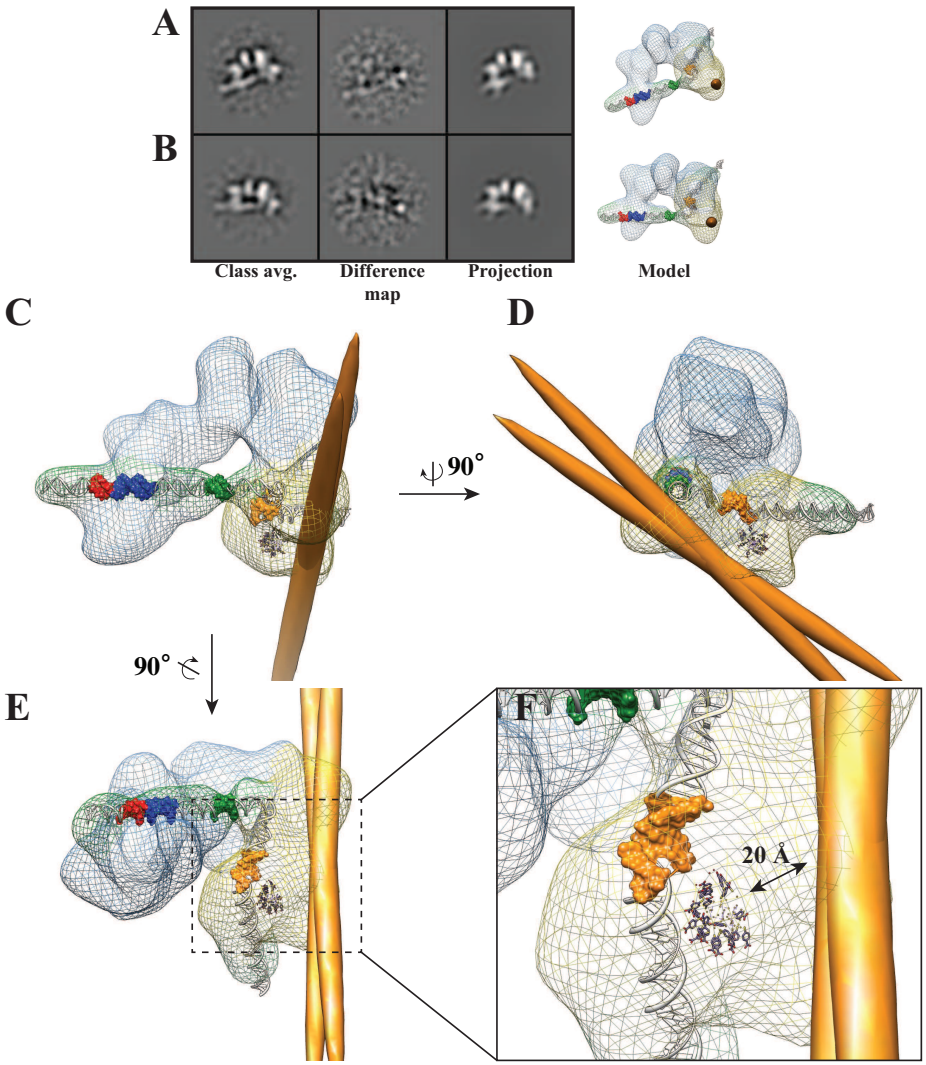


Figure 3.18: Localization of TATA-Nanogold within high defocus class averages. (A) & (B) 2D reference-free class averages for high defocus, dusted particles from TFIID-TFIIA-SCP(TATA gold) aligned with 2D difference map and model projection. The 2D difference map was obtained upon comparison of these averages with TFIID-TFIIA-SCP averages, where the difference density exhibits $> 4\sigma$ intensity value, indicating significance. Considering that these two averages were obtained from different Euler angles, back-projection of the difference density reveals location of TATA gold within lobe A (C - E). (F) A 70 atom gold nano-cluster [43] is docked as a reference for 1.4 nm-Nanogold size and location. Note that the location of Nanogold is approximately 20 Å away from the intersection of the back-projected difference densities.

alize the major and minor grooves of the promoter DNA. The structures presented here were unable to surpass 32Å resolution (Figure 3.5). Since these structures were calculated from a dataset containing 34,167 particles, we hypothesized that collecting a much larger dataset should increase the resolution of the structures. After collecting and analyzing a dataset that contained a combined total of 150,000 particles, 3D model refinements did not result in any detectable resolution improvement (data not shown). This suggests that high resolution structural studies of TFIID-TFIIA-SCP will require improving particle contrast within the cryo-EM images, since the presence of glycerol, sugars, and a carbon support likely limit the SNR of the individual particles. Recent advances in direct electron detection technology [16, 72] in combination with phase-plates [76] hold the promise of increasing the SNR for cryo-EM images and will likely be necessary tools to reach higher resolution.

Chapter 4

Biochemical and structural support for a general model of promoter binding by TFIID

The structural studies presented in Chapters 2 & 3 have provided the foundation for understanding the nature of TFIID's interaction with promoter DNA. However, in order to arrive at a complete description of TFIID bound to promoter DNA, biochemical and structural studies were pursued in an attempt to propose a model of TFIID-TFIIA-SCP complex formation. To facilitate the biochemical studies, we engaged in a productive collaboration with James T. Kadonaga's laboratory at University of California - San Diego, where George A. Kassavetis performed all footprinting experiments shown in this chapter. The synergy between structural studies of TFIID-TFIIA-SCP and biochemical footprinting have helped construct a detailed model describing potential mechanisms for regulating human TFIID's interaction with promoter DNA.

This chapter addresses potential mechanisms of promoter binding by TFIID through a variety of experimental designs. First, detailed footprinting analysis using DNase I and MPE-Fe were used to probe the extent of contacts that TFIID makes with promoter DNA in a TFIIA-dependent and -independent fashion. These footprinting results were extended by testing the structural effect that core promoter mutations have on TFIID's structure. After developing a detailed model of TFIID-TFIIA-DNA, the canonical conformation was studied to investigate its DNA binding ability. Finally, cryo-EM structural analysis and footprinting established that the rearranged conformation is the TFIIB-binding conformation of TFIID, suggesting that the rearranged state is the conformation that loads RNAPII at the TSS. These results were synthesized to propose a model that addresses the interplay between TFIID's conformational landscape and DNA binding.

4.1 Detailed footprinting of TFIID-TFIIA-SCP

Cryo-EM structural analysis of the TFIID-TFIIA-SCP suggested that the rearranged state interacted with the downstream DNA core promoter elements MTE and DPE using lobe C, whereas the TATA box and possibly the Inr are bound by lobe A. To test and extend this model, we performed footprinting analyses of the TFIID-SCP complex in the presence or absence of TFIIA.

DNase I footprinting of TFIID-SCP showed an extended region of protection from -7 to +41, which is in agreement with previous footprinting data on TFIID-SCP (Figure 4.1A) [44]. The addition of TFIIA did not alter the downstream interactions of TFIID with the Inr, MTE, and DPE core promoter elements, but did, however, result in a substantial increase in the binding of TFIID to the TATA box and flanking DNA sequences from -38 to -20 (Figures 4.1A, 4.2A & B). DNase I footprinting analysis of both DNA strands in the absence or presence of TFIIA revealed distinct patterns of protection and cleavage throughout the core promoter region. The results indicate that the DNA sequence between the Inr and MTE/DPE sites exhibits a phasing in DNase I sensitivity in which one face of the SCP DNA between Inr and MTE-DPE is susceptible to DNase I cleavage, whereas the opposite face remains protected (Figure 4.1C). Furthermore, the accessible face of the DNA exhibits DNase I hypersensitivity upon binding of TFIID (Figure 4.1A).

To obtain high resolution data on the interactions of TFIID with the core promoter, we carried out footprinting analyses with MPE-Fe, an intercalating agent that delivers Fe(II) for oxidation of the deoxyribose phosphate backbone of DNA and provides single bp resolution of protein-DNA contacts [36, 85, 110]. The MPE-Fe footprinting data on TFIID-SCP and TFIID-TFIIA-SCP revealed the extensive and continuous interaction of TFIID with DNA from the DPE through the Inr as well as the TFIIA-dependent protection of 8 bp (-31 to -24) of DNA encompassing the TATA box (Figures 4.1B, 4.2C & D). The strong stimulation of TFIID binding to the TATA box is consistent with the results of previous studies on the binding of TBP to DNA [27, 48, 50, 77]. The region of DNase I protection observed on only one face of the helix between the MTE/DPE and Inr also shows continuous protection from MPE-Fe(II) cleavage on both DNA strands, which is likely due to the inhibition of the DNA unwinding, necessary for MPE intercalation, that would occur as a result of protein bound to one side of the helix [109]. Thus, the DNase I and MPE-Fe footprinting results, summarized in Figure 4.1C, provide insight into the TFIID-DNA contacts that complements the cryo-EM data and contributes to the placement of the core promoter DNA on the TFIID structure.

The footprinting data provide new structural insight into the model of DNA through TFIID-TFIIA-SCP. On each strand between the Inr and MTE, there is a distinct 10 bp phasing of DNase I hypersensitive sites at -2, +18, and +28 (with probe DNA 5-labeled upstream) and at +4 and +13 (with probe DNA 5-labeled downstream) (Figure 4.1A). Docking the DNase I footprinting model of protection (Figure 4.1C) into the cryo-EM model of TFIID-TFIIA-SCP allows for one side of the DNA sequence to face away from the central cavity (DNase I sensitive sites) while the opposite side of the helix faces the inner cavity of TFIID (DNase I protected sites) (Figure 4.3). Hence, these data suggest that TFIID

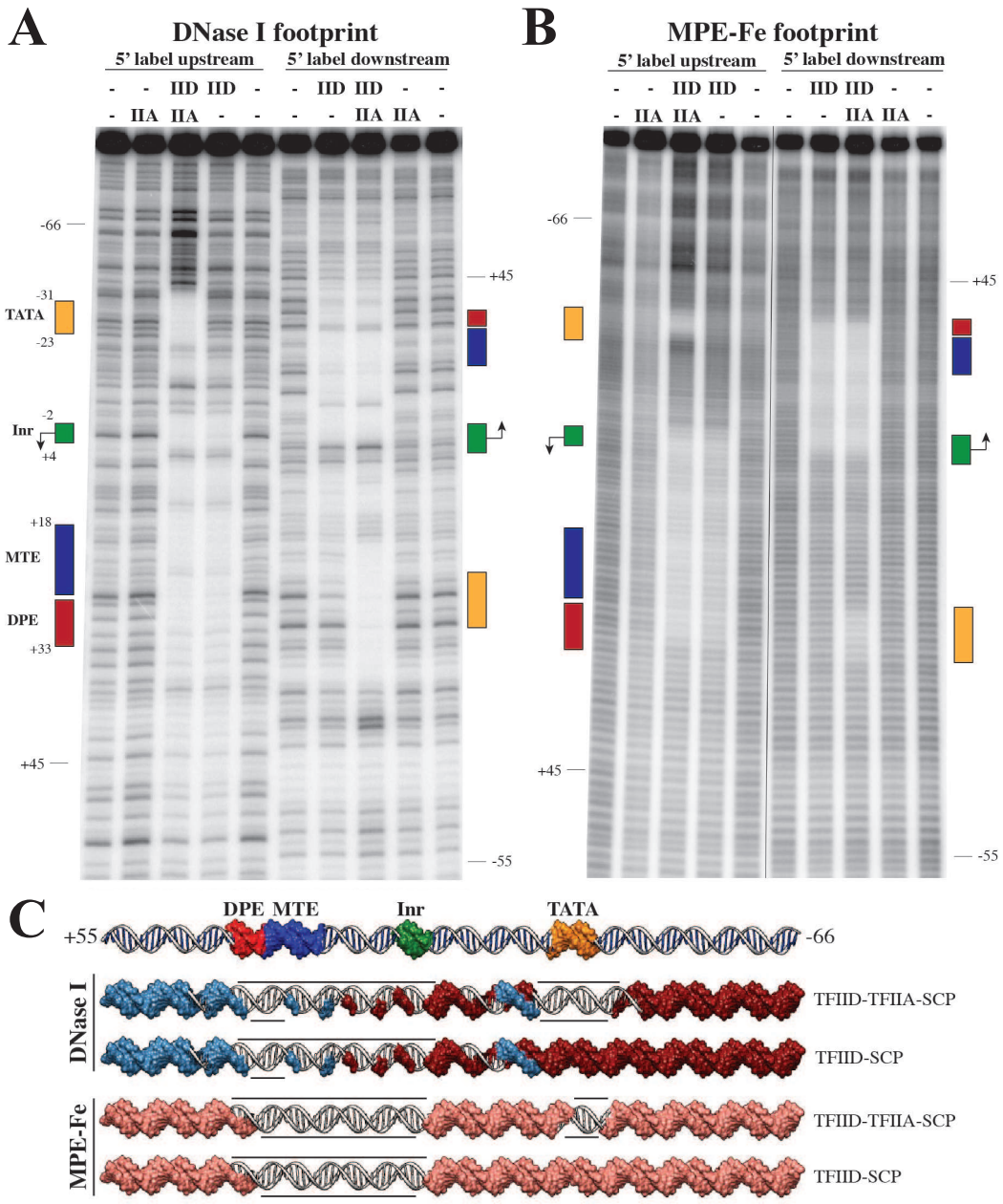


Figure 4.1: TFIID exhibits TFIIA-independent and TFIIA-dependent interactions with SCP DNA. DNase I (A) and MPE-Fe (B) footprinting of TFIID-SCP and TFIID-TFIIA-SCP. (C) B-DNA model of the SCP with positions of core promoter elements (top), DNase I, and MPE-Fe protection patterns (-66 to +55 shown) color-coded for TFIID-SCP and TFIID-TFIIA-SCP (four bottom rows). White base pairs indicate protection (also marked by black lines), blue surfaces indicate partial digestion and red surfaces indicate complete digestion by DNase I, and pink indicates digestion by MPE-Fe.

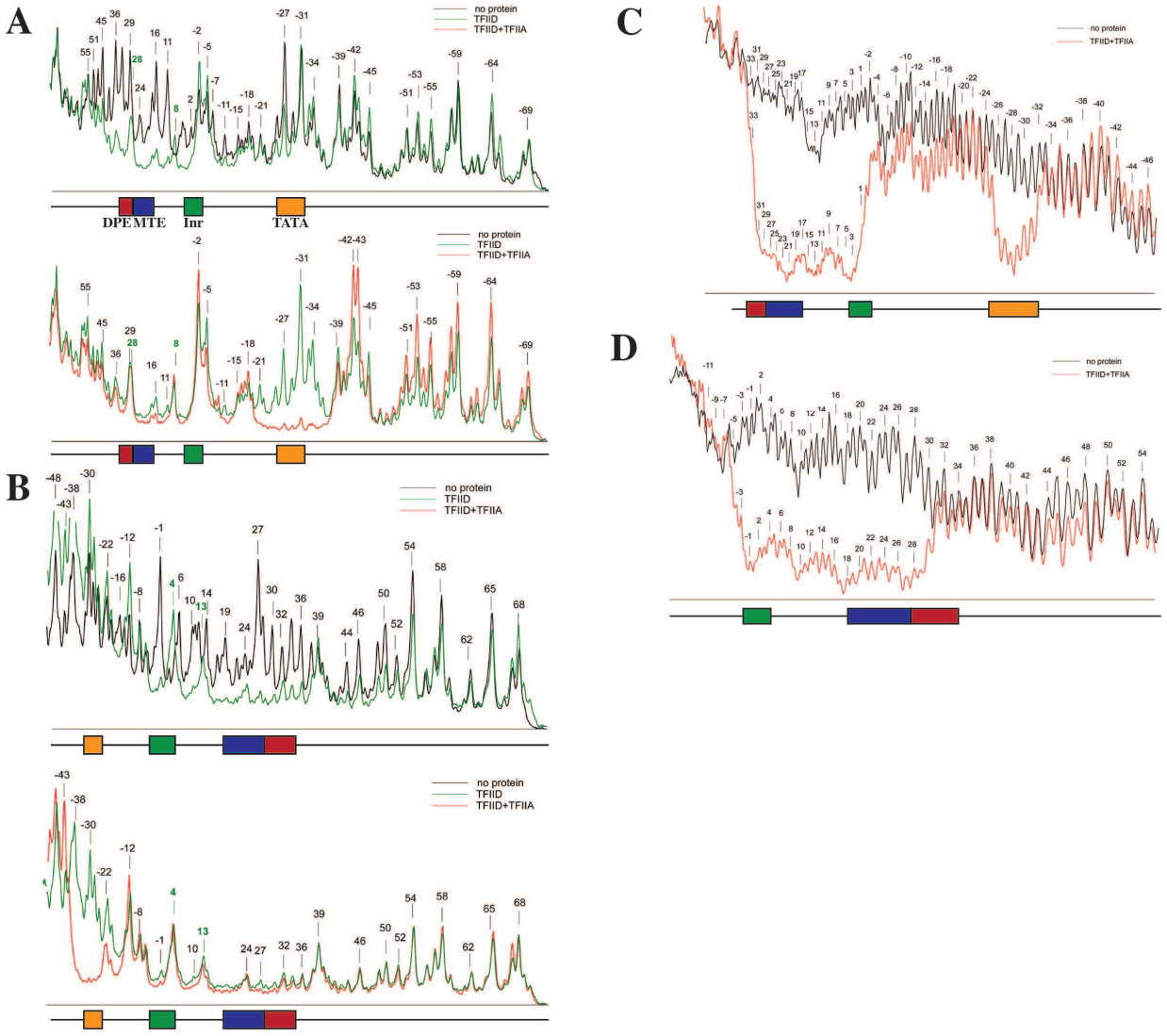


Figure 4.2: Line profiles for DNase I footprinting gels of TFIID-SCP and TFIID-TFIIA-SCP. (A) Line profile traces from DNase I footprinting gels for 5-label upstream (A) and downstream (B). Bold base pair numbers indicate hypersensitive sites. Line profile traces from MPE-Fe footprinting gels for 5-label upstream (C) and 5-label downstream (D).

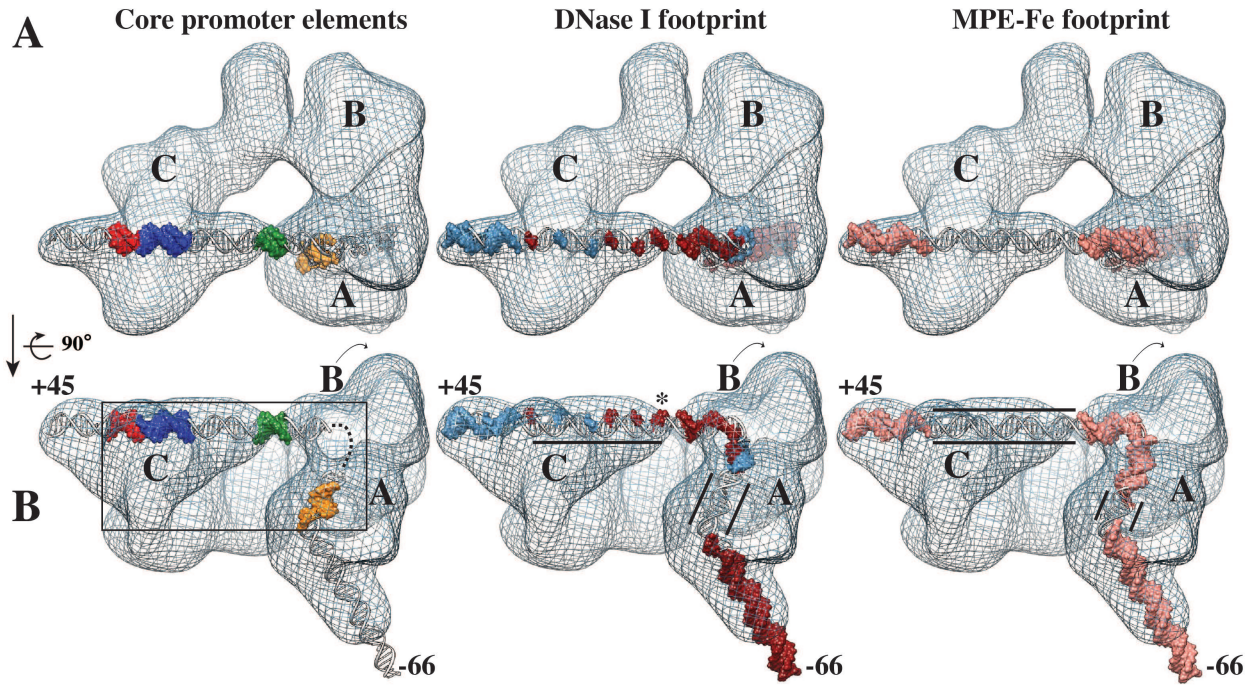


Figure 4.3: Structural model of DNase I and MPE-Fe footprinting results within the ternary TFIID-TFIIA-SCP(-66) complex. Promoter DNA for SCP(-66) docked into the TFIID-TFIIA-SCP(-66) map (shown in mesh). DNA models were taken from Figure 4.1C for sequences from -66 to +45. Asterisk in (B) indicates DNase I hypersensitive site at +3. Black lines in (B) indicate regions of continuous protection along SCP helix.

presents the DNase I-accessible face of this region of the core promoter to the bulk solution for interactions with RNAPII and other factors involved in transcription initiation.

As the DNA extends across the central channel towards lobe A, the footprinting data suggest that there are topological changes in the promoter DNA surrounding the TSS. For instance, the DNase I footprinting experiment shows a hypersensitive site at +3, indicating that the DNA has changed conformation for optimal cleavage by Dnase I only when bound to TFIID (Figures 4.1A). Given that the strength of this hypersensitive site appears to correlate with the strength of transcription initiation from a given promoter [106], it is interesting that this site is exposed within the central channel of TFIID, suggesting that TFIID-induced topological changes of DNA within the central channel may be relevant for later steps in transcription initiation.

The modeled path of promoter DNA through TFIID-TFIIA-SCP suggested that, in addition to the kinking of the TATA box by TBP, there is an additional bend in promoter DNA (Figure 4.3). While it is difficult to model the proposed DNA bend accurately at this resolution, the MPE-Fe footprinting indicate that the intervening DNA between TATA and Inr showed sensitivity to MPE-Fe, suggesting that the DNA helix is not distorted and there are not strong protein-DNA contacts. Therefore, while the DNA must bend through this part of the structure, it likely follows a gradual bending path as it enters into lobe A.

4.2 TFIID interacts with diverse promoter architectures through the rearranged conformation

While the SCP DNA has served as an important tool for structurally dissecting the structure of TFIID-TFIIA-SCP, the presence of four consensus promoter motifs within the SCP represents a non-physiological arrangement, since most human promoters contain only one or two consensus motifs [46]. To explore the effect of core promoter architecture on TFIID-promoter interactions, DNase I footprinting experiments were performed with mutant SCP DNA constructs in the presence or absence of TFIIA. Mutation of the TATA box within the SCP sequence (mTATA) resulted in a wild-type interaction with the promoter DNA from the Inr to the DPE, as seen previously (Figure 4.4) [44]. In addition, the inclusion of TFIIA resulted in a weak but detectable footprint over the mutant TATA box (Figure 4.5). The strong resemblance between the the protection patterns of the downstream Inr-MTE/DPE region from mTATA and the wild-type SCP suggested that TFIID is bound to the mTATA sequence in the rearranged conformation. To test this hypothesis, we collected cryo-EM data and visualized a sample of TFIID-TFIIA-SCP(mTATA) (Figure 4.6). This experiment revealed that TFIID binds to the mTATA promoter in a nearly identical conformation as that observed with the wild-type SCP sequence. Thus, the combined footprinting and EM data indicate that the rearranged state of TFIID serves as the predominant DNA binding

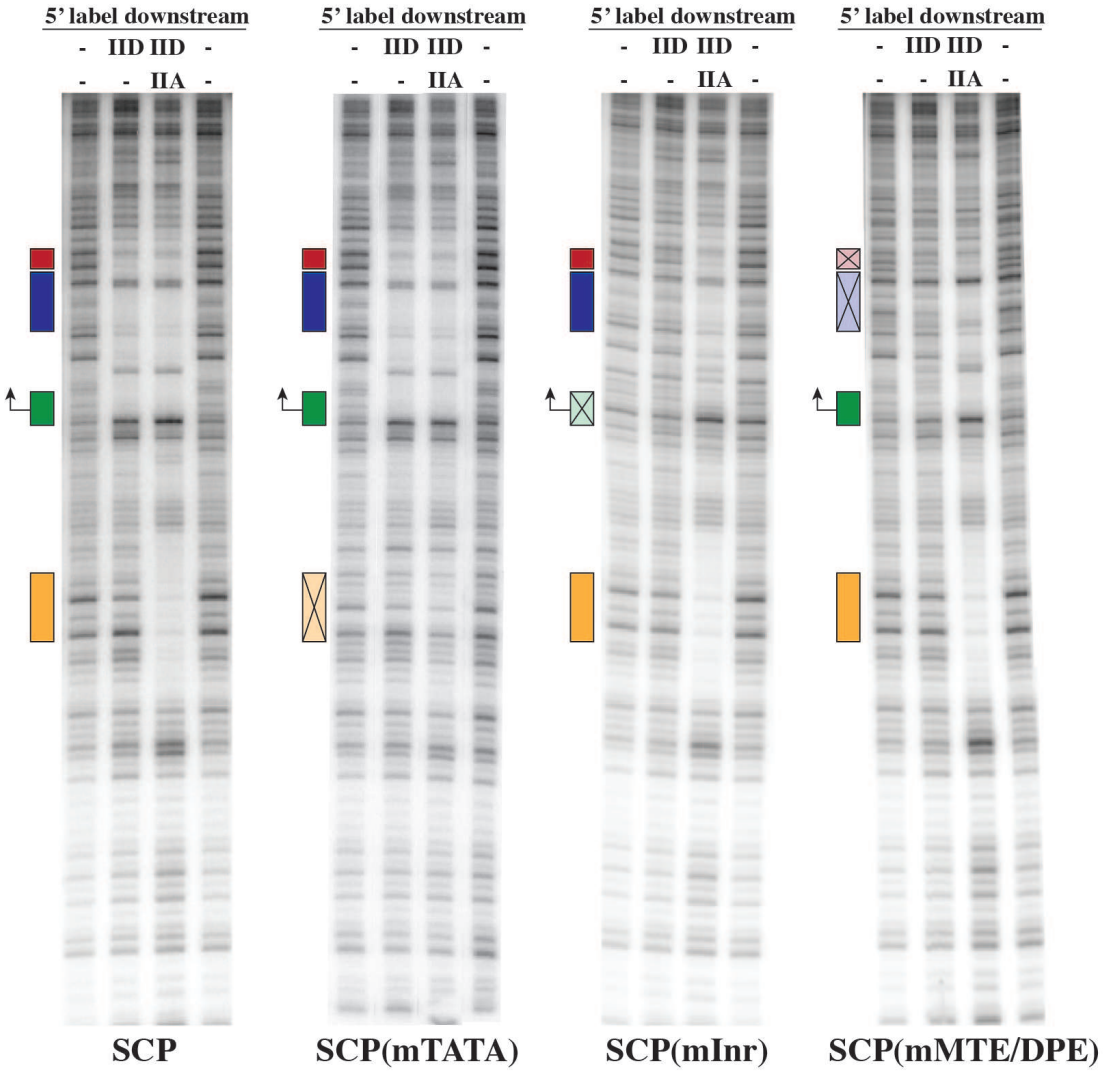


Figure 4.4: Core promoter architecture dictates TFIIA-dependent and TFIIA-independent interactions of TFIID with core promoter DNA. DNase I footprinting on wild-type and mutant SCP DNA sequences. 5'-labeled downstream probes were analyzed for DNase I protection in the presence or absence of TFIIA for wild type, mutant TATA (mTATA), Inr (mInr), and MTE/DPE (mMTE/DPE).

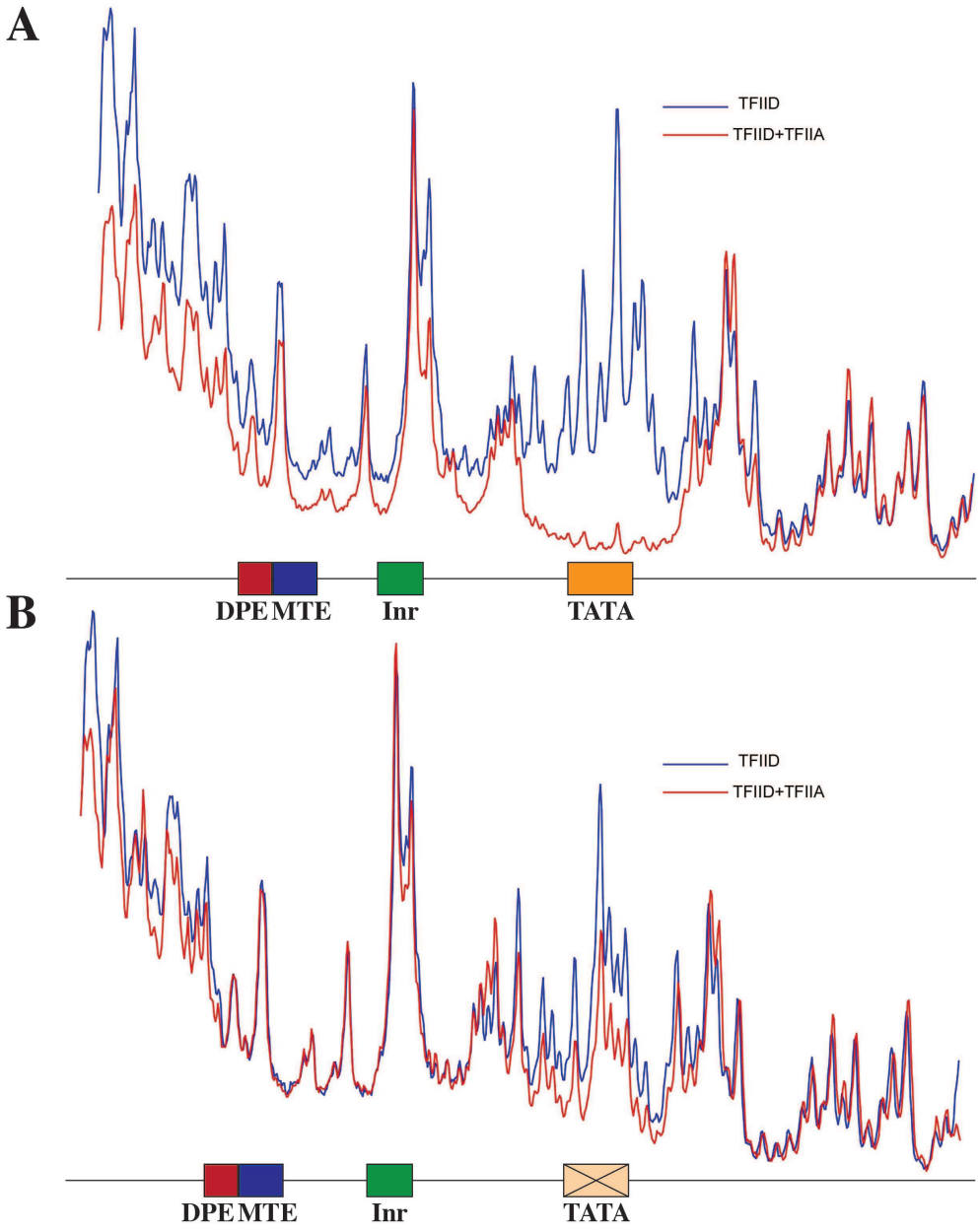


Figure 4.5: Line profiles for DNase I footprinting on mTATA promoter. (A) Line profiles for TFIID (blue) and TFIID-TFIIA (red) on wild-type SCP. (B) Line profiles for TFIID (blue) and TFIID-TFIIA (red) on mTATA promoter.

conformation for the SCP and mTATA promoter architecture.

The conformation of promoter-bound TFIID was further addressed by analysis of promoters that contain mutations in the Inr (mInr) or MTE/DPE (mMTE/DPE) promoter motifs (Figure 4.4). TFIID did not interact appreciably with either the mInr promoter or the mMTE/DPE promoter in the absence of TFIIA, as seen previously [44]. However, the addition of TFIIA resulted in strong binding of TFIID to the TATA box as well as to sequences from the Inr through the DPE regions. In the presence of TFIIA, the overall patterns of protection observed with the mInr and mMTE/DPE promoters are similar to those seen with the wild-type SCP. It thus appears likely that TFIID-TFIIA binds to the mInr and mMTE/DPE promoters in the rearranged conformation. Hence, the footprinting and EM data both suggest that TFIID binds to the wild-type and mutant SCPs in the newly discovered rearranged conformation.

With the 'wild-type' SCP as well as with the three 'mutant' (mTATA, mInr, mMTE/DPE) versions of the SCP, we observed that TFIIA stimulates the binding of TFIID to the TATA box region (Figure 4.4). This effect is consistent with the well-established TFIIA-mediated enhancement of TBP binding to the TATA box [107]. With the mTATA promoter, the primary interaction of TFIID with the DNA is via the Inr, MTE, and DPE motifs, and a weak stimulation by TFIIA of the binding of TFIID to the mutant TATA box region is also observed. With the mInr and mMTE/DPE promoters, it seems likely that TFIIA stimulates the binding of TBP to the TATA box and that the remainder of the TFIID complex then interacts with the Inr through the DPE region of the core promoter, irrespective of the presence of consensus Inr or MTE/DPE elements. These findings may be analogous to the previously observed stimulation of the binding of partially-purified TFIID to the downstream promoter region of the adenovirus major late promoter (which lacks MTE/DPE motifs) by the upstream stimulatory factor, USF [95, 111]. In this light, it is possible that other sequence-specific activators, as well as coactivators, may function in a related manner to stabilize TFIID on promoter DNA and thus promote the formation of the rearranged conformation.

4.3 DNA binding within the canonical conformation

While the rearranged conformation is the predominant form of TFIID bound to promoter DNA, we next wanted to investigate if the canonical state is capable of interacting with promoter DNA. Specifically, we wanted to know if the canonical state was 'inhibited' for binding to the MTE/DPE, given the close proximity of lobe A and the MTE/DPE binding sites within the canonical conformation. This investigation into DNA binding by the canonical state is similar in analysis to the localization of TFIIA within lobe A (Figure 4.7A), where 2D reference-free class averages of the canonical state from cryo-EM of TFIID-TFIIA-SCP(gold) were generated and analyzed.

After collecting larger datasets for both +45 and TATA-Nanogold labeled samples, class averages of the canonical conformation for TFIID-TFIIA-SCP showed that the promoter

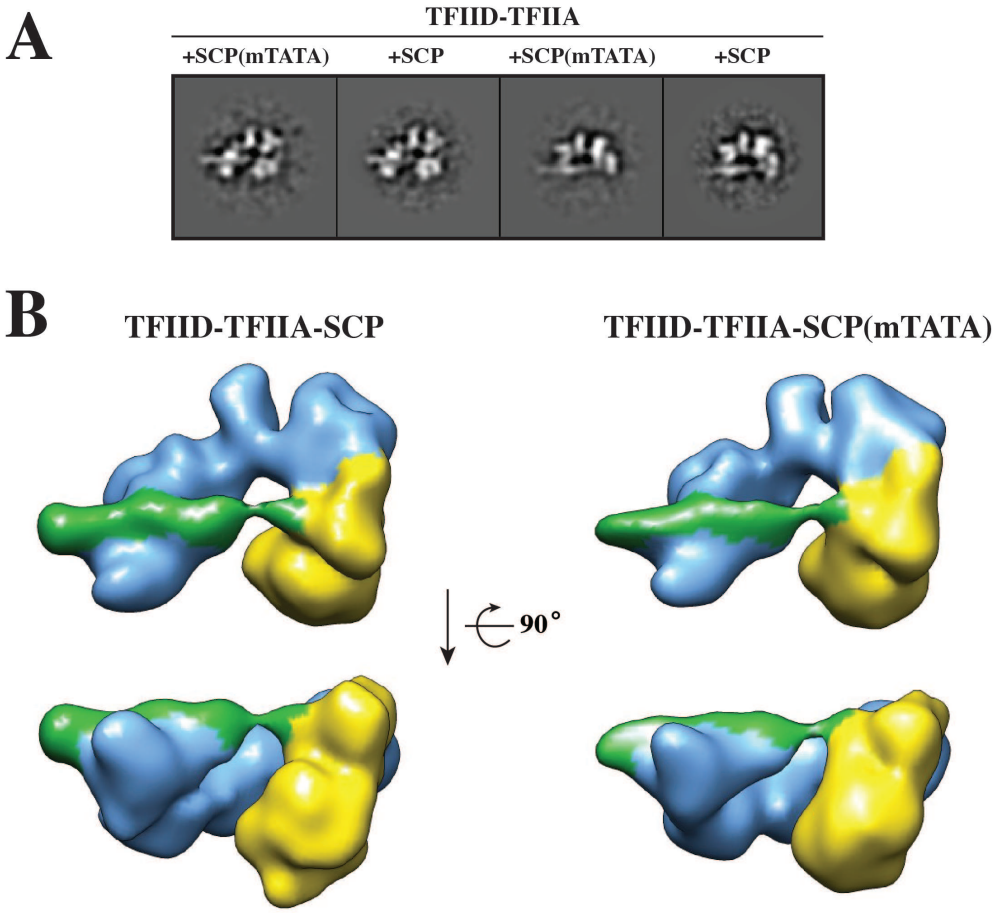


Figure 4.6: TFIID-TFIIA interacts with SCP(mTATA) within the rearranged conformation. (A) 2D reference-free class averages calculated from cryo-EM data of TFIID-TFIIA-SCP(mTATA) are shown alongside 2D reference-free class averages from TFIID-TFIIA-SCP. (B) 3D models of the rearranged conformation for TFIID-TFIIA-SCP and TFIID-TFIIA-SCP(mTATA) at 34Å.

DNA binds in a manner similar to the rearranged conformation. This conclusion comes from comparison of the low defocus thresholded average from +45 Nanogold (Figure 4.7B, right) with the high defocus class averages (Figure 4.7B, left). The presence of additional density extending away from lobe C in the high defocus average and the terminal localization of +45-Nanogold indicates that SCP DNA is bound. Furthermore, considering that the +45 is 15 bps away from lobe C within the rearranged conformation, the presence of a similar density extending away from lobe C in the canonical state suggests that the promoter DNA bound is in a near identical manner (Figure 4.7B). Unfortunately, the high degree of flexibility of lobe A did not allow careful comparison of lobe A's position and DNA binding, preventing conclusions from being drawn regarding potential inhibition of MTE/DPE binding by lobe A. However, despite this limitation, analysis of the canonical conformation from TFIID-TFIIA-SCP(+45 gold) indicated that the canonical conformation is capable of binding MTE/DPE DNA sequences within the same configuration as the rearranged state.

Since lobe C interacts with the MTE/DPE within the canonical state, we hypothesized that if lobe A engages the TATA box, it should occur in a position located away from lobe C. To test this, canonical class averages from TFIID-TFIIA-SCP(TATA gold) were analyzed for Nanogold localization (Figure 4.7C). Unlike the stable binding of MTE/DPE to lobe C, the Nanogold signal for TATA gold was diffusely localized along the trajectory of lobe A's rearrangement. Furthermore, considering the small cluster of peaks at a position opposite the central channel from lobe C, DNA binding in the canonical state appears to position the TATA box for binding by TBP within lobe A of the rearranged conformation (Figure 4.7C).

Analysis of the highly flexible canonical state using Nanogold labels for TFIIA and DNA has revealed insight into the dynamic process of DNA binding by TFIID (Figure 4.7). While it is difficult to know the position of lobe A within these Nanogold labeled class averages, the Nanogold labeling approach has provided key localizations in an otherwise flexible conformation. Despite the flexibility of lobe A, these labeling data indicate that the canonical conformation is competent for binding to promoter DNA. Lobe C binds to the MTE/DPE sequence in a similar fashion as the rearranged conformation, suggesting that its intrinsic DNA binding activity is preserved within the canonical state. Additionally, as suggested earlier, lobe A appears to exist as a modular TATA binding component of TFIID, where the TATA gold label localized to locations along the path of lobe A's rearrangement.

4.4 TFIID-TFIIA-TFIIB-SCP adopts the rearranged conformation on promoter DNA

The cryo-EM analysis presented thus far indicates that the predominant DNA binding form of TFIID is the rearranged conformation, suggesting that the remaining GTFs may bind the rearranged state for RNAPII loading. To test this hypothesis, cryo-EM samples were prepared with TFIID-TFIIA-TFIIB-SCP(-66) and analyzed (Figure 4.8A). 2D reference-free class av-

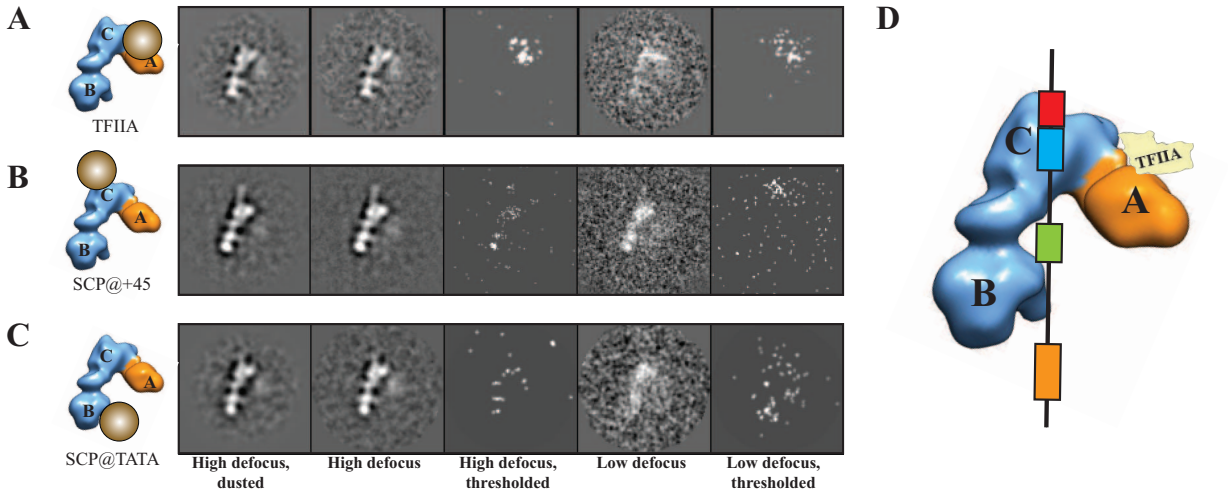


Figure 4.7: Lobe C interacts with MTE/DPE motifs within the canonical conformation. 2D reference-free class averages for the canonical conformation from cryo-EM data of TFIID-TFIIA-SCP with Nanogold labeled TFIIA (A), +45 (B), and TATA (C). Left of the averages is a model of the canonical state indicating location of Nanogold label. (D) Model of DNA path and TFIIA within the canonical conformation based upon gold labeling.

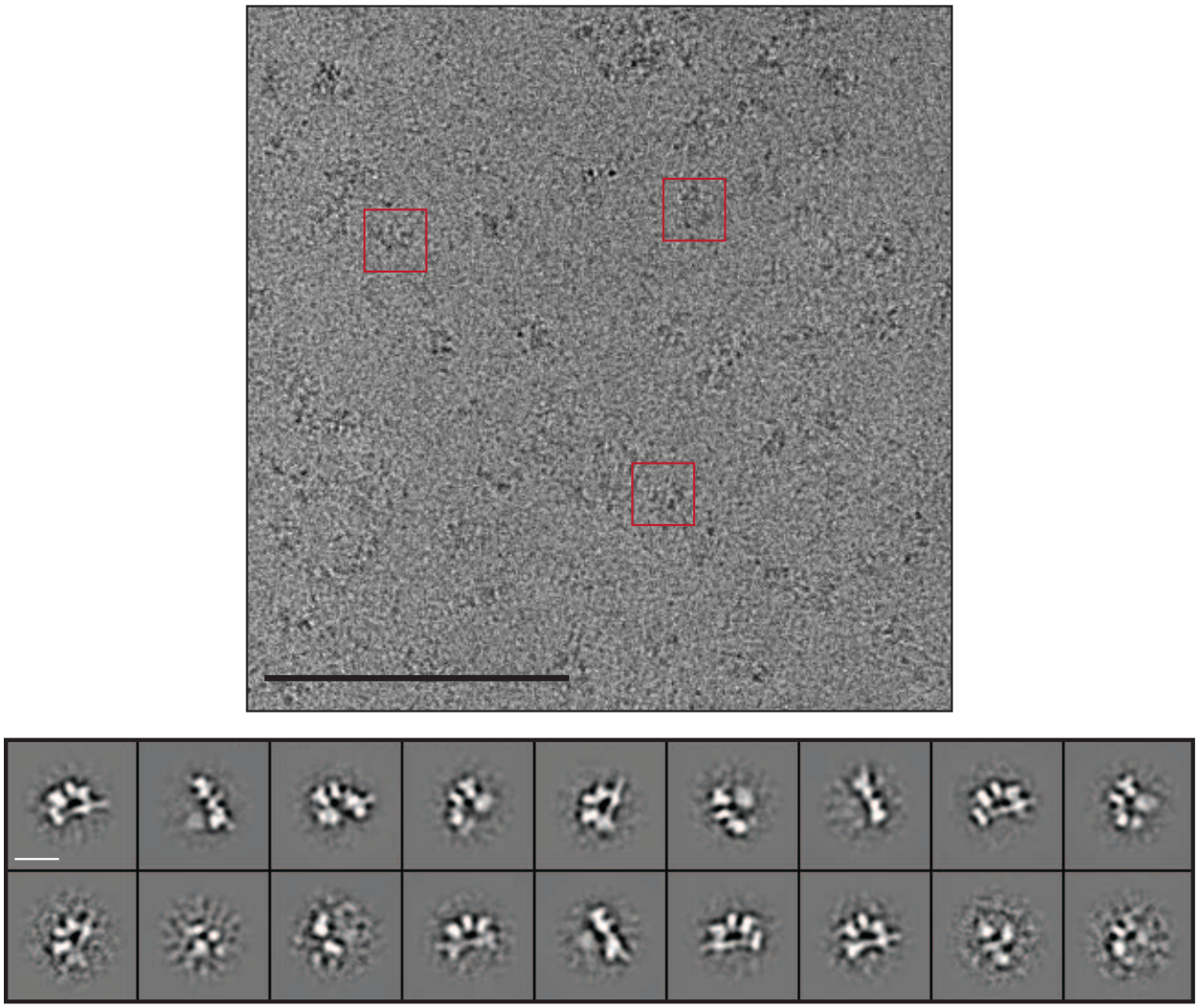


Figure 4.8: Cryo-EM of TFIID-TFIIA-TFIIB-SCP(-66). Representative micrograph (A) and 2D reference-free class averages (B). Scale bar is 200 nm in (A) and 200Å in (B).

erages showed that there was DNA bound to the rearranged conformation (Figure 4.8B), suggesting that TFIIB does not disrupt the DNA binding to the rearranged conformation.

To verify the binding of TFIIB to the rearranged state of TFIID-TFIIA-SCP, DNase I and MPE-Fe footprinting experiments were performed on TFIID-TFIIA-TFIIB-SCP. The footprinting results show that TFIIB extends the footprint surrounding the TATA, while the downstream contacts along the Inr, MTE, and DPE remain unchanged (Figure 4.9A & B). The DNase I patterns of protection indicate that an additional 8 - 10 bps are protected on either side of the TATA box, increasing the footprint around the TATA box from -45 to -18 (Figure 4.9A). This shows that nearly 30 bps of promoter DNA upstream of the TSS are sequestered within the TFIID-TFIIA-TFIIB-SCP complex. Furthermore, given that the

remaining downstream contacts with promoter DNA are unaffected by TFIIB, these data suggest that TFIIB preferentially interacts with the rearranged conformation. Considering that the TATA box was localized to lobe A, this result indicates that TFIIB localizes to lobe A within the cryo-EM structure of TFIID-TFIIA-TFIIB-SCP.

In addition to the DNase I footprinting experiments, MPE-Fe was used in order to provide high-resolution information on the protein-DNA contacts introduced by TFIIB within the TFIID-TFIIA-TFIIB-SCP complex. This analysis revealed that TFIIB makes intimate contacts with the flanking DNA sequences around the TATA box from -34 to -19 (Figure 4.9B). These results are consistent with the presence of a downstream BRE motif within the SCP sequence [44], a region of the promoter that is capable of engaging in sequence-specific contacts with TFIIB [21, 108]. Interestingly, however, there are upstream contacts from TATA box, even though the SCP sequence does not contain a functional upstream BRE motif [44]. Thus, TFIIB makes specific contacts with the upstream and downstream DNA sequences surrounding the TATA box, potentially stabilizing the TFIID-TFIIA-TFIIB-SCP complex due to an extended footprint on the DNA.

From the high affinity binding of TFIID-TFIIA-TFIIB-SCP to the TATA box and flanking sequences, we hypothesized that the this ternary complex would make stronger contacts with the SCP(mTATA) sequences. To test this, we performed DNase I footprinting of TFIID-TFIIA-TFIIB on wild-type and mutant TATA box promoters (Figure 4.10). The footprinting results revealed that mutation of the TATA box disrupts protein-DNA contacts along the TATA box and downstream BRE motifs in addition to the contacts upstream of the TATA box. Since TFIIB is unable to bind downstream BRE motifs in the absence TBP-TATA contacts (Figure 4.9A), the footprinting on SCP(mTATA) suggests a cooperative activity of TFIIB-TBP-TATA in order for high affinity binding to the TATA box and downstream BRE motifs. This explanation is consistent with the crystal structure of TBP-TFIIB-TATA, where the TFIIB simultaneously contacts TBP and the downstream BRE [108]. It should be noted, however, that there is slight protection of the TATA sequence, comparable to that previously observed (Figure 4.5), indicating that TBP may be within close proximity of the mutated TATA sequence. These footprinting data suggest that, within the context of a mutant TATA box, TFIIB does not stimulate stronger TATA box binding nor does it interact with the downstream BRE element.

The data above indicate that TFIIB binds the rearranged conformation in the presence of a functional TATA box and, given the extended footprinting surrounding the TATA box, the binding of TFIIB to TFIID-TFIIA-SCP may stabilize the ternary complex to enable high resolution structural studies using cryo-EM. Therefore, 3D refinements were performed on a sample of TFIID-TFIIA-TFIIB-SCP(-66) using previously obtained models corresponding to the canonical and rearranged conformations. After collecting and analyzing 56,000 cryo-EM particles of TFIID-TFIIA-TFIIB-SCP, the resulting 3D refined model for the rearranged state achieved a resolution of 36Å, comparable to that obtained previously for TFIID-TFIIA-SCP. To investigate if the binding of TFIIB to the ternary complex altered structural contacts between components within TFIID-TFIIA-SCP, the structure is shown alongside the previously obtained TFIID-TFIIA-SCP(-66) (Figure 4.11). The overall struc-

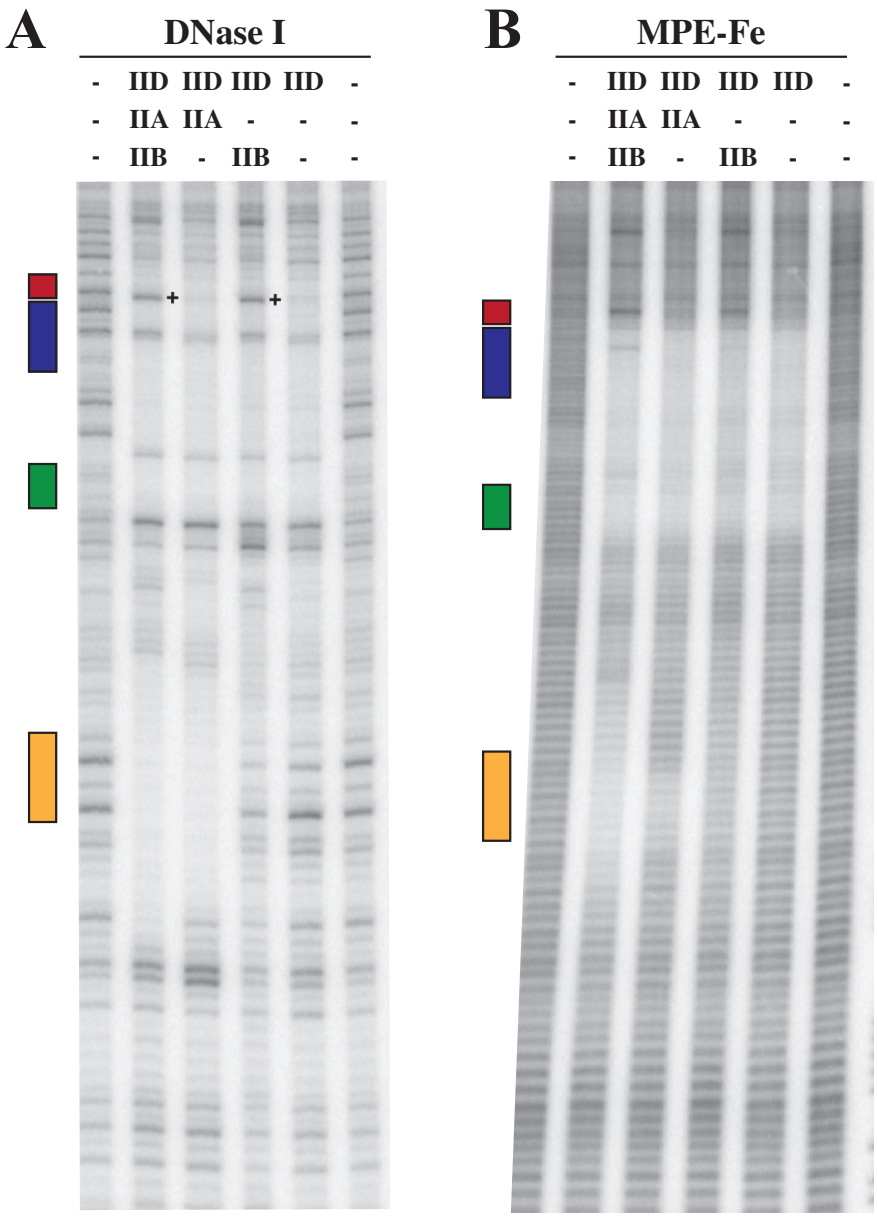


Figure 4.9: TFIIB interacts with TFIID-TFIIA-SCP within the rearranged conformation on promoter DNA. DNase I (A) and MPE-Fe (B) footprinting of TFIID-TFIIA-TFIIB-SCP on 5'-labeled downstream DNA.

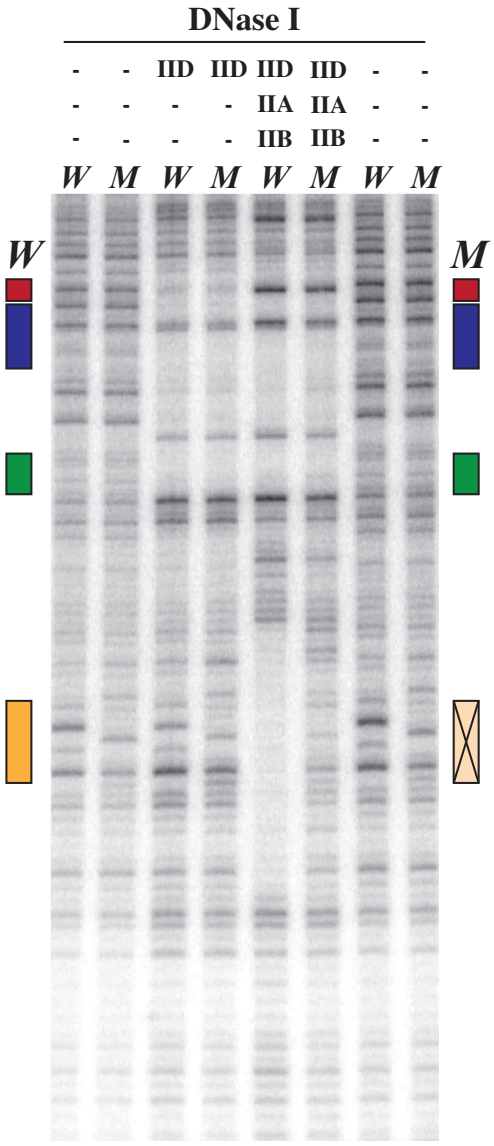


Figure 4.10: TFIIB does not induce strong TATA box protection within TFIID-TFIIA-TFIIB-SCP(mTATA). DNase I footprinting of TFIID-TFIIA-TFIIB on SCP(mTATA). *W* and *M* indicate wild-type SCP and SCP(mTATA) promoters, respectively.

tural features of these two models are nearly the same at this resolution, indicating that TFIIB does not introduce global changes in the structure of TFIID-TFIIA-SCP upon binding.

Despite these overall similarities, there are subtle changes in the DNA density within the TFIID-TFIIA-TFIIB-SCP(-66) structure. First, serving as a positive control, this sample was prepared with SCP(-66) to test the presence and location of the upstream DNA density. Since the initial 3D model used for refinement did *not* contain density corresponding to the upstream DNA, the presence of additional density exiting lobe A in a near-identical location suggests that position of upstream DNA does not change relative to the TATA box within TFIID-TFIIA-TFIIB-SCP(-66) (Figure 4.11). Interestingly, however, the shape of the DNA density appears to be narrower than the previously determined structure. This may be due to the increased rigidity of upstream DNA due to the TFIIB-DNA contacts surrounding the TATA box (Figure 4.9A). Unfortunately, even though there were changes in the upstream DNA, we were unable to detect any differences within lobe A that was consistent in size with TFIIB (30 kDa). For clarity, the previously proposed atomic model of promoter DNA, TBP, TFIIA, and TFIIB were docked into the TFIID-TFIIA-TFIIB-SCP(-66) structure to show the proposed TFIIB binding site within lobe A. While TFIIB likely interacts with the rearranged conformation, the low resolution of the structure prevents further conclusions to be drawn regarding structural consequences of TFIIB binding.

4.5 A general model of regulated DNA binding by TFIID

The structural work presented within this chapter has served to test and extend a model of promoter binding by TFIID. In order to describe TFIID's interaction with promoter DNA, the model must take into account the interplay between TFIID's conformational dynamics and promoter binding. Therefore, the following models have been proposed in an attempt to summarize the structural data presented throughout Chapters 2, 3 & 4 (Figures 4.12 - 4.14). While the rearranged state serves as the predominant high affinity DNA binding conformation for TFIID, the structural data presented here address alternative structural states that TFIID can adopt during the process of DNA binding. The proposed models show the interplay between core promoter architecture and TFIIA, where the arrangements of core promoter motifs dictate the mode of DNA binding by TFIID: SCP (Figure 4.12), SCP(mTATA) (Figure 4.13), and SCP(mMTE/DPE) (Figure 4.14). Broadly, we believe these models serve as a framework for understanding TFIID's pleiotropic interactions with promoter DNA in an activator-dependent manner.

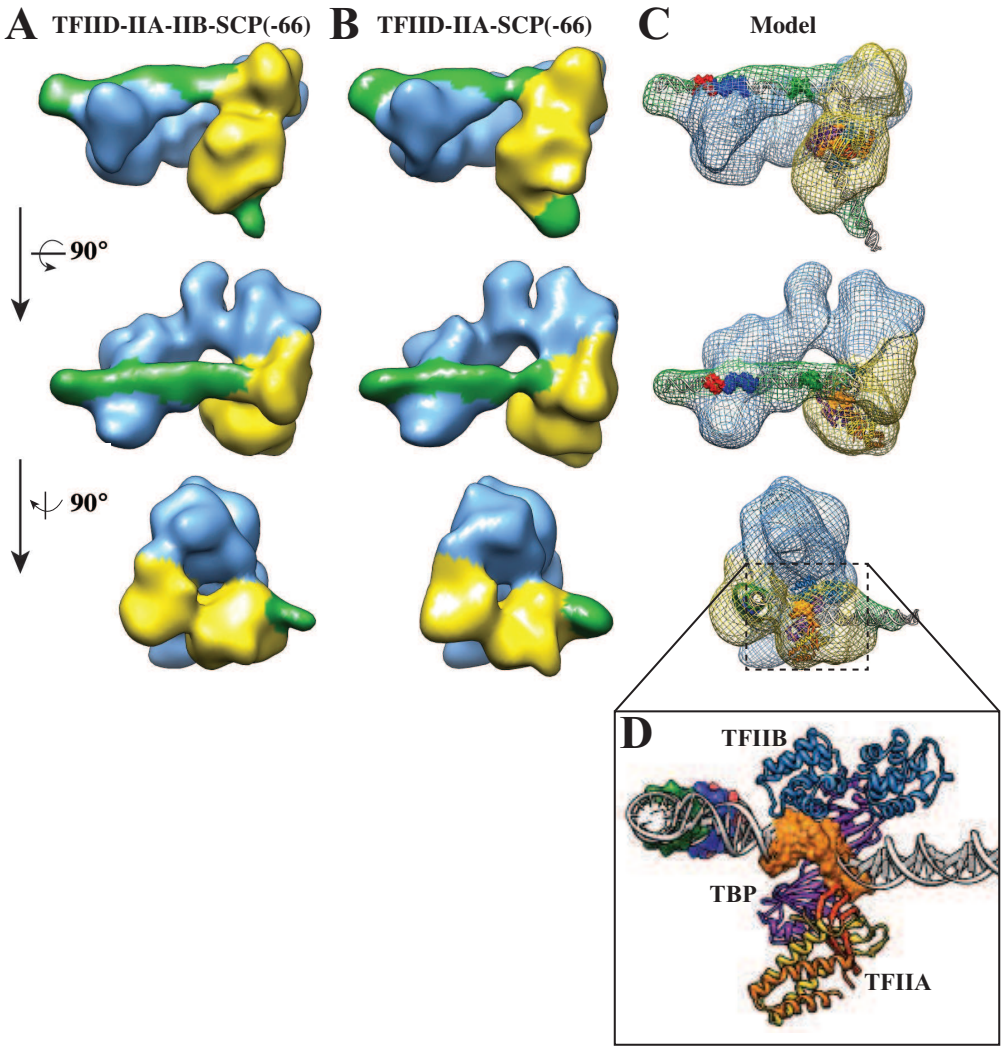


Figure 4.11: 3D reconstruction of the rearranged conformation of TFIID-TFIIA-TFIIB-SCP(-66) reveals similar topology as TFIID-TFIIA-SCP(-66). 3D models for rearranged conformation of TFIID-TFIIA-TFIIB-SCP(-66) (A) and TFIID-TFIIA-SCP (B) at 36Å. (C) Docked DNA model for -66 to +45 with TBP-TFIIB-TFIIA crystal structure model (D) from Figure 3.17.

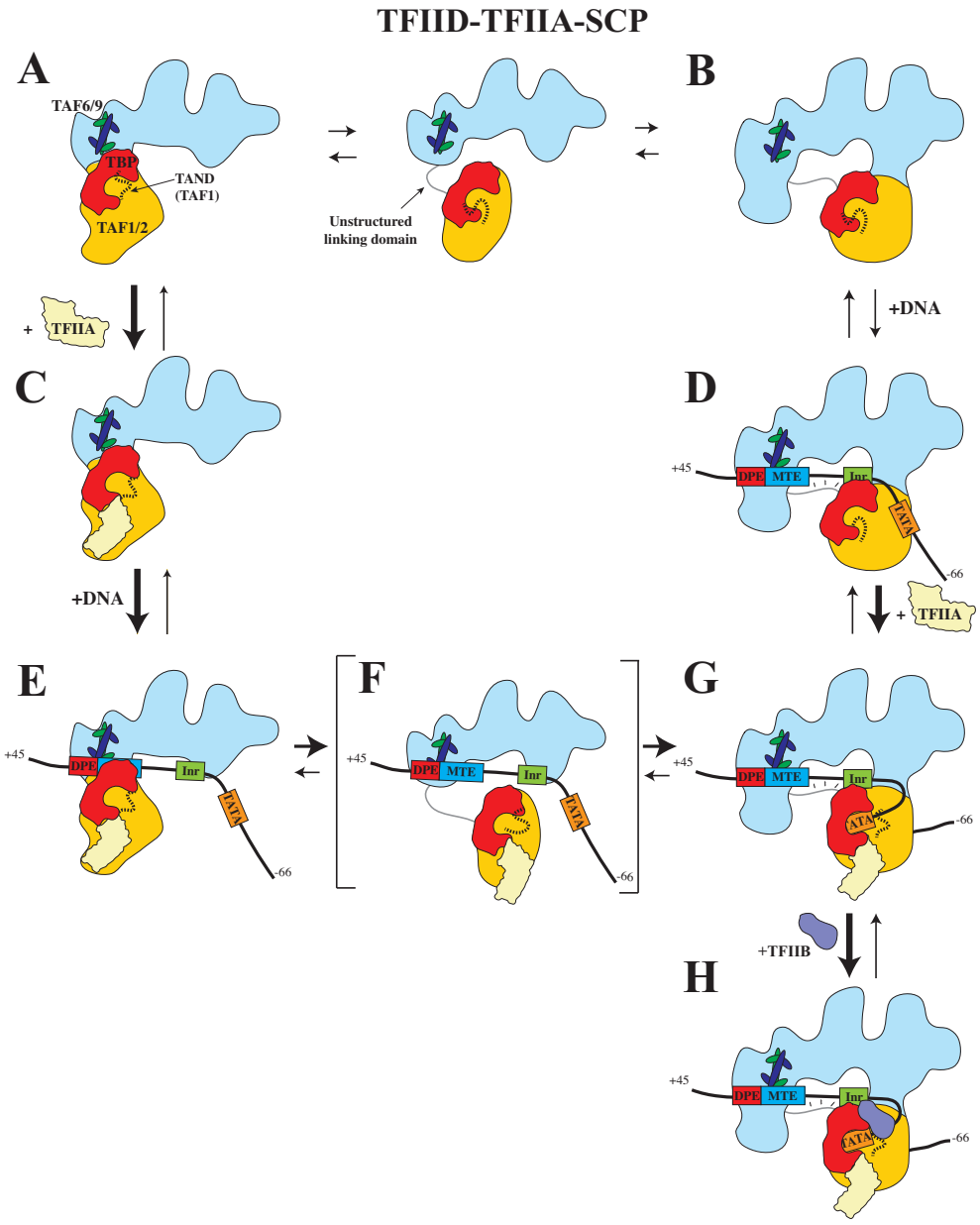


Figure 4.12: Model for SCP DNA binding by TFIID-TFIIA. TFIID undergoes conformational changes between canonical (A) and rearranged (B) states. The addition of TFIIA stabilizes the canonical conformation (C). Alternatively, the addition of SCP DNA leads to binding of the Inr-MTE/DPE to the rearranged conformation (D). The combined presence of TFIIA and SCP DNA leads to a stabilization of the rearranged conformation (G). While most particles adopt the rearranged state for TFIID-TFIIA-SCP, there is a small subset that is bound to SCP DNA and TFIIA within the canonical state (E). This state converts to the rearranged state through a likely intermediate state (F). The rearranged TFIID-TFIIA-SCP conformation can then be bound by TFIIB to load RNAPII.

TFIID-TFIIA-SCP

As previously suggested, TFIID's conformational state appears to be intimately connected to its promoter binding properties. Before binding DNA, however, TFIID already exhibits an unprecedented degree of conformational dynamics, where lobe A reorganizes between two distinct conformations: the canonical and rearranged states (Figures 4.12A & B). Measurements of lobe A position from both negative stain and cryo-EM sample preparations revealed that lobe A undergoes a structural transition between these two states. Surprisingly, these results indicated that approximately 50% of the TFIID molecules adopted the rearranged state before the addition of activators or promoter DNA (Figure 2.4). Therefore, within our model, lobe A coexists between the canonical (Figure 4.12A) and rearranged states (Figure 4.12B). In order to explain the lobe A's pendulum-like motion, we have modeled lobe A to contain an unstructured linker domain that remains stably attached to the BC core throughout the process of lobe A's reorganization.

From this existing equilibrium between the canonical and rearranged states, we propose that there are two alternative paths towards the high-affinity rearranged state bound to DNA. One path involves the binding of SCP DNA to the rearranged state of TFIID in the absence of TFIIA (Figure 4.12D). This was indicated by the identical footprinting patterns of the Inr-MTE/DPE within TFIID-SCP and TFIID-TFIIA-SCP. While the rearranged state is competent for DNA binding, the presence of SCP DNA alone is not sufficient to dramatically alter the conformational distribution of TFIID (Figure 2.6B). Therefore, while the conformational dynamics of lobe A remain unchanged, the presence of high affinity Inr and MTE/DPE motifs within the SCP allows TFIID to interact with the DNA within the rearranged conformation. Notably, TBP remains inhibited in the absence of TFIIA (Figure 4.1), which is likely due to the inhibitory N-terminal domain of TAF1 [2, 27, 66].

Given the ability of TFIIA to stimulate high affinity TFIID-DNA complexes, we hypothesized that TFIIA should stabilize TFIID within the rearranged state. Surprisingly, TFIIA stabilizes TFIID within the canonical state (Figure 2.6C). While we do not understand the functional role of this canonical state-stabilization, we propose that this conformation primes TFIID for TBP-TATA interactions through the release of the inhibitory N-terminal domain of TAF1 (Figure 4.12C).

These two parallel branches of TFIID's conformational dynamics converge when TFIID is in the presence of TFIIA and SCP DNA, where the rearranged conformation becomes preferentially stabilized. Given that a majority (60%) of the particles for TFIID-TFIIA-SCP adopted the rearranged state (Figure 2.6D) and were bound to promoter DNA (Figure 2.7A), we modeled the conformational landscape of TFIID-TFIIA-SCP to facilitate the formation of the rearranged state (Figure 4.12G). We believe that the transition from Figure 4.12D to Figure 4.12G involves the stabilization of TBP-TATA interactions through a TFIIA-mediated release of TBP inhibition. This is supported by the strong TATA box protection observed only within the ternary TFIID-TFIIA-SCP complex.

On the other hand, we believe that there are at least two structural transitions that are necessary to facilitate the reorganization from the canonical to rearranged state. Since

TFIIA localizes to lobe A (Figure 4.7A), we predicted that the TATA box would be bound by lobe A within the canonical state. This hypothesis was tested through Nanogold localization of SCP(+45 gold), where the SCP DNA was oriented in a manner that the MTE and DPE motifs were bound by lobe C (Figure 4.7B). This binding geometry was incompatible with TATA box binding by lobe A in the canonical state due to the rigidity of the DNA. This was shown through SCP(TATA gold) localization in a distribution of states *away* from lobe A in the canonical state (Figure 4.7C). Therefore, we have modeled the canonical state for TFIID-TFIIA-SCP to contain lobe C interacting with the MTE/DPE in a similar manner as in the rearranged state (Figure 4.12E). We believe that this conformation is primed for rearrangement due to the TFIIA-mediate release of TBP inhibition, allowing lobe A to bind the TATA box within the rearranged state. Given the flexible nature of lobe A's attachment to the BC core, the structural reorganization between the canonical and rearranged states likely involves an intermediate state, where DNA is bound to lobe C and TFIIA is within lobe A (Figure 4.12F). The culmination of these processes results in the net-stabilization of TFIID-TFIIA-SCP within the rearranged state (Figure 4.12G).

TFIID-TFIIA-SCP(mTATA)

The model presented for the conformational landscape of TFIID-TFIIA-SCP describes at least two structural pathways sampled by TFIID during the binding of SCP DNA. While these studies of the SCP have provided the framework for understanding the relationship between TFIID's conformational dynamics and promoter binding, the majority of promoters across the human genome likely contain Inr-MTE/DPE or TATA-Inr promoters [46]. To address TFIID's promoter selectively, we performed footprinting and cryo-EM visualization of TFIID bound to SCP(mTATA), an SCP construct lacking a functional TATA box. DNase I footprinting showed that TFIID retains its intrinsic affinity for Inr-MTE/DPE in the absence of TFIIA (Figure 4.4). Since the pattern of protection from the Inr to the MTE/DPE for TFIID-SCP(mTATA) is nearly identical to TFIID-SCP and TFIID-TFIIA-SCP, we can infer that TFIID binds the SCP(mTATA) sequence within the rearranged conformation (Figure 4.13C). As seen previously for wild-type SCP, the addition of TFIIA does not change Inr-MTE/DPE contacts. However, when the TATA box is mutated, strong DNA contacts along the TATA box are disrupted while retaining low levels of DNA protection within the mutant TATA box sequence (Figures 4.4 & 4.5). These footprinting data indicate that TFIID-TFIIA-SCP(mTATA) interacts with DNA through the rearranged conformation, but that there are minimal interactions between TFIID-TFIIA and the mutant TATA box sequence (Figure 4.13D).

We propose that the absence of a functional TATA box severely reduces the role of the canonical state for DNA binding (Figure 4.13). For wild-type SCP, we previously proposed that TFIIA binding to lobe A within the canonical state activates TBP for TATA box binding. Due to the lack of a functional TATA box and the minimal DNase I protection for TFIID-TFIIA-SCP(mTATA), we believe that TFIIA-mediated binding of the

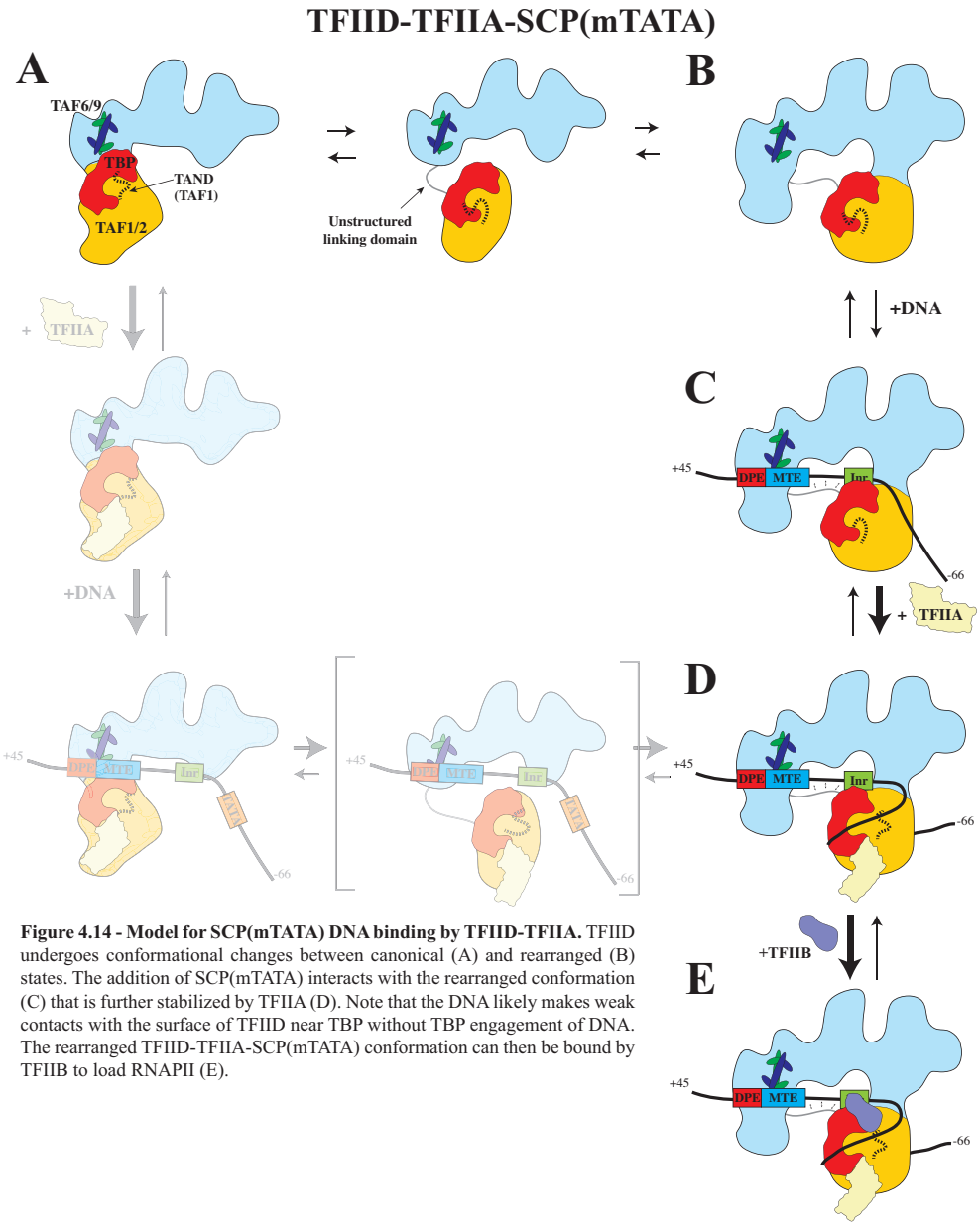


Figure 4.14 - Model for SCP(mTATA) DNA binding by TFIID-TFIIA. TFIID undergoes conformational changes between canonical (A) and rearranged (B) states. The addition of SCP(mTATA) interacts with the rearranged conformation (C) that is further stabilized by TFIIA (D). Note that the DNA likely makes weak contacts with the surface of TFIID near TBP without TBP engagement of DNA. The rearranged TFIID-TFIIA-SCP(mTATA) conformation can then be bound by TFIIB to load RNAPII (E).

Figure 4.13: Model for SCP(mTATA) DNA binding by TFIID-TFIIA. TFIID undergoes conformational changes between canonical (A) and rearranged (B) states. The addition of SCP(mTATA) interacts with the rearranged conformation (C) that is further stabilized by TFIIA (D). Note that the DNA likely makes weak contacts with the surface of TFIID near TBP without TBP engagement of DNA. The rearranged TFIID-TFIIA-SCP(mTATA) conformation should then be bound by TFIIB, although it is difficult to know the location of the TFIIB binding site (E).

SCP(mTATA) to lobe A does not facilitate high-affinity interactions in the same way that it stimulates TATA box binding for the wild-type SCP.

The well-characterized interactions between TFIIB, TBP, and DNA suggested that the addition of TFIIB may stimulate near wild-type binding of TFIID-TFIIA to the mutated TATA box sequence. While there may be TFIIB-TBP interactions present within TFIID-TFIIA-TFIIB-SCP(mTATA), there was only minimal protection of the mutant TATA box sequence, similar to that observed previously for TFIID-TFIIA-SCP(mTATA) (Figure 4.10). While the downstream interactions from the Inr to the MTE/DPE remain unchanged, the lack of any TFIIB-induced DNA protection near the TATA box suggests that either TFIIB is not bound to the TFIID-TFIIA-SCP(mTATA) complex or that TFIIB is anchored to the complex through protein-protein interactions with TBP. The extensive literature of TFIIB-TBP-TATA would suggest that these interactions would be maintained, even in the absence of consensus TATA box motifs [91]. Our results indicate TFIIB is not sufficient to induce DNase I protected sites within TFIID-TFIIA-TFIIB-SCP(mTATA), warranting further study of transcription initiation from non-TATA containing promoters.

TFIID-TFIIA-SCP(mMTE/DPE)

We further tested our model by studying the structural effects of a TATA-Inr promoter within the context of TFIID in the presence or absence of TFIIA. Surprisingly, our results showed that mutation of the MTE/DPE motifs abolished DNA binding of TFIID to promoter DNA for TFIID-SCP(mMTE/DPE) (Figure 4.4). This indicates that TFIID's interactions with promoter DNA through the rearranged conformation in the absence of TFIIA was dependent on functional MTE/DPE motifs. The addition of TFIIA to TFIID-SCP(mMTE/DPE) resulted in a strong protection of the TATA box in conjunction with near wild-type binding to the mutated MTE/DPE sequences (Figure 4.4). Considering that the protection patterns of TFIID-TFIIA-SCP(mMTE/DPE) strongly resembled wild-type interactions, we propose that TFIID-TFIIA-SCP(mMTE/DPE) adopts the rearranged conformation. However, unlike previous models, the footprinting data suggest that TFIIA-induced binding of TBP to the TATA box facilitates the reorganization of lobe A into the rearranged state. Therefore, we propose that TFIIA, bound to lobe A within the canonical state (Figure 4.14C), stimulates TATA box binding by TBP (Figure 4.14D) and thereby facilitates the movement of lobe A into the rearranged state (Figure 4.14E & F).

In addition to the results obtained from footprinting on SCP(mMTE/DPE), similar results were obtained for a mutant Inr (Figure 4.4), suggesting that TFIIA-mediated stimulation of TATA box binding is sufficient to drive rearrangement of lobe A in the absence of functional Inr or MTE/DPE motifs. We believe that these results are generally applicable for TFIID's interaction with low affinity promoter DNA sequences, where the binding of additional factors, anchoring TFIID to the promoter DNA, will stimulate DNA binding through an increase in TFIID's avidity for these low affinity sites. Given TFIID's ability to interact with a variety of upstream activators [29, 63], in addition to histone modifications

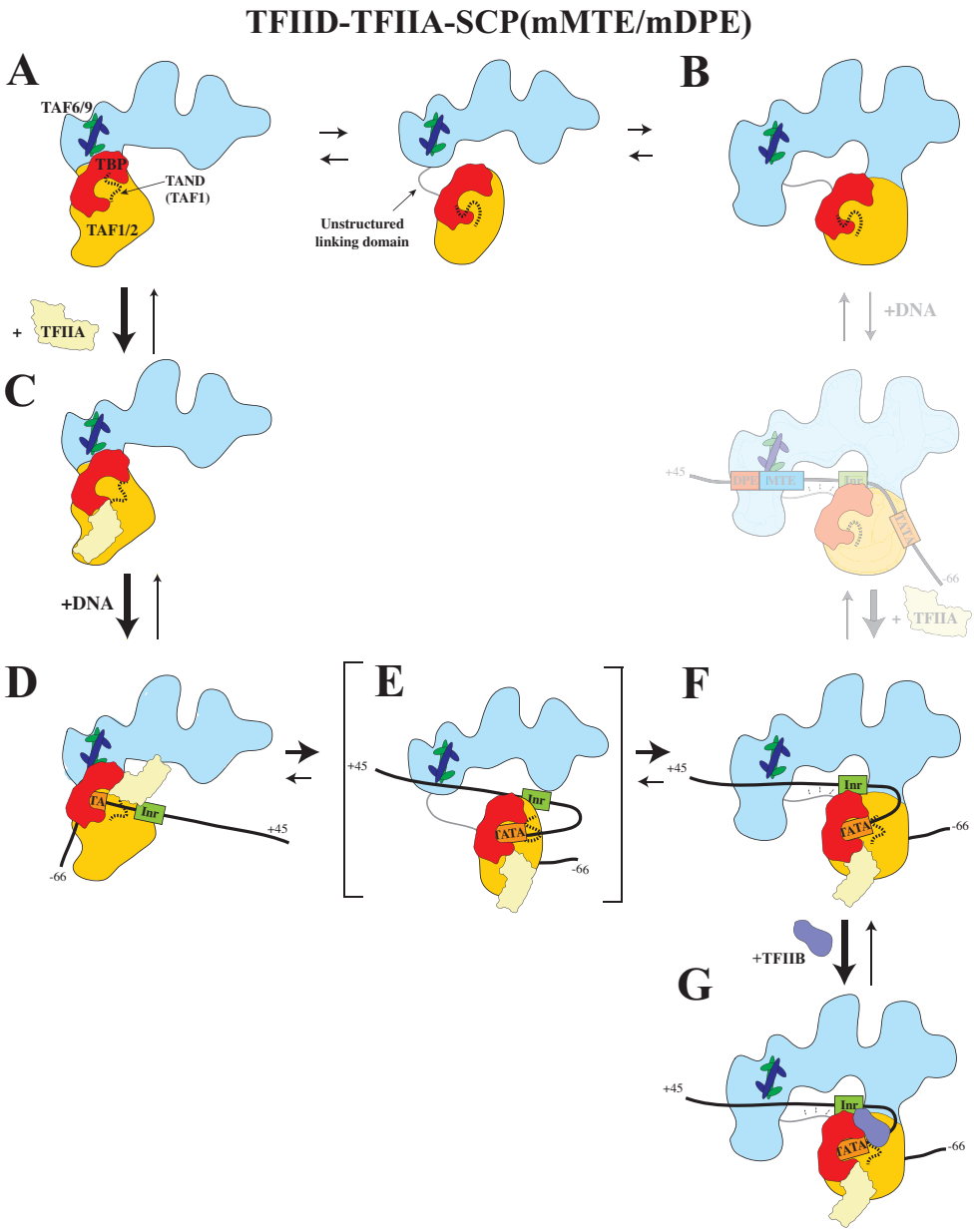


Figure 4.14: Model for SCP(mMTE/DPE) DNA binding by TFIID-TFIIA. TFIID undergoes conformational changes between canonical (A) and rearranged (B) states. The addition of TFIIA stabilizes the canonical conformation (C). The addition of SCP(mMPE/DPE) leads to the formation of the rearranged conformation (F) through the proposed intermediates of (D) and (E). The rearranged TFIID-TFIIA-SCP(mMTE/DPE) conformation can then be bound by TFIIB to load RNAPII (G).

(e.g. [42, 113]), we believe that the combined effect of TFIID's low affinity interactions with binding partners and DNA sequences is to stabilize TFIID bound to promoter DNA within the rearranged conformation.

Implications for RNAPII loading at the core promoter

Our studies suggest that TFIID introduces specific changes in the structure and topology of promoter DNA that may contribute to its role in transcription initiation. The cryo-EM structure of TFIID-TFIIA-SCP provides insight into a previously observed DNase I hypersensitive at +3, localizing this cleavage site to DNA spanning the central channel (Figure 4.3, denoted with *). This hypersensitive site has been observed in a wide range of core promoters upon binding of purified TFIID from Drosophila [13, 53, 65, 86, 89, 106] as well as human sources [20, 44, 47, 64, 78]. The strength of this hypersensitive site at +3 is positively correlated with the strength of *in vitro* transcription initiation, suggesting that the topological changes that TFIID induces at this site may be important for subsequent steps in transcription initiation.

Chapter 5

Conclusions and future outlook

A mechanistic description of metazoan transcription is essential for understanding the molecular processes that govern cellular decisions. To provide structural insights into the DNA recognition step of transcription initiation, we used single particle electron microscopy to visualize human TFIID with promoter DNA. This analysis revealed that TFIID co-exists in two predominant and distinct structural states that differ by a 100Å translocation of TFIID's lobe A. The transition between these structural states is modulated by TFIIA, as the presence of TFIIA and promoter DNA facilitates the formation of a novel rearranged state of TFIID capable of promoter recognition and binding. DNA-labeling and footprinting, together with cryo-EM studies, mapped the locations of the TATA, Inr, MTE, and DPE promoter motifs within the TFIID-TFIIA-DNA structure. The existence of two structurally and functionally distinct forms of TFIID suggests that the different conformers may serve as specific targets for the action of regulatory factors.

5.1 Discovery of the rearranged conformation

The flexibility of lobe A was previously overlooked because we assumed that TFIID would maintain a specific 3D shape. Since previous work in the lab had determined the three-dimensional structure of the canonical state [31, 67, 68], we attempted to characterize the binding of TFIID to SCP DNA. After 3 - 4 years of unsuccessful attempts to determine the structure of TFIID-TFIIA-SCP within the lab, we finally had the breakthrough discovery that lobe A repositions itself within the DNA bound state.

Since all of our previous negative stain and cryo-EM models were unable to describe the topology of the protein density seen within these DNA bound class averages, we attempted angular reconstitution of the cryo-EM 2D class averages. Working with Richard Hall, we calculated many different angular reconstitution models from the cryo-EM class averages of TFIID-TFIIA-SCP. These attempts were unsuccessful because we mistakenly tried to merge class averages from the canonical and rearranged state. Therefore, all of the resulting 3D models were unable to refine the cryo-EM data.

The key control experiment was the implementation of OTR for TFIID-TFIIA-SCP. However, the first time we performed OTR, we merged class volumes before 3D model refinement. It is likely that the merged model was the result of incorrect combination of class volumes, where either models from the rearranged were incorrectly aligned or models from the canonical and rearranged states were merged together. Therefore, when this incorrectly merged model was refined against the untilted particle data, the resulting 3D model was no longer in agreement with the 2D reference-free averages, suggesting that something was wrong in our analysis. When I returned to this OTR dataset 6 months later, individual class volumes were *not* merged together before 3D refinement. Subsequent inspection of the refined volumes revealed, much to our surprise, that lobe A adopted the rearranged state in almost all of the refined 3D structures.

5.2 The rearranged state as *the* DNA binding configuration of TFIID

In addition to mapping the conformational landscape of TFIID, this work shows that the rearranged conformation likely serves as the general mode of promoter binding by TFIID. Footprinting experiments of TFIID on promoters containing mutations in the TATA box, Inr, and MTE/DPE showed that TFIID retains the same overall footprint on DNA regardless of the underlying sequence. For instance, we observed that the mutation of the MTE/DPE resulted in the loss of promoter binding by TFIID in the absence of TFIIA. However, this loss was fully restored when TFIIA was added to the reaction. This suggested that TFIIA anchors TFIID to the TATA box, allowing TFIID to bind the low affinity downstream DNA sites in a near-identical manner. Overall, these data indicate that the rearranged state serves as TFIID's conformation capable of multi-valent interactions with promoter DNA.

EM structural studies have confirmed that TFIID binds promoters with varying architecture through the rearranged state. While we have not performed cryo-EM on all mutant promoters tested, cryo-EM of TFIID-TFIIA-SCP(mTATA) showed that it was bound through the rearranged conformation, similar in topology to TFIID-TFIIA bound to the wild-type SCP. Further support came from cryo-EM of an assembled TFIID-p53-TFIIA complex on native promoter DNA from the hdm2 gene. Even though this promoter only has a TATA box, preliminary cryo-EM visualization by Weili Liu (Albert Einstein College of Medicine) showed 2D reference-free class averages of TFIID bound to promoter DNA through the rearranged state. Considering the low affinity of TFIID alone for this hdm2 promoter, we propose that the role of activators (p53 and TFIIA) is to facilitate the formation of the rearranged state by anchoring TFIID to the promoter DNA.

Beyond the studies presented here, we believe that the characteristic footprinting pattern for human TFIID - observed for the past 30 years - corresponds to the rearranged state. Comparison of TFIID's footprinting patterns on the SCP(mMTE/DPE) and the previously

published AdML promoter using partially purified (e.g. [95, 111]) or fully purified TFIID (e.g. [13, 20, 78, 120] shows the overall similarity of the results. In all cases, there is a TFIIA-mediated protection of the TATA that is accompanied by downstream contacts at positions +1 and +30 relative to the TSS. Therefore, we propose that the previously observed 'isomerizations' and 'structural changes' correspond to the formation of the rearranged state of TFIID bound to promoter DNA.

Considering that the rearranged state appears to be the dominant DNA binding configuration of TFIID, the recognition of TFIID-TFIIA-SCP by TFIIB suggests that this state loads RNAPII at the TSS. While EM studies were able to unambiguously identify the binding site for TFIIB within the TFIID-TFIIA-TFIIB-SCP structure, biochemical footprinting showed that TFIIB makes intimate contacts with the flanking DNA sequences around the TATA box. Considering that the TATA box is located within lobe A, we believe that TFIIB is binding to lobe A. Furthermore, since the downstream contacts along the Inr and MTE/DPE were unchanged, TFIIB is binding to TFIID through the rearranged state. This indicates that, in addition to providing TFIID with high affinity DNA interactions, the rearranged state may serve as a structural signal to load RNAPII at the TSS.

5.3 Future directions

This thesis has laid the groundwork for mechanistic experiments to probe the contributions that TFIID makes during the formation of the pre-initiation complex. Additionally, for the first time, we have a detailed model of promoter binding and regulation of human TFIID activity that can be tested through a variety of experimental designs. The two most important research questions that remain will be obtaining a high resolution structure (10 - 15Å) of promoter-bound TFIID and determining the mechanism for RNAPII loading at the TSS.

High resolution studies of TFIID should be pursued through two orthogonal lines of research: 1) cryo-EM of a purified complex of TFIID-TFIIA-TFIIB-SCP and 2) biochemical isolation and EM analysis of the BC core. Given the ability of TFIIB to stabilize DNA contacts within TFIID-TFIIA-TFIIB-SCP, this complex should be co-immunoprecipitated with TFIID and pursued for high resolution studies. By visualizing this protein complex using high resolution optics (e.g. Titan EM microscopes by FEI), detectors (e.g. direct electron detectors), and phase-plate technologies, much higher levels of contrast should be possible for single particles of TFIID-TFIIA-TFIIB-SCP, allowing improved particle alignments. Especially considering the fact that we have these technologies installed on the shared Titan Microscope, a dedicated effort should be undertaken to get all of these devices working in concert to image single particles in vitreous ice.

Considering the difficulties associated with using delicate pieces of equipment, an orthogonal strategy should be taken in order to increase the resolution of the BC core. Since the BC core appears to be the most well-ordered protein density within TFIID, biochemical experiments should be performed to try and isolate this subcomplex away from the holo-TFIID

complex. A purified BC core will be a strong candidate for high resolution studies given its unique shape and apparent rigidity. To obtain such a sample, limited proteolysis should be performed with a wide-range of proteases to identify the proper conditions for specific cleavage of lobe A away from the BC core. It should be noted that previous work has shown that TAF1 (a proposed component of lobe A) appears to be sensitive to limited proteolysis, suggesting that it may be possible to separate lobe A from the BC core [81].

Alternatively, recent work from Imre Berger's lab (EMBL, Grenoble) has purified and begun structural characterization of a subcomplex of human TFIID that comprises TAF6, -9, -4, -5, and -12. This work is currently unpublished, but after publication, the constructs should be requested and we should begin purification of these sub-complexes of TFIID. While the EM structure of this subcomplex does not appear (at first glance) to be related to the BC core, additional biochemical work and structural to define the relationship between the BC core and the reconstituted subcomplex should yield important information regarding the organization of subunits within TFIID.

In addition to these high resolution structural studies of TFIID, mechanistic studies should be pursued with TFIID-TFIIA-TFIIB-SCP and RNAPII-TFIIF. Based upon the ability of Yuan He to purify and structurally characterize intermediates along the assembly pathway for the formation of a minimal pre-initiation complex nucleated by TBP, similar strategies should be taken to purified a complex of TFIID-TFIIA-TFIIB-SCP-RNAPII-TFIIF. While this is a non-trivial purification, it has the potential to offer unprecedented insight into the loading of RNAPII at the promoter. This will require extensive biochemical screening in addition to advanced image sorting methodologies, considering that there may be sample heterogeneity due to the presence of RNAPII.

Appendix A

Antibody labeling of TFIID

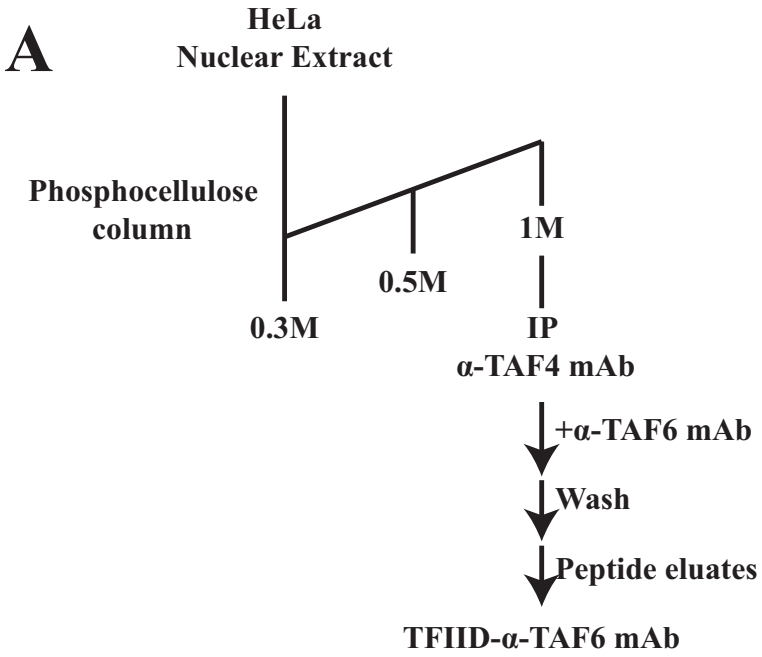
Given the lack of information regarding the subunit organization within TFIID, there has been a concerted effort by a number of laboratory members to generate and analyze TFIID-antibody datasets. Most of these datasets were prepared and analyzed by Wei-Li Liu (Assistant Prof., Albert Einstein College of Medicine) and Patricia Grob (staff scientist). Antibodies directed against TBP, TAF4, and TAF6 were bound to TFIID during the purification procedure, allowing the elution of stoichiometric TFIID-antibody complexes. Unfortunatley, all previous attempts to localize these antibodies were unsuccessful.

After describing the extent of lobe A's flexibility within the context of TFIID alone, we hypothesized that the flexibility of lobe A hampered previously analysis of antibody-bound TFIID particles. Therefore, TFIID-antibody datasets for α-TBP (unpublished), α-TAF4 [67], α-TAF6 (unpublished) were reanalyzed after classification of lobe A into discrete conformational sub-states. While we were unable to recover any density within 2D class averages or 3D models for α-TBP and α-TAF4 antibodies (data not show), we were able to localize significant antibody density for α-TAF6 bound to the BC core. This localization supports the model of DNA-bound TFIID in addition to proposing a pseudo-symmetry within the BC core that necessitates further study.

A.1 Purification of a ternary TFIID-α-TAF6 complex

As previously stated, the purification and analysis of TFIID-α-TAF6 was performed previously by Wei-Li Liu and Patricia Grob, respectively. The general strategy for antibody labeling of TFIID was to purify TFIID from Hela nuclear extract using established protocols, but, before elution of TFIID from the agarose beads, labeling antibodies were incubated with TFIID. After extensive washing, these TFIID-antibody complexes were eluted from the beads and analyzed using SDS-PAGE to verify antibody binding.

Using this strategy, Wei-Li Liu purified a stoichiometric TFIID-α-TAF6 complex with a monoclonal antibody directed against TAF6 (Figure A.1A). Analysis of this labeling through SDS-PAGE revealed strong binding by the α-TAF6 antibody that could be co-immunpurified



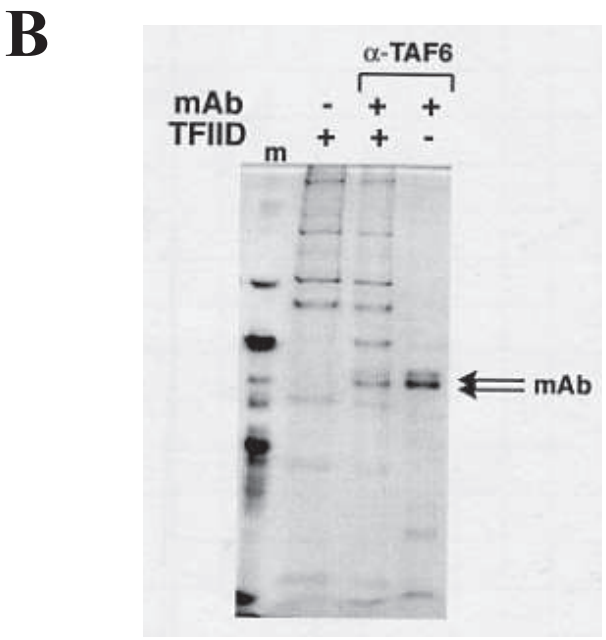


Figure A.1: Purification strategy for TFIID-α-TAF6. (A) Strategy for purification of a ternary TFIID-α-TAF6 complex. (B) SDS-PAGE of TFIID-α-TAF6. Note that this was photo-copied from Patrica Grob's laboratory notebook.

with TFIID (Figure A.1B).

A.2 α-TAF6 antibody localization

Cryo-EM analysis showed that lobe C interacts with the MTE and DPE, suggesting that TAF6 is found within lobe C since TAF6 cross-links to both MTE and DPE sequences [14, 65]. Therefore, we expected to find the TAF6 antibody bound to lobe C. But, considering the large degree of conformational flexiblity exhibited by TFIID alone (Figure 2.6A), focused classification of lobe A was first performed before any attempt at antibody localization. To this end, the purified TFIID-α-TAF6 complex was prepared for negative stain single particle EM by Patricia Grob. Particles were picked, extracted, and phase-flipped prior to storage of the dataset within the archives of the Nogales lab. These particles were subjected to careful 2D reference-free image analysis and processed as previously described (Figures 2.5). Briefly, extracted particles corresponding to characteristic view were re-aligned to the rigid BC core, where the overall average of the re-aligned particles showed the characteristic flexibility of lobe A (Figure A.2A). After focused classification of lobe A and measurement of lobe A's position relative to the BC core, the histogram of lobe A positions indicated that the canonical and rearranged conformations were equally populated (Figure A.2B), as seen previously for TFIID alone (Figure 2.6A).

Since the position of lobe A within the canonical state is bound to lobe C in close proximity to the DNA binding site, we hypothesized that the epitope recognized by α-TAF6 may be occluded within the canonical conformation. Therefore, particles were separated into groups corresponding to the canonical and rearranged conformations before further analysis. After systemtically moving a circular mask around lobe C and the BC core, we found strong antibody density only when the mask was positioned around lobe C (Figure A.2C). This density was found in two different sub-classified averages, providing evidence that TAF6 is localized to lobe C.

During sub-classification of the canonical state, we noticed additional density that appeared bound to lobe B. Further analysis confirmed that, within the canonical state, lobe B interacts with the TAF6 antibody (Figure A.2D). This result is surprising considering that lobe C is the MTE/DPE interacting domain within TFIID. However, considering that a number of subunits within TFIID may exist as multiple copies, we believe that this localization may provide insight into the apparanet pseudo-symmetry of the BC core.

A.3 Comparison of antibody labeling between yeast and human TFIID

The data presented within this Appendix indicate that the α-TAF6 antibody localizes to lobes B and C within the canonical and rearranged states, respectively. These α-TAF6 antibody densities are the strongest example of antibody localization within human TFIID to

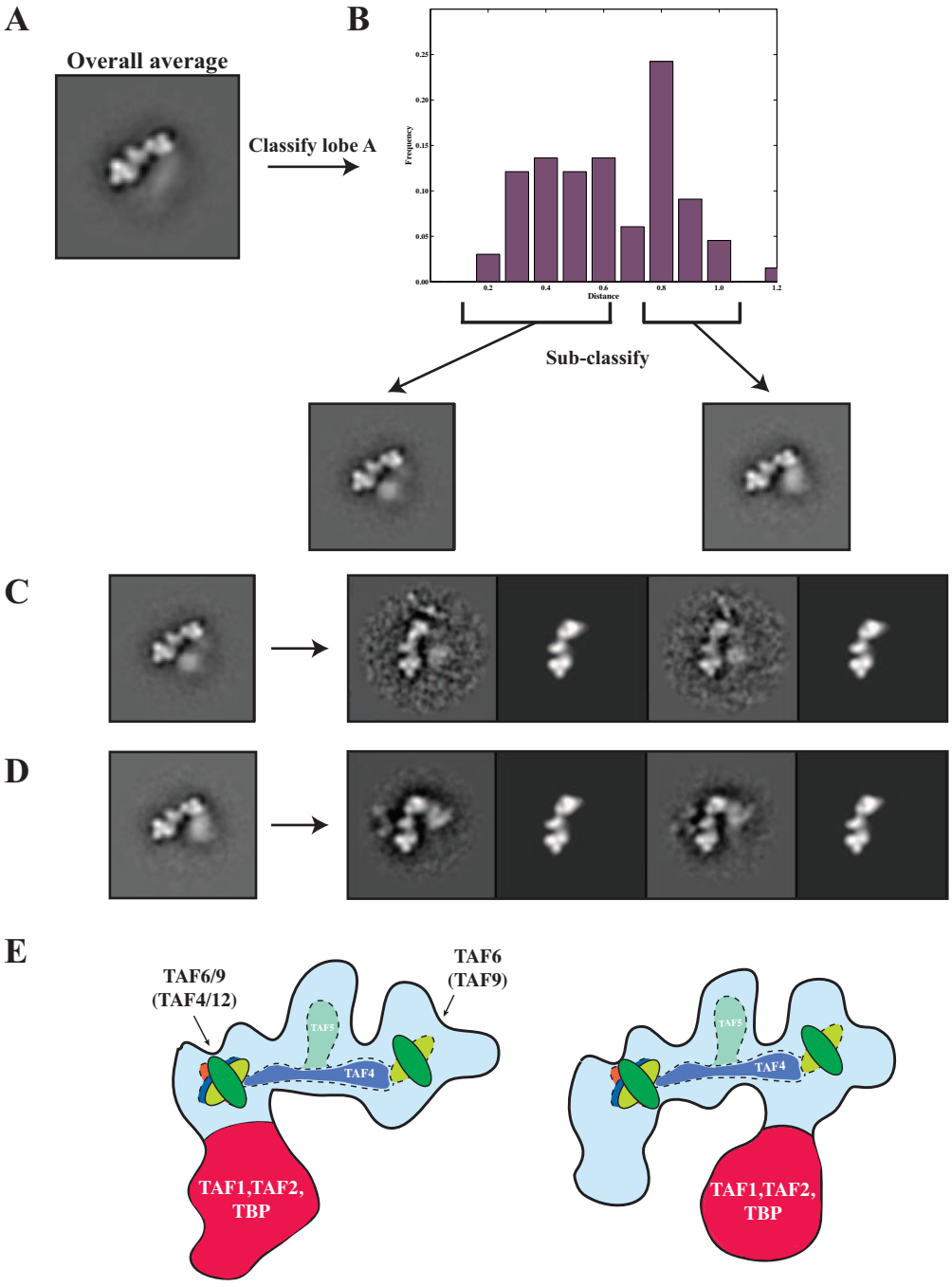


Figure A.2: α-TAF6 localizes to lobes B and C within the BC core. 2D reference-free analysis was performed previously described in Figure 2.6. (A) Overall average of re-aligned particles within characteristic view. (B) Histogram of lobe A positions relative to the BC core. (C) Sub-classification of the rearranged conformation. Note additional density bound to lobe C. Averages were aligned to projections of the rigid BC core. (D) Sub-classification of the canonical conformation. Note additional density bound to lobe B. (E) Model of TFIID's subunits. Tentativity locations of TAF4, -5, -9, and -12 are shown as dotted lines.

date. The antibody localization was successful only after extensive 2D reference-free classification and subclassification of TFIID within distinct structural states. We believe that this strategy will be generally applicable for localization of additional antibodies and other TFIID-interacting factors.

Previous research on yeast TFIID has proposed the existance of TAFs existing in multiple copies within TFIID [62, 61]. Indeed, mass spectrometry has supported this proposal, providing biochemical evidence for this model [94]. Combining the biochemical characterization with antibody labeling of TFIID, a number of papers have proposed yeast TFIID to exist as a nearly-symmetric molecule [62, 82]. However, considering that the antibody-labeling of yeast TFIID employed the same strategies as previously employed on human BTAF1 and TFIIH, we propose that this error-prone strategy, combined with conformational flexibility, led to an over-symmetrized model of TFIID subunit organization.

The same research group that performed antibody labeling of yeast TFIID also utilized a similar stategy on human BTAF1 [87] and human TFIIH [98], where subsequent studies of these protein complexes revealed that these subunit locations were inaccurate. For instance, the BTAF1 labeling proposed that TBP localizes to the extension protruding away from the core of the molecule [87], however, x-ray crystallographic studies of yeast Mot1-TBP revealed that TBP is bound within the core domain [87]. This suggests that, within the defined molecular system of BTAF1 (two subunits), the authors inaccurately localized TBP, the key component within BTAF1.

To tackle an even more challenging system, the authors performed antibody localization for human TFIIH, a multi-subunit GTF responsible for promoter opening by RNAPII [98]. It should be stated that this was a challenging system and it provided the first 'glimpse' into TFIIH architecture, however, all antibody labeling performed ([98]: Figure 5) was proven incorrect through structural analysis of an assembled pre-initiation complex containing RNAPII bound to promoter DNA and TFIIH (Yuan He, Nogales laboratory, unpublished). For TFIIH, the authors did not realize that the kinase domain was extremeley flexibile, which made it nearly invisible for a number of different class averages shown ([98]: Figure 5A - F). Furthermore, when the kinase domain was visible, it was assigned to antibody density ([98], Figure 5G & H). As previously stated, this was a challenging sample, however, it serves as another instance of incorrect antibody localization by this research group.

Given these considerations, we hesitate to extend the yeast TFIID antibody labeling studies into our model of human TFIID. While the authors are not intentionally misrepresenting their data for antibody labeling of yeast TFIID, we believe that their stategy has a false-positive rate of 40 - 50% based upon their past work [87, 98]. We believe that this false-positive rate is the result of two error-prone methods employed by this research group: 1) failure to purify ternary antibody-protein complexes, and 2) 'single-particle' counting of antibodies. For human BTAF1 [87], human TFIIH [98], and yeast TFIID [62, 61], the authors did not purify ternary complexes of antibody-protein complexes, instead mixing the antibody in 5 - 10 molar excesses with the protein prior to grid preparation. We believe that this method will provide false positives given the high molar excess of antibodies. While non-specific antibody-protein complexes should be averaged out when 2D reference-free class

averages are calculated, the authors do not calculate averages of their labeled complexes. Instead, the authors perform 'single-particle' counting of antibody-protein complexes, where individual particle images are inspected for potential antibody density and, after investigating 600 - 1000 individual particles, locations are proposed. Therefore, by adding the antibodies in excess and relying on localization within low SNR images of single particles, the authors inherently bias their analysis towards a high false-positive rate.

Antibody labeling is a notoriously difficult technique for subunit localization. The above papers represent hontest attempts to identify subunit locations within biologically important protein complexes. However, for reasons listed above, it is difficult to assess the quality of subunit location assignments. Indeed, antibody localization is extremeley difficult for human TFIID as well. The localization presented here for TAF6 was the only success out of three different datsets. While the localization of TAF6 may not be entirely accurate, it serves as a starting point for future studies, where orthogonal approaches should be used to verify antibody labeled locations. For example, considering that lobe C binds the MTE/DPE , a region of DNA known to interact with TAF6 through biochemical studies, while also showing density for the α-TAF6 antibody, we have a high degree of confidence that TAF6 is localized within lobe C.

A.4 Implications of a pseudo-symmetric BC core

While we need to continue to verify the location of TAF6 within lobe B using alternative strategies, it is tempting to speculate on this pseduo-symmetry within the BC core. As previously noted, studies of yeast TFIID proposed the existance of a nearly-symmetric TFIID molecule, due to the presence of antibody density on both lobes. We believe that this symmetric model of yeast TFIID is the result of mislocalization of antibodies within yeast TFIID. Lobe A within yesat TFIID may exhibit the same large degree of flexibility as observed for human TFIID, necessitating that all analysis of yeast TFIID labeled with antibodies be obtained from class averages.

While the symmetric model of yeast TFIID may not be accurate, the labeling of TAF6 to both lobes B an d C suggests that the BC core may exist as a pseudo-symmetric backbone of TFIID. Initial structural inspection of the BC core revealed that it almost appeared as a C2 symmetrical structure, with the symmetry axis located along the ridge connecting lobes B and C. This observation is not sufficient for concluding a pseudo-symmetric BC core, but the dual labeling of the TAF6 antibody to both lobes B and C provides biochemical evidence for such a model. This observation of a dual-antibody labeled TFIID molecules was only apparent *after* calculating 2D averages of human TFIID and performed careful sub-classification, a detail that would be overlooked when investigating single particles for yeast TFIID.

The presence of TAF6 within both lobes B and C may provide a binding site for lobe A within the canonical and rearranged states. Since the discovery of lobe A's dramatic flexibilty, we posited that TFIID may contain similar binding sites for lobe A within both

canonical and rearranged conformations. Therefore, after localizing TAF6 to both lobes B and C, we propose that TAF6 may serve as the common binding site for lobe A within the canonical (lobe C) and rearranged (lobe B) conformations. While this hypothesis remains untested, it serves as a reasonable molecular explanation for lobe A's ability to form the canonical and rearranged states.

The pseudo-symmetry of the BC has been incorporated into a model describing the organization of human TFIID (Figure A.2E). TAF6 occupies a position within both lobes B and C, where it is likely bound to TAF9 in both cases through its highly conserved histone-fold domain [99]. From biochemical evidence, we believe that TAF6 and TAF9 further interact with TAF4 and TAF12 within lobe C to form a histone-like tetramer, an interaction shown to stabilize DPE-interactions within a TAF6/9/4/12 complex [99]. We propose that lobe C is connected to the BC core through an extension of TAF4 from the histone-tetramer. This is consistent with TAF4 serving a scaffolding role within *Drosophila* TFIID, where RNAi knockdown of TAF4 results in the loss of remaining subunits within holo-TFIID [118]. Serving to facilitate the connections between lobes C and B, we believe that TAF5's dimerization domain may be localized within the ridge connecting lobes C and B. While this model remains to be tested, we believe that it serves as a valid initial description of TFIID's organization, providing a molecular basis for the pseudo-symmetry of the BC core and dual docking site for lobe A in both states.

Appendix B

Materials & Methods

B.1 Protein preparation

Purification of human TFIID and TFIID-TFIIA-SCP was performed as described previously [31, 68]. Briefly, nuclear extract was fractionated with phosphocellulose P11 (P-Cell) resins [1, 75]. P-Cell column fractions eluting at 1 M KCl/HEMG buffer (pH 7.9, [20 mM Hepes, 0.2 mM EDTA, 2 mM $MgCl_2$, 10% glycerol] plus 1 mM DTT and 0.5 mM PMSF) were pooled, dialyzed to 0.3 M KCl/HEMG buffer (containing 0.1% NP40 and 10 μM leupeptin) and immunoprecipitated overnight at 4°C with an anti-TAF4 mAb covalently conjugated to protein G beads (GE Healthcare). TAF4-immunoprecipitates were extensively washed with 0.65 M KCl/HEMG buffer, 0.3 M KCl/HEMG buffer and 0.1M KCl/HEMG buffer (containing 0.1% NP40 buffer and 10 μM leupeptin) prior to the addition of human recombinant TFIIA and SCP DNA. Purified recombinant TFIIA (see below) and SCP DNA (IDT) were incubated (2 hrs) with 2.5 μg of the washed TFIID complex bound to 250 μl of protein G sepharose beads containing covalently conjugated TAF4 mAb. Assembled TFIID-TFIIA-SCP complexes were washed 5 times with 0.1 M KCl/HEMG buffer (containing 0.1% NP40) and were eluted with a peptide (1 mg/ml in 0.1 M KCl/HEMG buffer/0.1% NP40) recognized by the TAF4 mAb. For purified TFIID used in Nanogold-labeling experiments, the final elution step included 1 mM TCEP instead of 1 mM DTT. The eluates were concentrated with a microcon-10 concentrator, immediately applied to grids or flash frozen, and negatively stained as described below. Assembly of the TFIID-TFIIA-SCP complex involved adding a 10X-excess of recombinant TFIIA [104] and SCP DNA (IDT) prior to elution from the resin.

Recombinant human TFIIA was purified as described previously [67, 75, 104]. TFIIA-α/β and TFIIA-γ constructs were co-transformed into BL21 Star competent *E. coli* (Invitrogen) and were grown until $OD_{600} = 0.6$ 0.8. After induction with 0.4M IPTG, the following temperature and times were used: TFIIA-α/β was induced at 25°C for 5 hours and TFIIA-γ was induced at 37°C for 1 hour. Cells for each construct were pelleted and washed twice in 1X PBS with 0.1 M PMSF prior to flash freezing in liquid nitrogen. Pellets were resuspended

in lysis buffer (50 mM Tris-CL pH = 7.9, 10% glycerol, 20 mM beta-mercaptoethanol, 1 mM PMSF, 100 mM NaCl) and sonicated with 5 x 10 second prior to centrifugation to reduce viscosity. The pellets were washed once in lysis buffer before they were extracted with lysis buffer (6 M GuHCl) for 1 hour at room temperature. Ni^{2+}-NTA resin that was previously equilibrated in lysis buffer (6 M GuHCl) was added to the resuspended pellets and incubated overnight at 4°C. The resin was washed using a gravity flow column at 4°C using lysis buffer (6M GuHCl) prior to elution with elution buffer (lysis buffer + 6 M GuHCl + 300 mM imidazole). The purity of the purified and denatured TFIIA-α/β and TFIIA-γ were analyzed by SDS-PAGE. TFIIA-α/β and TFIIA-γ were renatured together by dialysis. First, they were dialyzed for 4 hrs against elution buffer (HEMG: 20 mM HEPES pH = 8, 10% glycerol, 2 mM $MgCl_2$, 100 mM KCl, 0.1 mM PMSF, 20 mM beta-mercaptoethanol) with the addition of 2 M GuHCl. Using a peristaltic pump, TFIIA was renatured through the slow addition (1.3 mL/min) of elution buffer lacking GuHCl over the course of 8 hours. Refolded TFIIA was then dialyzed against HEMG (5 mM beta-mercaptoethanol), purified over a Poros20 HQ column (Applied Biosystems), and eluted using a linear gradient of 0.1 M to 1 M KCl over 10 column volumes.

For Nanogold labeling reactions, TFIIA was labeled with 1.8 nm Ni^{2+}-NTA-Nanogold (Nanoprobes Inc.) using prescribed protocols. Briefly, TFIIA was incubated with a 10-fold molar excess of Ni^{2+}-NTA-Nanogold in a final volume of 54 μl for 20 minutes at 4°C. Prior to sample loading, a S200 size exclusion column (GE Healthcare) on an Ettan Analytical LC system was equilibrated with HEMG (1 mM TCEP). After comparison of A_{420}/A_{280} signals, there was 50% labeling efficiency for TFIIA. Given the labile nature of a Nanogold, Nanogold-TFIIA was used immediately for cryo-EM grid preparation.

B.2 Nucleic acid preparation

The DNA sequence used for the SCP was taken directly from the originally published sequence of SCP1 [44]: GTACTTATATAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGCCGAGCAGACGTGCCTACGGACCG. For SCP(-66), the following sequence was added immediately upstream of the SCP sequence: CTCGCGCCACCTCTGTTTTCCCAGTCACGA. Mutant promoter sequences were taken directly from previous mutational analysis of SCP1 [44].

All DNA oligonucleotides were ordered through Integrated DNA Technologies (IDT) and were annealed in 10 mM Tris-HCl pH = 8.0. For 1.4 nm-monomaleiomido-Nanogold (Nanoprobes Inc.) labeling experiments, SCP oligonucleotides were ordered with a 5'- or 3'-six carbon linked thiol functional group. Prior to incubating DNA with Nanogold, DNA was incubated in buffer containing 1 mM TCEP for 15 minutes at room temperature to ensure that the thiol functional groups were fully reduced. The DNA was added to a 50-fold molar excess of Nanogold that was previously solubilized in 0.2 mL distilled water and incubated for 2 hours at room temperature. Prior to purification, the sample was concentrated using

a 10 MWCO spin-column concentrator (Sartorius) from 200 μl to approximately 30 μl. The concentrated sample was added to a S200 size exclusion column (GE Healthcare) that was previously equilibrated in running buffer (0.02 M sodium phosphate, 150 mM NaCl pH = 7.4). After purification of a labeled DNA-Nanogold complex, comparison of A_{420}/A_{260} signal indicated that the sample was labeled with a 70% labeling efficiency. Before the Nanogold-DNA sample could be added to TFIID, the sample was concentrated 5-fold and then used immediately for cryo-EM grid preparation.

B.3 EM sample preparation

For negative stain microscopy, four microliters of TFIID sample was directly applied to glow discharged holey carbon film covered with a continuous thin-carbon support on a 400 mesh copper grid (Electron Microscopy Sciences). For cryo-EM, samples were vitrified using a Vitrobot (Gatan, Inc.) that was set to 100% humidity at 4°C. Four microliters of sample was incubated for 30 sec. - 1 min. on carbon-thickened C-flat grid (Protochips) with 4 μm holes that were spaced 2 μm apart that had a thin-carbon support within the holes, blotted for 6.5 s, and then plunge-frozen in liquid ethane.

Negative stain data for TFIID alone, TFIID-TFIIA, TFIID-SCP, TFIID-TFIIA-SCP, TFIID-p53, TFIID-sp1, TFIID-c-Jun, and TFIID-α-TAF6 were collected using a Tecnai T12 bio-TWIN transmission EM operating at 120 keV under low dose (15 $e^-/Å^2$) conditions on Kodak SO163 film at a nominal magnification of 30,000X from defocuses ranging from -0.70 μm to -1.20 μm. Micrographs were digitized in a Nikon Super Coolscan 8000 at a pixel size of 12.7 μm (4.23 Å/pix). Cryo-EM data were collected on a Tecnai F20 TWIN transmission EM operating at 120 keV using a dose of 25 $e^-/Å^2$ on a Gatan 4k x 4k CCD at a nominal magnification of 80,000X (1.501 Å/pixel). Leginon software was used to automatically focus and collect exposure images [103].

B.4 Single particle image analysis

Negative stain data were manually picked using Boxer (EMAN) [69], CTF-estimated using CTFFIND3 [73], and phase-flipped using SPIDER [26]. Cryo-EM data were prepared using the Appion image processing environment [55] where particles were automatically picked using Signature [19], CTF-estimated using CTFFIND3 [73], phase flipped using SPIDER [26], and normalized using XMIPP [101].

For negative stain and cryo-EM data, 2D reference-free image analysis was performed using an iterative routine implementing a topology representing network [79] in combination with mutli-reference-alignment within IMAGIC [34].

3D refinements were performed on phase-flipped particles using an iterative projection matching and 3D reconstruction approach that was performed using libraries from EMAN2

and SPARX software packages [38, 105]. 34,167 particles from cryo-EM grids prepared from TFIID-TFIIA-SCP sample were refined against low-pass filtered models of the canonical and rearranged state. For TFIID, 30,800 particles were collected and refined. For SCP(mTATA), TFIID was incubated with 10X excess of SCP(mTATA) and TFIIA, under identical conditions as the TFIID-TFIIA-SCP sample preparation, where 27,500 particles were collected and analyzed. Following this global projection matching routine, all data were further refined in FREALIGN [30] and low-pass filtered at the final resolution while applying a negative B-factor using *bfactor.exe*.

Orthogonal tilt reconstruction of TFIID-IIA-SCP

The substantial structural differences between the canonical and rearranged states of TFIID rendered the previously reported 3D reconstruction of the canonical state human TFIID unable to be used as a reference for projection matching and 3D reconstruction. We thus generated an *ab initio* model for the rearranged state by implementing the orthogonal tilt reconstruction (OTR) technique [59] on negatively stained samples of TFIID-TFIIA-SCP.

Tilt pairs at +/- 45° were collected on a TFIID-IIA-SCP sample prepared on glow discharged holey carbon film [7] on a 400 mesh copper grid (Electron Microscopy Sciences) with a thin continuous carbon layer using a Tecnai F20 TWIN transmission electron microscope operating at 120 keV using a dose of 25 $e^-/Å^2$ on a Gatan 4k x 4k CCD at a nominal magnification of 80,000X (1.501 Å/pixel). Particles were manually picked from the tilt pairs using *xmipp mark* within XMIPP [101]. The XMIPP coordinates (.pos) were converted to BOXER coordinates (.box) prior to particle extraction using batchboxer within EMAN [69]. The tilt angles for each micrograph were calculated using CTFTILT [73]. Extracted particles from a given tilt were subjected to 2D reference-free classification and alignment using multivariate-statistical-analysis and multi-reference-refinement within IMAGIC [34]. To calculate individual class volumes, euler angles were calculated using the equations previously described [58], 3D reconstructions were performed using 'bp 3f' within SPIDER [26], and the volumes were filtered based upon FSC = 0.5.

Since OTR preserves three-dimensional features of individual class volumes without suffering from the 'missing-cone' problem of random conical tilt [18], individual class volumes were refined against untilted negative stain data for TFIID-IIA-SCP. These refined models were then quantitatively compared to the characteristic class averages of TFIID in six characteristic conformations and models were excluded based upon Fourier Ring Correlation (FRC) [96] values below 60 Å. A majority (8/10) of the refined, validated models were found to be in the rearranged state, whereas a minority (2/10) were in the canonical conformation.

Focused classification of lobe A

The overall data processing strategy for assessing lobe A's conformational heterogeneity is mapped out in Figure 2.5. Briefly, negatively stained particles for each dataset - TFIID (16,000 particles), TFIID-SCP (9,859 particles), TFIID-IIA-SCP (19,678 particles), and TFIID-IIA (20,152 particles) were aligned and classified in a reference-free fashion using a topology representing network [79] and multi-reference-alignment within IMAGIC [34], where class average reference images were aligned to each other and mirrors generated after each iteration using the prep-mra-refs command within IMAGIC. After 8-9 iterations, all particles corresponding to the characteristic view of TFIID were extracted, mirrored if necessary, and placed into a single stack of particles. These particles were then aligned to a masked reference of the BC core using multi-reference-alignment within IMAGIC [34]. Multivariate-statistical-analysis was performed within a mask drawn around lobe A. The resulting eigen-images reflected the conformational flexibility of lobe A and were used to perform hierarchical ascendant classification. Subsequently generated classums (10-20 particles/class) showed a range of positions for lobe A relative to the rigid BC core.

To measure lobe A's position, individual classums were measured using BOXER within EMAN [69] and plotted within MATLAB. Boxes were placed, in order, on lobes B, C, and A. The coordinates were used to calculate a projection of lobe A's position on the B-C axis that were normalized to the length of the BC core. A histogram of lobe A positions were calculated within MATLAB and the distribution was fitted using a maximum-likelihood estimate of a mixed Gaussian distribution ('gmdistribution.fit') and was plotted using 'pdf.' Wilcoxon rank sum tests were performed within MATLAB ('ranksum') to test for statistical differences between distributions of lobe A positions from the different datasets.

Cryo-EM of Nanogold complexes & image analysis

For localization of Nanogold labels within cryo-EM images of TFIID-IIA(gold)-SCP (10,075 particles), TFIID-IIA-SCP(Nanogold at TATA) (851 particles), and TFIID-IIA-SCP(Nanogold at +45) (6,502 particles), focal pair images were collected using Leginon [103] at defocuses of -3 μm and -0.5μm on a Tecnai F20 TWIN transmission electron microscope operating at 120 keV using a dose of 25 $e^-/Å^2$ on a Gatan 4k x 4k CCD at a nominal magnification of 80,000X (1.501 Å/pixel). Particles were manually picked using BOXER, shifts between focal pairs were calculated using alignhuge and applied to the particle coordinate files prior to particle extraction with batchboxer within EMAN [69]. The high defocus particles were then normalized and dusted using *xmipp normalize* within XMIPP [101] to remove gold signal $> 3.5\sigma$. These dusted, high defocus particles were subjected to reference-free 2D classification using iterative classification and alignment within IMAGIC with multivariate statistical analysis and multi-reference-alignment [34]. The rotation and shifts applied to the high-defocus particles were applied to the following stacks of particles using 'equivalent-rotation' within IMAGIC: 1) un-dusted high defocus particles termed 'high defocus particles'

that were normalized using 'edgenorm' within the *proc2d* command of EMAN; 2) thresholded high defocus particles where pixels $> 4\sigma$ were converted to 1 and all pixels below 4σ were converted to 0 for each un-dusted high defocus particle; 3) un-dusted low defocus particles termed 'low defocus particles' that were normalized using 'edgenorm' within the *proc2d* command of EMAN; and 4) thresholded low defocus particles where pixels $> 4\sigma$ were converted to 1 and all pixels below 4σ were converted to 0 for each un-dusted low defocus particle. By averaging low defocus particles and low defocus thresholded particles according to the 2D reference-free alignment of dusted high defocus particles, Nanogold density could be localized with high confidence given the high contrast of Nanogold at low defocus values.

Image simulations of gold nanoclusters

Gold simulations utilized a crystalline gold model that was face-centered cubic with a lattice parameter a = 4.08Å. Due to the high contrast of the gold crystal, the model was oriented away from the zone axis. The gold nanocluster was imbedded into a GROMACS box of water molecules and exported into .xyz format for simulations within multislice [51]. After modeling Fresnel propagation through this sample, a variety of defocuses were simulated ranging from 0.0 to 5.0 mum using a 0.15 mum step size. For the 'noise added' simulations, Poisson noise was added to simulate an electron dose of 25 $e^-/Å^2$. Peak intensity was calculated using a Gaussian fitting routine written by Colin Ophus (LBL, NCEM) within MATLAB [70]. The peak height was plotting as: Peak height = -(Gold intensity)/(Standard deviation of the image), since the gold particles were black on a white background.

The image processing package SPIDER [26] was used to create synthetic images to simulate top hat and hard-edge functions (Figure 3.11). For the top hat function (Figure 3.11B), a circle was created using the model image command ('MO') with a diameter of 9 pixels (14Å). The image contrast was inverted so that the pixels within the circle had a value of -1 while all other values were 0. The model image was then multiplied in Fourier space by simulated CTF functions ('TF CT') over a range of defocuses. The image was then normalized so that the average pixel intensity was 0. From this normalized image, line profiles were calculated through the central pixel ('LI D') and plotted within MATLAB. For the hard-edge function (Figure 3.11C), the same approach was utilized except using a circle of much larger radius (100 pixels using a box size of 256 pixels). Subsequent line profiles were shifted to align the edges from the hard-edge and the simulated gold images.

B.5 DNase I and MPE-Fe footprinting

Extended-length M13 sequencing primers (28 nt) with 5'-ends corresponding to basepair -109 and +110 (relative to the start site of SCP as +1) were purified by denaturing gel electrophoresis, 5'-end-labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase, desalted on a 900 μl Sephadex G50 fine column, extracted with phenol/$CHCl_3$/isoamyl alcohol (IAA),

precipitated with ethanol, and the pellet was suspended in 10 mM Tris-Cl pH 8.0, 0.1 mM EDTA. DNA probes with either a 5'-upstream or a 5'-downstream labeled end were generated by PCR using the appropriate combination of labeled and unlabeled primers with pUC119-SCP1 [44] as template. The 220 basepair product was isolated by native gel electrophoresis, passively eluted into 10 mM Tris-Cl pH 8.0, 0.1 mM EDTA, 0.2% SDS, 1M LiCl followed by phenol/$CHCl_3$/IAA extraction, ethanol precipitation, and resuspension in 10 mM Tris-Cl pH 8.0, 0.1 mM EDTA. A+G, T, and A chemical sequencing ladders used for mapping cleavage sites followed published methods [40, 93].

Protein-DNA complexes were formed for 20 min. at 30°C in 20 μl of binding buffer containing 20 mM KHEPES pH 7.8, 4 mM $MgCl_2$, 0.2 mM EDTA, 0.05% (v/v) NP-40, 8% (v/v) glycerol, 100 μg/mL BSA, and 1 mM (DNase I footprinting) or 2 mM DTT (MPE-Fe footprinting) upon final assembly containing 2 nM TFIID, 20 nM TFIIA, and 0.2 nM (DNase I footprinting) or 0.75 nM (MPE-Fe footprinting) DNA probe. Proteins were diluted in a diluent consisting of binding buffer with 20% glycerol and 200 μg/mL BSA. DNase I digestion was initiated by the addition of 2.2 μl of 0.38 - 0.75 mU/μl DNase I (in diluent containing 5 mM $CaCl_2$) and terminated 30 seconds later by the addition of 158 μl of a stop solution (10 mM Tris-Cl pH 8.0, 3 mM EDTA, 0.2% SDS). MPE-Fe(II) was generated by combining equal volumes of 1 mM methidiumpropyl EDTA (Sigma, discontinued) and 1 mM $NH_4Fe(II)SO_4$ (Aldrich) for 5 min. followed by dilution in water. MPE-Fe(II) cleavage was initiated by the sequential addition of 1.2 μl 23 mM NaAscorbate and 1.2 μl 46 μM MPE-Fe(II) and terminated 2 min. later by the sequential addition of 2 μl 120 μM bathophenanthroline and 156 μl stop solution. Samples were processed by phenol/$CHCl_3$/IAA extraction and ethanol precipitation. Digestion products were resolved on 10% polyacrylamide (37.5:1) containing 8.3 M urea until xylene cyanol migrated 80% down the gel. The phosphor image (Typhoon Trio; GE Healthcare) was analyzed using Image Gage (Fuji Film).

Bibliography

[1] 3rd Andel F. et al. "Three-dimensional structure of the human TFIID-IIA-IIB complex". In: *Science* 286.5447 (1999), pp. 2153–6.

[2] S. Bagby et al. "TFIIA-TAF regulatory interplay: NMR evidence for overlapping binding sites on TBP". In: *FEBS Lett* 468.2-3 (2000), pp. 149–54.

[3] L. A. Baker et al. "Arrangement of subunits in intact mammalian mitochondrial ATP synthase determined by cryo-EM". In: *Proc Natl Acad Sci U S A* 109.29 (2012), pp. 11675–80.

[4] S. Bhattacharya, S. Takada, and R. H. Jacobson. "Structural analysis and dimerization potential of the human TAF5 subunit of TFIID". In: *Proc Natl Acad Sci U S A* 104.4 (2007), pp. 1189–94.

[5] M. Bleichenbacher, S. Tan, and T. J. Richmond. "Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes". In: *J Mol Biol* 332.4 (2003), pp. 783–93.

[6] N. Boisset et al. "Three-dimensional reconstruction of a complex of human alpha 2-macroglobulin with monomaleimido Nanogold (Au1.4nm) embedded in ice". In: *J Struct Biol* 109.1 (1992), pp. 39–45.

[7] D.E Bradley. "The preparation of specimen support films." In: (1965).

[8] M. Brand et al. "Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC". In: *Science* 286.5447 (1999), pp. 2151–3.

[9] A. T. Brunger. "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures". In: *Nature* 355.6359 (1992), pp. 472–5.

[10] L. Bumba et al. "Localization of the PsbH subunit in photosystem II from the Synechocystis 6803 using the His-tagged Ni-NTA Nanogold labeling". In: *J Struct Biol* 152.1 (2005), pp. 28–35.

[11] S. Buratowski et al. "Five intermediate complexes in transcription initiation by RNA polymerase II". In: *Cell* 56.4 (1989), pp. 549–61.

[12] S. Buratowski et al. "Function of a yeast TATA element-binding protein in a mammalian transcription system". In: *Nature* 334.6177 (1988), pp. 37–42.

[13] T. W. Burke and J. T. Kadonaga. "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters". In: *Genes Dev* 10.6 (1996), pp. 711–24.

[14] T. W. Burke and J. T. Kadonaga. "The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila". In: *Genes Dev* 11.22 (1997), pp. 3020–31.

[15] S. K. Burley and R. G. Roeder. "Biochemistry and structural biology of transcription factor IID (TFIID)". In: *Annu Rev Biochem* 65 (1996), pp. 769–99.

[16] M. G. Campbell et al. "Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy". In: *Structure* (2012).

[17] G. E. Chalkley and C. P. Verrijzer. "DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator". In: *EMBO J* 18.17 (1999), pp. 4835–45.

[18] P. Chandramouli et al. "Validation of the orthogonal tilt reconstruction method with a biological test sample". In: *J Struct Biol* 175.1 (2011), pp. 85–96.

[19] J. Z. Chen and N. Grigorieff. "SIGNATURE: a single-particle selection system for molecular electron microscopy". In: *J Struct Biol* 157.1 (2007), pp. 168–73.

[20] T. Chi and M. Carey. "Assembly of the isomerized TFIIA–TFIID–TATA ternary complex is necessary and sufficient for gene activation". In: *Genes Dev* 10.20 (1996), pp. 2540–50.

[21] W. Deng and S. G. Roberts. "A core promoter element downstream of the TATA box that is recognized by TFIIB". In: *Genes Dev* 19.20 (2005), pp. 2418–23.

[22] D. J. DeRosier and A. Klug. "Reconstruction of Three Dimensional Structures from Electron Micrographs". In: *Nature* 217 (1968), pp. 130–3.

[23] B. D. Dynlacht, T. Hoey, and R. Tjian. "Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation". In: *Cell* 66.3 (1991), pp. 563–76.

[24] H. Elmlund et al. "Cryo-EM reveals promoter DNA binding and conformational flexibility of the general transcription factor TFIID". In: *Structure* 17.11 (2009), pp. 1442–52.

[25] J. Frank. *Three-dimensional electron microscopy of macromolecular assemblies*. Academic Press, San Diego, CA, 1996.

[26] J. Frank et al. "SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields". In: *J Struct Biol* 116.1 (1996), pp. 190–9.

[27] J. H. Geiger et al. "Crystal structure of the yeast TFIIA/TBP/DNA complex". In: *Science* 272.5263 (1996), pp. 830–6.

[28] P. Goodman and A. F. Moodie. "Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method". In: *Acta Crystallographica Section A* 30.2 (1974), pp. 280–290.

[29] J. A. Goodrich and R. Tjian. "Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation". In: *Nat Rev Genet* 11.8 (2010), pp. 549–58.

[30] N. Grigorieff. "FREALIGN: high-resolution refinement of single particle structures". In: *J Struct Biol* 157.1 (2007), pp. 117–25.

[31] P. Grob et al. "Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues". In: *Structure* 14.3 (2006), pp. 511–20.

[32] J. F. Hainfeld and R. D. Powell. "New frontiers in gold labeling". In: *J Histochem Cytochem* 48.4 (2000), pp. 471–80.

[33] R. J. Hall, E. Nogales, and R. M. Glaeser. "Accurate modeling of single-particle cryo-EM images quantitates the benefits expected from using Zernike phase contrast". In: *J Struct Biol* 174.3 (2011), pp. 468–75.

[34] M. van Heel et al. "A new generation of the IMAGIC image processing system". In: *J Struct Biol* 116.1 (1996), pp. 17–24.

[35] R. Henderson et al. "Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy". In: *J Mol Biol* 413.5 (2011).

[36] R. P. Hertzberg and P. B. Dervan. "Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses". In: *Biochemistry* 23.17 (1984), pp. 3934–45.

[37] A. Hoffmann et al. "A histone octamer-like structure within TFIID". In: *Nature* 380.6572 (1996), pp. 356–9.

[38] M. Hohn et al. "SPARX, a new environment for Cryo-EM image processing". In: *J Struct Biol* 157.1 (2007), pp. 47–55.

[39] J. Y. Hsu et al. "TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription". In: *Genes Dev* 22.17 (2008), pp. 2353–8.

[40] B. L. Iverson and P. B. Dervan. "Adenine specific DNA chemical sequencing reaction". In: *Nucleic Acids Res* 15.19 (1987), pp. 7823–30.

[41] R. H. Jacobson et al. "Structure and function of a human TAFII250 double bromodomain module". In: *Science* 288.5470 (2000), pp. 1422–5.

[42] R. H. Jacobson et al. "Structure and function of a human TAFII250 double bromodomain module". In: *Science* 288.5470 (2000), pp. 1422–5.

[43] P. D. Jadzinsky et al. "Structure of a thiol monolayer-protected gold nanoparticle at 1.1 A resolution". In: *Science* 318.5849 (2007), pp. 430–3.

[44] T. Juven-Gershon, S. Cheng, and J. T. Kadonaga. "Rational design of a super core promoter that enhances gene expression". In: *Nat Methods* 3.11 (2006), pp. 917–22.

[45] T. Juven-Gershon, J. Y. Hsu, and J. T. Kadonaga. "Caudal, a key developmental regulator, is a DPE-specific transcriptional factor". In: *Genes Dev* 22.20 (2008), pp. 2823–30.

[46] T. Juven-Gershon and J. T. Kadonaga. "Regulation of gene expression via the core promoter and the basal transcriptional machinery". In: *Dev Biol* 339.2 (2010), pp. 225–9.

[47] J. Kaufmann and S. T. Smale. "Direct recognition of initiator elements by a component of the transcription factor IID complex". In: *Genes Dev* 8.7 (1994), pp. 821–9.

[48] J. L. Kim, D. B. Nikolov, and S. K. Burley. "Co-crystal structure of TBP recognizing the minor groove of a TATA element". In: *Nature* 365.6446 (1993), pp. 520–7.

[49] T. H. Kim et al. "A high-resolution map of active promoters in the human genome". In: *Nature* 436.7052 (2005), pp. 876–80.

[50] Y. Kim et al. "Crystal structure of a yeast TBP/TATA-box complex". In: *Nature* 365.6446 (1993), pp. 512–20.

[51] E.J. Kirkland. *Advanced Computing in Electron Microscopy*. Plenum Press, New York, NY, 1998.

[52] D. Kostrewa et al. "RNA polymerase II-TFIIB structure and mechanism of transcription initiation". In: *Nature* 462.7271 (2009), pp. 323–30.

[53] A. K. Kutach and J. T. Kadonaga. "The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters". In: *Mol Cell Biol* 20.13 (2000), pp. 4754–64.

[54] T. Lagrange et al. "New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB". In: *Genes Dev* 12.1 (1998), pp. 34–44.

[55] G. C. Lander et al. "Appion: an integrated, database-driven pipeline to facilitate EM image processing". In: *J Struct Biol* 166.1 (2009), pp. 95–102.

[56] G. C. Lander et al. "Complete subunit architecture of the proteasome regulatory particle". In: *Nature* 482.7384 (2012), pp. 186–91.

[57] W. C. Lau and J. L. Rubinstein. "Structure of intact Thermus thermophilus V-ATPase by cryo-EM reveals organization of the membrane-bound V(O) motor". In: *Proc Natl Acad Sci U S A* 107.4 (2010), pp. 1367–72.

[58] A. E. Leschziner. "The orthogonal tilt reconstruction method". In: *Methods Enzymol* 482 (2010), pp. 237–62.

[59] A. E. Leschziner and E. Nogales. "The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles". In: *J Struct Biol* 153.3 (2006), pp. 284–99.

[60] A. E. Leschziner and E. Nogales. "Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions". In: *Annu Rev Biophys Biomol Struct* 36 (2007), pp. 43–62.

[61] C. Leurent et al. "Mapping histone fold TAFs within yeast TFIID". In: *EMBO J* 21.13 (2002), pp. 3424–33.

[62] C. Leurent et al. "Mapping key functional sites within yeast TFIID". In: *EMBO J* 23.4 (2004), pp. 719–27.

[63] M. Levine and R. Tjian. "Transcription regulation and animal diversity". In: *Nature* 424.6945 (2003), pp. 147–51.

[64] P. M. Lieberman, J. Ozer, and D. B. Gursel. "Requirement for transcription factor IIA (TFIIA)-TFIID recruitment by an activator depends on promoter structure and template competition". In: *Mol Cell Biol* 17.11 (1997), pp. 6624–32.

[65] C. Y. Lim et al. "The MTE, a new core promoter element for transcription by RNA polymerase II". In: *Genes Dev* 18.13 (2004), pp. 1606–17.

[66] D. Liu 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". In: *Cell* 94.5 (1998), pp. 573–83.

[67] W. L. Liu et al. "Structural changes in TAF4b-TFIID correlate with promoter selectivity". In: *Mol Cell* 29.1 (2008), pp. 81–91.

[68] W. L. Liu et al. "Structures of three distinct activator-TFIID complexes". In: *Genes Dev* 23.13 (2009), pp. 1510–21.

[69] S. J. Ludtke, P. R. Baldwin, and W. Chiu. "EMAN: semiautomated software for high-resolution single-particle reconstructions". In: *J Struct Biol* 128.1 (1999), pp. 82–97.

[70] MathWorks. "MATLAB". In: (2011).

[71] T. Matsui et al. "Multiple factors required for accurate initiation of transcription by purified RNA polymerase II". In: *J Biol Chem* 255.24 (1980), pp. 11992–6.

[72] A. C. Milazzo et al. "Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy". In: *J Struct Biol* 176.3 (2011), pp. 404–8.

[73] J. A. Mindell and N. Grigorieff. "Accurate determination of local defocus and specimen tilt in electron microscopy". In: *J Struct Biol* 142.3 (2003), pp. 334–47.

[74] L. Montesano-Roditis et al. "Cryo-electron microscopic localization of protein L7/L12 within the Escherichia coli 70 S ribosome by difference mapping and Nanogold labeling". In: *J Biol Chem* 276.17 (2001), pp. 14117–23.

[75] A. M. Naar et al. "Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro". In: *Genes Dev* 12.19 (1998), pp. 3020–31.

[76] K. Nagayama. "Another 60 years in electron microscopy: development of phase-plate electron microscopy and biological applications". In: *J Electron Microsc (Tokyo)* 60 Suppl 1 (2011), S43–62.

[77] D. B. Nikolov et al. "Crystal structure of a TFIIB-TBP-TATA-element ternary complex". In: *Nature* 377.6545 (1995), pp. 119–28.

[78] T. Oelgeschlager, C. M. Chiang, and R. G. Roeder. "Topology and reorganization of a human TFIID-promoter complex". In: *Nature* 382.6593 (1996), pp. 735–8.

[79] T. Ogura, K. Iwasaki, and C. Sato. "Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking". In: *J Struct Biol* 143.3 (2003), pp. 185–200.

[80] S. Ohtsuki, M. Levine, and H. N. Cai. "Different core promoters possess distinct regulatory activities in the Drosophila embryo". In: *Genes Dev* 12.4 (1998), pp. 547–56.

[81] J. Ozer et al. "Transcription factor IIA derepresses TATA-binding protein (TBP)-associated factor inhibition of TBP-DNA binding". In: *J Biol Chem* 273.23 (1998), pp. 14293–300.

[82] G. Papai, P. A. Weil, and P. Schultz. "New insights into the function of transcription factor TFIID from recent structural studies". In: *Curr Opin Genet Dev* 21.2 (2011), pp. 219–24.

[83] G. Papai et al. "Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID". In: *Structure* 17.3 (2009), pp. 363–73.

[84] G. Papai et al. "TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation". In: *Nature* 465.7300 (2010), pp. 956–60.

[85] A. G. Papavassiliou. "Chemical nucleases as probes for studying DNA-protein interactions". In: *Biochem J* 305 (Pt 2) (1995), pp. 345–57.

[86] T. J. Parry et al. "The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery". In: *Genes Dev* 24.18 (2010), pp. 2013–8.

[87] L. A. Pereira et al. "Molecular architecture of the basal transcription factor B-TFIID". In: *J Biol Chem* 279.21 (2004), pp. 21802–7.

[88] B. F. Pugh and R. Tjian. "Mechanism of transcriptional activation by Sp1: evidence for coactivators". In: *Cell* 61.7 (1990), pp. 1187–97.

[89] B. A. Purnell, P. A. Emanuel, and D. S. Gilmour. "TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes". In: *Genes Dev* 8.7 (1994), pp. 830–42.

[90] M. Radermacher et al. "Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli". In: *J Microsc* 146.Pt 2 (1987), pp. 113–36.

[91] H. S. Rhee and B. F. Pugh. "Genome-wide structure and organization of eukaryotic pre-initiation complexes". In: *Nature* 483.7389 (2012), pp. 295–301.

[92] P. B. Rosenthal and R. Henderson. "Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy". In: *J Mol Biol* 333.4 (2003), pp. 721–45.

[93] Fritsch E.F Maniatis T. Sambrook J. "Molecular cloning: a laboratory manual". In: 3 (1989).

[94] S. L. Sanders, K. A. Garbett, and P. A. Weil. "Molecular characterization of Saccharomyces cerevisiae TFIID". In: *Mol Cell Biol* 22.16 (2002), pp. 6000–13.

[95] M. Sawadogo and R. G. Roeder. "Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region". In: *Cell* 43.1 (1985), pp. 165–75.

[96] W. O. Saxton and W. Baumeister. "The correlation averaging of a regularly arranged bacterial cell envelope protein". In: *J Microsc* 127.Pt 2 (1982), pp. 127–38.

[97] S. H. Scheres et al. "Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization". In: *Structure* 17.12 (2009), pp. 1563–72.

[98] P. Schultz et al. "Molecular structure of human TFIIH". In: *Cell* 102.5 (2000), pp. 599–607.

[99] H. Shao et al. "Core promoter binding by histone-like TAF complexes". In: *Mol Cell Biol* 25.1 (2005), pp. 206–19.

[100] S. T. Smale and D. Baltimore. "The "initiator" as a transcription control element". In: *Cell* 57.1 (1989), pp. 103–13.

[101] C. O. Sorzano et al. "XMIPP: a new generation of an open-source image processing package for electron microscopy". In: *J Struct Biol* 148.2 (2004), pp. 194–204.

[102] A. Stewart and N. Grigorieff. "Noise bias in the refinement of structures derived from single particles". In: *Ultramicroscopy* 102.1 (2004), pp. 67–84.

[103] C. Suloway et al. "Automated molecular microscopy: the new Leginon system". In: *J Struct Biol* 151.1 (2005), pp. 41–60.

[104] X. Sun et al. "Reconstitution of human TFIIA activity from recombinant polypeptides: a role in TFIID-mediated transcription". In: *Genes Dev* 8.19 (1994), pp. 2336–48.

[105] G. Tang et al. "EMAN2: an extensible image processing suite for electron microscopy". In: *J Struct Biol* 157.1 (2007), pp. 38–46.

[106] J. W. Theisen, C. Y. Lim, and J. T. Kadonaga. "Three key subregions contribute to the function of the downstream RNA polymerase II core promoter". In: *Mol Cell Biol* 30.14 (2010), pp. 3471–9.

[107] M. C. Thomas and C. M. Chiang. "The general transcription machinery and general cofactors". In: *Crit Rev Biochem Mol Biol* 41.3 (2006), pp. 105–78.

[108] F. T. Tsai and P. B. Sigler. "Structural basis of preinitiation complex assembly on human pol II promoters". In: *EMBO J* 19.1 (2000), pp. 25–36.

[109] K. Uchida et al. "High resolution footprinting of EcoRI and distamycin with Rh(phi)2(bpy)3+, a new photofootprinting reagent". In: *Nucleic Acids Res* 17.24 (1989), pp. 10259–79.

[110] M. W. Van Dyke, R. P. Hertzberg, and P. B. Dervan. "Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA: DNA cleavage-inhibition patterns with methidiumpropyl-EDTA.Fe(II)". In: *Proc Natl Acad Sci U S A* 79.18 (1982), pp. 5470–4.

[111] M. W. Van Dyke, R. G. Roeder, and M. Sawadogo. "Physical analysis of transcription preinitiation complex assembly on a class II gene promoter". In: *Science* 241.4871 (1988), pp. 1335–8.

[112] M. Van Heel. "Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction". In: *Ultramicroscopy* 21.2 (1987), pp. 111–23.

[113] M. Vermeulen et al. "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4". In: *Cell* 131.1 (2007), pp. 58–69.

[114] C. P. Verrijzer et al. "Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II". In: *Cell* 81.7 (1995), pp. 1115–25.

[115] J. D. Watson and F. H. Crick. "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid". In: *Nature* 171.4356 (1953), pp. 737–8.

[116] S. Werten 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)". In: *J Biol Chem* 277.47 (2002), pp. 45502–9.

[117] E. de Wit and W. de Laat. "A decade of 3C technologies: insights into nuclear organization". In: *Genes Dev* 26.1 (2012), pp. 11–24.

[118] K. J. Wright, 2nd Marr M. T., and R. Tjian. "TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter". In: *Proc Natl Acad Sci U S A* 103.33 (2006), pp. 12347–52.

[119] X. Xie et al. "Structural similarity between TAFs and the heterotetrameric core of the histone octamer". In: *Nature* 380.6572 (1996), pp. 316–22.

[120] P. Yakovchuk et al. “RNA polymerase II and TAFs undergo a slow isomerization after the polymerase is recruited to promoter-bound TFIID”. In: *J Mol Biol* 397.1 (2010), pp. 57–68.