

Another piece in the transcription initiation puzzle

Francisco J Asturias

A new report provides evidence that the TFIIB-RNAPII interaction depends on the presence of additional factors and highlights the importance of structural characterization of the entire preinitiation complex.

Control of gene expression underlies cellular differentiation and development, and constitutes an essential checkpoint in signaling and oncogenesis. Therefore, functional and structural characterization of the complex machinery responsible for transcription and its regulation in higher eukaryotic cells is crucial. Our understanding of the centerpiece of the eukaryotic transcription apparatus, RNA polymerase II (RNAPII), has advanced tremendously with the determination of its high-resolution structure by X-ray crystallography^{1–3}. The structure revealed the organization of the 12 RNAPII subunits and provided a solid basis for interpreting results from numerous functional studies. Further information came from X-ray studies of RNAPII–DNA–RNA complexes, which revealed much about the RNA chain elongation mechanism^{4,5}. However, less was learned about the crucial initiation step, on which regulation is largely focused. EM studies of frozen-hydrated single RNAPII particles⁶ revealed the conformation of the enzyme in solution and suggested how the enzyme might initially interact with DNA, but the picture remained incomplete.

The real issue is that, despite its complexity, RNAPII must interact with several additional proteins to form a preinitiation complex (PIC) that is able to recognize a promoter and initiate transcription^{7–9}. The additional proteins, known as the general transcription factors, position RNAPII at the transcription start site, unwind a portion of the double-stranded DNA to allow access to the template strand, and eventually help RNAPII to clear the promoter and switch to the elongation phase. The PIC constitutes the target of the Mediator complex, a large (1 MDa) complex that includes ~20 different polypeptides and functions as the interface between the PIC and activators or repressors that convey signals from enhancer and silencer elements^{10–15}. In essence, understanding the

assembly and structure of the preinitiation complex formed by RNAPII and the general transcription factors is the key to understanding regulation during the initiation process. Unfortunately, other than the X-ray structure of the TFIIB–TBP–TATA element complex published over ten years ago^{16,17}, little direct structural information about the PIC or its components has been obtained.

The first models of PIC organization precede the RNAPII X-ray structure, and were largely based on sequence-specific DNA–protein photocrosslinking data that provided information about the interaction of polymerase and the general transcription factors with different segments of a promoter^{18,19}. It became apparent that promoter DNA made extensive contacts with both the polymerase and some of the transcription factors, but interpretation of the results in the absence of structural data was challenging. These early models, along with information from low-resolution EM studies, suggested that DNA could wrap around RNAPII, and that multiple copies of some of the transcription factors might be included in the PIC. Functional characterization studies that have assigned specific functions to individual general transcription factors have been used to validate and interpret structural analysis of complexes of RNAPII with individual factors, and in the last year, a cryo-EM structure of the yeast RNAPII–TFIIF complex²⁰ and an X-ray structure of the yeast RNAPII–TFIIB complex²¹ have been published. Interaction of RNAPII with these two factors is particularly interesting because a complex containing only IIB, IIF, and RNAPII can engage in a TBP-mediated interaction with a promoter and initiate transcription under conditions where mechanical stress facilitates promoter opening^{22,23}.

The cryo-EM structure of the RNAPII–TFIIF complex revealed substantial similarity between TFIIF and the bacterial σ factor. By considering the structures of bacterial RNAP holoenzyme–DNA and TFIIB–TBP–TATA complexes, and the location of TFIIB suggested by preliminary X-ray analysis of the RNAPII–TFIIB complex, a model for arrangement of the RNAPII–TFIIB–TBP–DNA complex was proposed²⁰.

The approximate path for DNA could be predicted from the cryo-EM structure of the RNAPII–TFIIF complex and comparison with the structure of the bacterial holoenzyme–DNA complex, but the assignment of the location of TFIIB was tentative. A few months later, more detailed information about the location of the zinc ribbon and B-finger domains of TFIIB (Fig. 1a) came from X-ray crystallography analysis of the RNAPII–TFIIB complex. Most significant was the revelation that the B-finger reaches deep into the active site cleft of RNAPII, where it probably plays a role in determining the transcription start site²¹. However, the position of the linker and core domains could only be inferred, as they were not ordered in the structure. Accurately determining the position of the core TFIIB domain is particularly important because it interfaces with the TATA-box element and TBP^{16,17}, and therefore its location will largely determine the topology of the PIC. Nonetheless, information from the RNAPII–TFIIB X-ray structure, along with consideration of the RNAPII–TFIIF and bacterial holoenzyme–DNA complexes, was used to propose a model for PIC organization, which also included rather speculative locations for factors IIE and IIH²¹.

Now, new and very significant information has come from a study that differs from previous ones in a very important respect: it examines the interaction between TFIIB, TFIIF and RNAPII in the context of the entire PIC²⁴. Hahn and colleagues have long been using immobilized template assays to study transcription in yeast^{25–27}. Their approach relies on attachment of a promoter to a substrate, followed by exposure to yeast nuclear cell extracts that contain all of the components of the transcription machinery. Under appropriate conditions, functional complexes can be assembled and conveniently characterized. They have now carried out their PIC assembly studies using nuclear extracts containing TFIIB mutants with hydroxyl radical-generating groups engineered into conserved TFIIB sites^{28,29}. After formation of the PIC, the location of cleavage sites, induced by the site-specific generation

*The author is at the Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.
e-mail: asturias@scripps.edu*

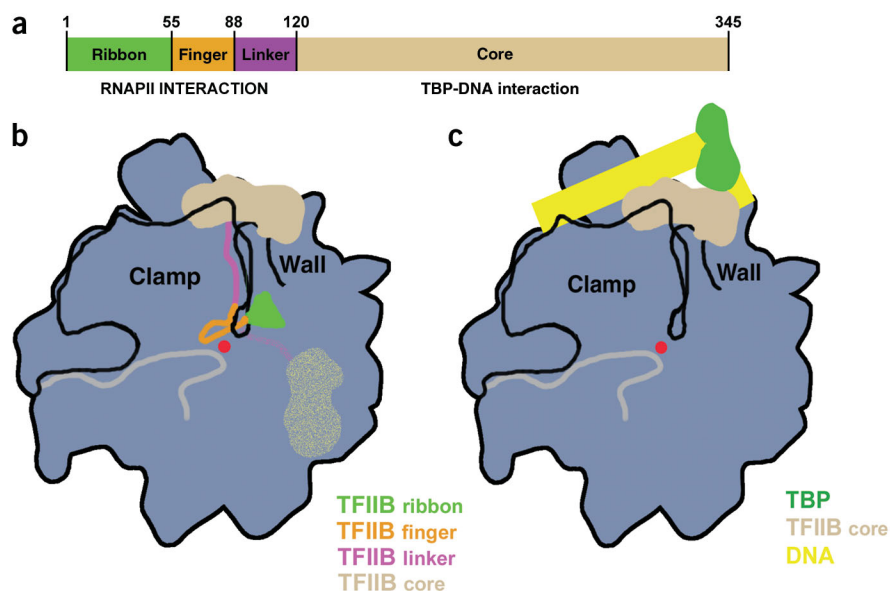


Figure 1 The position of TFIIB on the RNAPII preinitiation complex. **(a)** Domain structure of TFIIB. Approximate amino acid boundaries between different domains are indicated. **(b)** Interaction of TFIIB with RNA polymerase II. The RNAPII structure is shown in a side view, where the clamp domain is in front (subunits Rpb4 and Rpb7 have been left out for clarity). The active site (red dot) is located in a deep cleft formed between the clamp and the rest of the polymerase structure (gray line, bottom of the active site cleft). The cleft is blocked at the end by the wall domain. Different TFIIB domains are color-coded as indicated. The N-terminal ribbon domain interacts with the dock portion of the RNAPII structure. The B-finger reaches from there into the active site cleft, where it plays an important role in transcription start site selection. The interaction of these two domains with RNAPII must be strong, as their location does not depend on the presence of additional components of the PIC. The linker and core domains are disordered in a RNAPII–TFIIB complex, which led to incorrect prediction of their location (indicated by the fuzzy colored areas) from X-ray crystallography data. The linker and core domains become ordered when other components of the PIC (minimally DNA and TFIIF) are present. The linker moves upward along the internal face of the wall, and the core is positioned above the wall, at the upstream end of the active site cleft of the polymerase. **(c)** Position of the TFIIB core–TBP–TATA box DNA complex on the RNAPII structure. The expected position of the TFIIB core domain was used to model the position of the X-ray structure of the TFIIB core–TBP–TATA box DNA complex on the surface of polymerase. Promoter DNA runs down the back face of RNAPII and (after bending at the TATA box by TBP) along the top of the clamp, above the active site cleft. Promoter melting would allow the template strand to reach the polymerase active site at the bottom of the cleft. Features of the polymerase structure are the same as in **b**.

of hydroxyl radicals, could be mapped to the surface of the RNAPII X-ray structure, providing information about interactions between different TFIIB domains and RNAPII. Parallel studies using immobilized RNAPII provided information about RNAPII–TFIIB interactions in an RNAPII–TFIIB complex. Interestingly, but not entirely surprisingly, the results were not the same. In agreement with the X-ray studies, the zinc ribbon domain interacts with the dock domain of RNAPII, and the B-finger reaches deep into the active site cleft of RNAPII. The interaction of these two TFIIB domains with RNAPII must be strong and not dependent on other factors. However, the locations of the linker and the core domains are context-dependent. The positions of the linker and the core domains are poorly defined in the RNAPII–TFIIB complex, and these domains

become localized only in the presence of additional PIC components, notably promoter DNA (Fig. 1b).

Armed with more precise information about the location of the core TFIIB domain, Chen and Hahn²⁴ have revisited the organization of the RNAPII–TBP–TFIIB–DNA complex. Some reassuring similarities, especially with the model based on the cryo-EM structure of the RNAPII–TFIIF complex²⁰, are noted. Upstream promoter DNA runs along the back face of the polymerase until bending by TBP causes it to change direction. The location of the TFIIB–TBP–TATA complex has now been shifted closer to the upstream end of the active site cleft, near the wall domain, and (as seen in the bacterial holoenzyme fork junction DNA structure³⁰, and proposed in the cryo-EM model of the PIC²⁰) this causes promoter DNA downstream of the TATA box

to run along the top of the active site cleft. Positioning of the promoter is consistent with the results from several crosslinking studies, and would allow the transcription start site on the template strand to reach the active site located at the bottom of the narrow polymerase active site cleft (Fig. 1c). Like previous models of the PIC, the one presented by Chen and Hahn²⁴ depends critically on the presence of the TATA-box element and the consequent bending of promoter DNA by TBP. We still need to understand what happens in the case of many TATA-less promoters. Other components of the transcription machinery seem to have structural elements that could be involved in bending DNA, but it is not clear how such factors could complement or entirely replace the precise action of TBP.

Hahn and his group²⁴ were also interested in determining which other factors might interact closely with TFIIB and RNAPII. To this end, they conducted site-specific photocrosslinking studies that revealed extensive interactions between the TFIIB finger, linker and core domains, and the Tfg1 and Tfg2 TFIIF subunits. The interaction of the TFIIB core with Tfg2 near the upstream entrance to the active site cleft is consistent with the localization of Tfg2 in the cryo-EM structure of the RNAPII–TFIIF complex²⁰, and is particularly interesting, as Tfg2 has been proposed to function as a structural homolog of the bacterial σ factor and to play a role in directing promoter DNA to the polymerase active site cleft. However, Chen and Hahn²⁴ also detected extensive interaction of the finger and linker TFIIB domains with Tfg1 in the active center cleft, where no Tfg1 density was detected by the cryo-EM analysis. In fact, the cryo-EM reconstruction of the RNAPII–TFIIF complex localized only ~70% of the TFIIF density, and it was suggested that TFIIF would perhaps be fully ordered only in the presence of additional components of the PIC. The results presented by the Hahn group support that suggestion, and emphasize the importance of TFIIF for PIC assembly. Perhaps the most important contribution from the work presented by the Hahn group is the direct, strong evidence that the large multicomponent complexes involved in transcription are like puzzles made of flexible pieces, in which the picture only makes sense when all of the components are present. This realization should guide the design and interpretation of any further structural studies of the eukaryotic transcription machinery.

ACKNOWLEDGMENTS

The author acknowledges helpful comments from A. Jawhari about the manuscript and figures.

- Cramer, P. *et al. Science* **288**, 640–649 (2000).
- Cramer, P., Bushnell, D.A. & Kornberg, R.D. *Science* **292**, 1863–1876 (2001).
- Bushnell, D.A. & Kornberg, R.D. *Proc Natl. Acad. Sci. USA* **100**, 6969–6973 (2003).
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. & Kornberg, R.D. *Science* **292**, 1876–1882 (2001).
- Westover, K.D., Bushnell, D.A. & Kornberg, R.D. *Science* **303**, 1014–1016 (2004).
- Craighead, J., Chang, W. & Asturias, F. *Structure* **10**, 1117–1125 (2002).
- Hahn, S. *Nat. Struct. Mol. Biol.* **11**, 394–403 (2004).
- Dvir, A., Conaway, J.W. & Conaway, R.C. *Curr. Opin. Genet. Dev.* **11**, 209–214 (2001).
- Woychik, N.A. & Hampsey, M. *Cell* **108**, 453–463 (2002).
- Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M. & Kornberg, R.D. *Science* **283**, 985–987 (1999).
- Boube, M., Joulia, L., Cribbs, D.L. & Bourbon, H.M. *Cell* **110**, 143–151 (2002).
- Kelleher III, R.J., Flanagan, P.M. & Kornberg, R.D. *Cell* **61**, 1209–1215 (1990).
- Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D. *Cell* **77**, 599–608 (1994).
- Lewis, B.A. & Reinberg, D. *J. Cell. Sci.* **116**, 3667–3675 (2003).
- Malik, S. & Roeder, R.G. *Trends Biochem. Sci.* **25**, 277–283 (2000).
- Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. *Nature* **365**, 512–519 (1993).
- Nikolov, D.B. *et al. Nature* **377**, 119–128 (1995).
- Kim, T.K. *et al. Proc. Natl. Acad. Sci. USA* **94**, 12268–12273 (1997).
- Robert, F. *et al. Mol. Cell* **2**, 341–351 (1998).
- Chung, W.H. *et al. Mol. Cell* **12**, 1003–1013 (2003).
- Bushnell, D.A., Westover, K.D., Davis, R. & Kornberg, R.D. *Science* **303**, 983–988 (2004).
- Tyree, C.M. *et al. Genes Dev.* **7**, 1254–1265 (1993).
- Parvin, J.D. & Sharp, P.A. *Cell* **73**, 533–540 (1993).
- Chen, H.T. & Hahn, S. *Cell* (in the press).
- Reeves, W.M. & Hahn, S. *Mol. Cell. Biol.* **23**, 349–358 (2003).
- Yudkovsky, N., Logie, C., Hahn, S. & Peterson, C.L. *Genes Dev.* **13**, 2369–2374 (1999).
- Yudkovsky, N., Ranish, J.A. & Hahn, S. *Nature* **408**, 225–229 (2000).
- Chen, H.T. & Hahn, S. *Mol. Cell* **12**, 437–447 (2003).
- Datwyler, S.A. & Meares, C.F. *Trends Biochem. Sci.* **25**, 408–414 (2000).
- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. & Darst, S.A. *Science* **296**, 1285–1290 (2002).

Opening the GAP

Diacylglycerol is a lipid second messenger involved in cell signaling and phorbol esters are natural products that mimic that action. Diacylglycerol mediates signaling from a variety of hormones through G protein-coupled receptors, from growth factors via receptor tyrosine kinases, and many other intra- and extracellular agents. The binding of diacylglycerol or phorbol esters to the protein kinase C homology 1 (C1) domain of a target protein results in its translocation to the membrane and subsequent allosteric activation. The mechanism for this lipid-mediated membrane association has been examined by studies of C1 domains from the protein kinase C (PKC) family. Several other C1-containing proteins including the α - and β -chimerins have recently been identified.

The chimerins are GTPase-activating proteins (GAP) that bind phorbol esters with nanomolar affinity in the presence of acidic phospholipids. β 2-chimerin is composed of three conserved domains: an N-terminal SH2 domain for phosphotyrosine binding to an unknown partner, a C1 domain, and a C-terminal Rac-specific GAP domain. Rac is a small G protein that regulates an array of cellular activities, including the cell cycle, actin dynamics and transcription. Rac binding to RacGAP promotes its GTPase activity. To understand what conformational changes occur upon lipid binding and how this leads to Rac activation, Hurley and colleagues have solved the structure of β 2-chimerin (*Cell*, in the press, 2004).

The three domains of β 2-chimerin are arranged with the C1 (blue) sandwiched between RacGAP (green) and SH2 (red). The C1 and RacGAP domains have the canonical fold of their representative family members, while the C-terminal part of the SH2 domain resembles that of the adaptor protein APS.

The N-terminus (beige wire) runs the length of the protein, covering the RacGAP active site and the lipid-binding pocket of C1. The C1 domain makes extensive contacts with the SH2 and RacGAP domains, and the residues in C1 that form the basic face for putative membrane association are solvent-exposed. However, the hydrophobic residues implicated in diacylglycerol and phorbol ester binding are buried and form intramolecular interactions with the rest of the protein. This implies that a dramatic conformational change is required prior to the binding of lipid second messengers to the C1 domain. Mutational analysis indicates that the destabilization of C1 interactions with the remainder of the protein favors the conformational changes that occur upon membrane binding. This suggests that the β 2-chimerin intramolecular contacts compete with phorbol ester and phospholipids for binding to the C1 domain.

Modeling of the β 2-chimerin structure with that of Rac-GDP, combined with comparison to the liganded p50RhoGAP domain structure, shows how β 2-chimerin may activate Rac. When lipid binding occurs, a large conformational change causes the N-terminus of β 2-chimerin to move away from the RacGAP domain, allowing Rac to bind. The RacGAP domain interacts directly with the Rac active site, stabilizing its transition state for GTP hydrolysis. In essence, the N terminus of β 2-chimerin serves as an inhibitory domain similar to the N-terminal pseudosubstrate sequences used by the PKCs. However, whereas the PKCs have two C1 domains—one specialized for translocation and the other for activation— β 2-chimerin has a more rudimentary system whereby both functions are accomplished by a single C1 domain. It is remarkable that proteins with such differing targets would have similar modes of activation.

Michelle Montoya

