TRANSCRIPTION INITIATION: A COMPLEX GAME WITH MANY PLAYERS

Transcription of eukaryotic genes is a complex process requiring the action of a myriad of proteins. Central to the process is the large multi-subunit enzyme RNA polymerase II (Pol II), which contains 12 subunits (Lewin, 1997). In spite of its complexity, Pol II requires a series of general transcription factors (GTFs) to recognize, bind, and clear the core promoter of each gene. These GTFs are TFIID, TFIIA, TFII B, TFIIE, TFIIF, and TFIIH (Conway and Conway, 1993; Maldonado and Reinberg, 1995) (Fig. 1). Regulation of gene transcription occurs at all stages, initiation, elongation, and termination. Control of transcription at the initiation stage relies on the existence of multiple specific sequences that have regulatory duties. In front of the start of the coding region is the core promoter of the gene, the place where the transcriptional machinery assembles. In addition, there are enhancer and silencer sequences, sometimes far from the coding region, to which activators and repressors bind and, respectively, facilitate and inhibit the transcription of a gene. In the absence of activators bound to enhancer elements, the core transcription complex made up of the six GTFs and Pol II can accurately initiate basal levels of RNA synthesis in vitro. In the presence of gene-selective enhancer- and promoter-binding activators, significantly elevated levels of transcription initiation can be achieved (Zawel and Reinberg, 1995). This gene-specific regulation is thought to be fine-tuned by the synergistic effect of several activators/repressors acting simultaneously. One current view of mammalian gene regulation is that activators stimulate transcription largely by promoting assembly of the general transcriptional machinery at the core promoter (Ptashne and Gann, 1997). Alternatively, activators could have a direct effect on the conformation of some of the GTFs and/or DNA structure (Chi and Carey, 1996; Roberts and Green, 1994). As Pol II escapes from the promoter and elongates through the gene, it allows the entrance of another polymerase so that multiple mRNAs can be produced. Some studies support the idea that a major effect of activators is to stimulate promoter escape (Sandaltzopoulos and Becker, 1998).

Of the six essential GTFs mentioned above, TFIID, which contains the TATA-binding protein (TBP), is the first to bind to the core promoter and to facilitate the assembly of the complete initiation machinery. In addition to TBP, TFII D contains up to 12 other proteins, TBP-associated factors (TAFs), which are involved in the recognition of promoter elements other than the TATA box and are important for the regulation of the general machinery by activators (Goodrich and Tjian, 1994). In the current model of initiation machinery assembly, TFII A and TFII B contact DNA and TBP, increasing the stability of TBP binding. Then a complex of TFII F–Pol II is recruited, followed by TFII E and TFII H. TFII H is required to melt the DNA at the start site and for Pol II to clear the promoter and initiate transcription. Although this is a generally accepted model, the order of assembly, and the stability of any intermediate complex bound or unbound to DNA, still needs direct structural confirmation. An alternative model is that the general machinery assembles in solution and binds to the DNA...
酶作为一个异聚酶，激活子促进其招募。一些激活子已被证明可以直接与 TAFs 在 TFII D 中相互作用。在 TAFs 中，它们是蛋白质复合物，通常被称为辅助因子或’中介体’，被认为是激活酶和 GTFs 的桥梁。这些最特征化的就是 yeast 调节器，其作用是通过元件和 Pol II 调节器组成。几个哺乳动物调节器复合物已被隔离，包括 TRAP（或 DRIP）复合物，人类调节器，CRSP 和 ARC (Malik 和 Roeder, 2000; Hampsey 和 Reinberg, 1999)。

**STRUCTURES OF RNA POLYMERASES**

虽然在 eukaryotes 中的核内转录是通过复合物，multienzyme Pol II 及其一般转录因子（Tjian, 1996），细菌转录，以及其他作为辅助因子或 ‘中介体’，它们被认为是桥梁的调节器和 GTFs。这些最常见的就是 yeast 调节器，其作用是通过元件和 Pol II 调节器组成。几个哺乳动物调节器复合物已被隔离，包括 TRAP（或 DRIP）复合物，人类调节器，CRSP 和 ARC (Malik 和 Roeder, 2000; Hampsey 和 Reinberg, 1999)。

在 recent years several crystal structures of T7RNAP have been determined, including a ligand-free structure (Sousa et al., 1993) and structures of T7RNAP complexed to a transcriptional inhibitor (lysosyme) (Jeruzalmi and Steitz, 1998), an open promoter DNA fragment (Cheetham et al., 1999) and a promoter DNA being transcribed (Cheetham and Steitz, 1999; Cheetham and Steitz, 2000). T7RNAP is similar in structure to the Pol I DNA polymerase family (Doublé et al., 1998; Kiefer et al., 1998). It resembles a right hand, in which the finger, thumb and palm domains are arranged around a central cleft where the active site, formed by highly conserved residues, is located.

In eukaryotes there are three types of nuclear RNA polymerase: Pol I synthesizes ribosomal RNA, Pol II synthesizes messenger RNAs, and Pol III synthesizes transfer RNA and other small RNAs. Recently, the structure of the core RNA polymerase (RNAP) of *Thermus aquaticus* has been determined at 3.3 Å resolution (Zhang et al., 1999), and more recently Cramer and co-workers solved that of the large yeast Pol II to 3.5 Å (Cramer et al., 2000; Fig. 2A,B). The bacterial core RNAP resembles very closely the corresponding core in the larger, eukaryotic Pol II. The two largest subunits, β and β’ (Rpb1p and Rpb2p in yeast Pol II), interact to form the central part of the enzyme, enclosing a large channel at the base of which sits the catalytic site. The α dimer (Rpb3p and Rpb1p in yeast Pol II) does not contribute to the channel, and its location supports the view that it functions to aid the assembly of the β and β’ subunits. In the eukaryotic Pol II, Rpb5p, Rpb9p and part of Rpb1p surround the DNA-binding channel (Poglitsch et al., 1999), forming a pair of jaws that could enhance the stability and processivity of the complex by gripping the DNA template. A pore or secondary channel below the catalytic site Mg2+ ion might function as a conduit for the entry of ribonucleotide triphosphates in both the bacterial and the eukaryotic polymerases. A previously identified hinge domain (Fu et al., 1999) is formed by Rpb1p, Rpb2p and Rpb6p, and could act as a sliding clamp that locks the DNA template in place but allows movement during elongation. The two structures have allowed development of a model for the DNA and RNA paths through the enzyme that awaits confirmation by high-resolution structures of the elongating complex.

**COMPLEXES OF POLYMERASE, DNA AND GENERAL TRANSCRIPTION FACTORS**

TFIIB is thought to play an important role in transcription by bridging Pol II to the TBP promoter complex and determining the transcription start site (Buratowski et al., 1989; Li et al., 1994). The solution structure of the human TFIIB core domain is characterized by a pseudo-twofold symmetry and similarity to the cyclin A fold (Bagby et al., 1995). An initial structure of TFIIB in a ternary complex with TBP and DNA obtained by Nikolov and co-workers showed that the N-terminal domain of core TFIIB forms a downstream surface that could fix the transcription start site (Nikolov et al., 1995). Tsai and Sigler have recently obtained the structure of a human TFIIB-TBP bound to an extended adenovirus major late promoter (Tsai and Sigler, 2000). This structure reveals that TFIIB binds both to the major groove upstream and in the minor groove downstream of the TATA box. This extended, asymmetrical binding could be the determinant for the unidirectional assembly of the transcription initiation machinery.

Leuther et al. obtained a projection structure of TFIIB bound to yeast Pol II from 2-D co-crystals and image analysis a few years ago (Leuther et al., 1996). Difference mapping allowed the localization of the TFIIB density relative to the polymerase. Because TFIIB binds directly to TBP, which sits near the center of the TATA sequence (Nikolov et al., 1995), the polymerase-bound TFIIB identifies the approximate location of the TATA box in a transcription initiation complex. In the studies of Leuther and co-workers, TFIIB and the active site in Pol II, at the base of the channel, were separated by 110 Å in projection, which corresponds to 32 bp of B-form DNA. Given that the distance between the TATA box and the start site is typically 30 bp for Pol II promoters, the authors proposed that the DNA follows a straight path on the surface of the polymerase to the active center. Leuther and co-workers also localized the binding position of TFII E on the polymerase to near one of the claws engulfing the DNA-binding cleft. On the bases of this they proposed that TFII E promotes closure of the polymerase (Leuther et al., 1996). Recently Okuda et al. obtained the structure of the core domain of TFII E by NMR (Okuda et al., 2000). The structure showed that this domain is formed by three
helices and β hairpin, which resembles winged-helix proteins. Surprisingly, the dsDNA-binding activity is located on a positively charged furrow that is on the opposite side of that reported for other winged-helix motifs (Okuda et al., 2000).

Poglitsch et al. have studied the interaction of Pol II with DNA at low resolution by 2-D crystallography of stalled complexes (Poglitsch et al., 1999). They used a DNA tail template consisting of 33 bp of duplex DNA that had a 15-residue deoxycytidylate tail at one 3' end and a biotinilated residue at the other 3' end. In the absence of UTP, transcription paused 11 bp into the duplex region. Comparison of the 3-D structure of the elongation complex and that of Pol II alone showed extra densities that were assigned to the DNA bubble occupying the polymerase channel, and a tetrameric streptavidin marking the downstream end of the DNA. Cramer et al., have integrated this and additional information with the crystallographic structure of yeast Pol II to propose a model of nucleotide-protein interaction during transcriptional elongation (Cramer et al., 2000).

MEDIATOR

Some preliminary structural information exists on yeast mediator from low-resolution projections of this complex alone and bound to Pol II (Asturias et al., 1999). In the absence of Pol II, yeast mediator has a well-defined, compact conformation. The averaged projection at 40 Å showed a roughly triangular shape that has a distinct, dense domain at the base. Upon binding to Pol II, yeast mediator partially unfolds, adopting an extended conformation that partially surrounds the polymerase. The extended structure contains three domains: the head domain (proposed to correspond to the dense base of the compact form), the middle domain, and the tail domain (Fig. 3). Both the head and middle domains seem to contact the polymerase. Manual alignment of one of the views from the 16-Å 3-D reconstruction of Pol II (Darst et al., 1991) with the density assigned to it in the holoenzyme (Pol II + mediator) indicates that the middle domain of yeast mediator contacts polymerase near the C-terminal domain, whereas the head domain contacts it near the DNA-binding channel. Mediator isolated from murine cells has a similar general appearance, which suggests an overall conservation of the mediator structure across eukaryotes.

STRUCTURES OF TFIID AND RELATED FACTORS

X-ray diffraction and NMR studies have revealed the structures
of various subunits contained within TFII D. For example, the high-resolution structures of TBP bound to TATA box DNA and domains of either TFIIA or TFII B have been determined (Geiger et al., 1996; Nikolov et al., 1995; Tan et al., 1996). The crystal structure of TBP bound to the TATA box reveals that binding dramatically bends and unwinds the DNA (Burley and Roeder, 1996) and demonstrates an induced-fit mechanism of protein-DNA recognition. Crosslinking and DNase footprinting have shown that TAFs, in conjunction with TFIIA, change the behavior of TBP alone and induce conformational changes in the complex that some speculate lead to the wrapping of the core promoter around TFII D (Hoffmann et al., 1997).

The structures of some of the smaller TAFs, and domains of some of the larger TAFs, have also been obtained by X-ray crystallography and NMR studies. The co-crystal structure of Drosophila dTAF42-dTAF62 (Xie et al., 1996) showed that these TAFs are folded into a classical core-histone protein motif, in which a central helix is flanked by a coil and short helix segment. In the crystal structure, dTAF42 and dTAF62 form a heterodimer in which they interact with one another in a head-to-tail fashion and a heterotetramer like that formed by the H3-H4 histones.

Recently, Jackson and coworkers solved the structure of the bromodomains of human TAF250 by X-ray crystallography (Jacobson et al., 2000). This largest subunit of TFII D contains an intrinsic histone-acetyltransferase activity (Mizzen et al., 1996). The activity localizes to a double bromodomain module, which is invariant in TFII Ds from yeast to humans. The core of each bromodomain is formed by a left-handed four-helix bundle. At the center of this bundle is a deep, hydrophobic pocket that constitutes the putative acetyl-lysine-binding sites. The two bromodomains are rotated 108° with respect to the axis of the helical bundle, giving rise to a V-shaped structure in which the N- and C-termi ni are closed together at the vertex of the V, and the active sites lie at opposite ends and are separated by 25 Å. This arrangement is optimally suited to bind acetyl groups separated by seven amino acid residues, the acetylation pattern present in H4 histones. This ability of the bromodomains of TAF250 to recognize histone H4 tails that have transcriptionally relevant patterns of acetylation (Jacobson et al., 2000) led the authors to propose that TFII D is targeted to promoters near or within regions of DNA packed into nucleosomes. Thus, the role of nucleosomes in gene transcription might be more dynamic than previously anticipated.

In spite of this progress, both the size of the complex and the inherent difficulties in obtaining large quantities of purified holo-TFII D have precluded conventional X-ray diffraction studies of the full complex. Consequently, the overall shape and relative positions of the components within TFII D, and the architecture of even larger initiation complexes, remain unknown. How are the different TAFs organized in TFII D with respect to one another, TBP and DNA? How does TFII D interact with the other basal factors and with RNA polymerase to position the latter at the starting point of transcription? How do TAFs interact with activators or repressors, both directly or through the action of cofactors, and how is this information relayed to the polymerase? Structural information on several of these supracomplexes will be required if we are to address these questions directly, the final aim being to understand the ‘molecular mechanics’ of the transcription initiation complex and the process of gene regulation.

Recently, low-resolution structures of TFII D and the TBP-free TAF-containing complex (TFTC) were obtained by electron microscopy and image-reconstruction techniques (Andel et al., 1999; Brand et al., 1999a). At 35-Å resolution TFII D has a horseshoe shape in which three main structural lobes surround a central cavity (Andel et al., 1999; Fig. 4). The lobes are connected by narrow regions of density, which...
suggests that relative movement of the lobes (which would change the size and shape of the central cavity) can be easily accomplished, perhaps upon binding of the complex to DNA or upon interaction with activators or other GTFs. Mapping of TBP-specific antibodies localized the position of the TBP in the central density lobe, where it faces the central cavity (Andel et al., 1999). This supports the idea that the cavity is a major site for DNA binding. The X-ray structural analysis of the double bromodomain of TAF250 mentioned above, combined with quantitative binding assays, revealed that this TAF specifically recognizes a di-acetylated histone-H4 tail peptide (Jacobson et al., 2000). The central cavity in the TFIID reconstruction is large enough to accommodate an entire nucleosome, further supporting the possibility that TFIID does not require binding to ‘naked’ DNA but may instead interact specifically and intimately with appropriately modified chromatin templates.

My co-workers and I have studied dimeric and trimeric complexes containing TFIIB and both TFIIB and TFIIA to identify by difference mapping the relative binding locations of these GTFs on the TFIID structure (Andel et al., 1999). TFIIA and TFIIB bind on opposite lobes of TFIID across the central cavity, in both cases near narrow density bridges across lobes (Fig. 4). Their relative positions with respect to TBP correspond approximately to those expected from the crystal structures of TBP-TFIIA (Geiger et al., 1996; Tan et al., 1996) and TBP-TFIIB (Nikolov et al., 1995), although the three elements appear somehow more separated. A segment of the density attributed to TFIIA occupies part of the central cavity, which suggests a possible increase in the protein surface interacting with DNA upon binding of this factor; this is consistent with the known increase in DNA affinity of TFIID in the presence of TFIIA (Kobayashi et al., 1995).

The positions of TFIIA and TFIIB on the larger TFIID, together with the existing knowledge of the interaction between different TAFs and other components of the initiation complex, allow generation of a model for the distribution of TAFs on the three lobes of the TFIID structure (Fig. 4): TFIIB is known to contact Pol II (Li et al., 1994; Malik et al., 1993) and to interact with TAF32 (Klemm et al., 1995); TAF130 is known to interact with TFIIA (Yokomori et al., 1993); and TAF250 interacts with TBP and is known to be central to assembly of the complex. Finally, TAF18, TAF20 and TAF28 interact directly with TBP. In addition, a considerable number of TAF-TAF contacts have been revealed by a variety of biochemical studies (Burley and Roeder, 1996).

The TBP-free complex TFTC is able to direct preinitiation complex formation and initiation of transcription on both TATA-containing and TATA-less promoters in in vitro transcription assays (Wieczorek et al., 1998). The polypeptide composition of TFTC suggests that TFTC is recruited by activators to acetylated histones and thus may potentiate initiation and activation of transcription (Brand et al., 1999b). The structure of TFTC (obtained at somewhat lower resolution) is similar to that of TFIID overall, but contains an extra domain that gives it a more elongated shape (Brand et al., 1999a) (Fig. 5). The seven TAFs common to TFTC and TFIID (accounting for 58% of the molecular mass of TFIID) are probably organized into a similar core around the central cavity. The extra domain present in TFTC occupies a volume that would correspond to ~400 kDa and is thus similar to that expected for the TRTC-specific subunit TRRAP.

**THE STRUCTURE OF TFIIH**

Two different groups have recently reported the structure of TFIIH, a complex shared by the transcription initiation machinery and the transcription-coupled DNA repair machinery (Drapkin et al., 1994; Feaver et al., 1993). This complex has several known enzymatic activities. The core TFIIH complex contains the 3′-5′ helicase XPB, the 5′-3′ helicase XPD, and four more additional subunits. The holoenzyme in addition has a kinase complex, containing a catalytic subunit, CDK7, and two regulatory proteins (cyclin H and MAT1). This full complex is essential for the phosphorylation of the C-terminal domain of Pol II required for promoter clearance. Studies of 2-D crystals of a yeast core TFIIH lacking one of the helicases (Radp25p, a homologue of XPB) have rendered an 18-Å resolution view of this part of the complex (Chang and Kornberg, 2000) (Fig. 6, red/blue/green). Using previous genetic and biochemical evidence the authors have made a tentative assignment of the five subunits within the structure. They propose that Tbp1p (p62), Tbp2p (p52) and Tbp3p (MAT1) form a ring, connected to two ellipsoidal densities, which were assigned to Ssl1p (p44) and Rad3p (XPD). In an accompanying paper, Schultz and co-workers obtained the structure of the human holo-TFIIH at 38-Å resolution, using single-particle methodology (Schultz et al., 2000) (Fig. 6, yellow). The TFIIH is organized into a ring-like structure from which a large protein domain protrudes. A subcomplex assembled from five recombinant core subunits forms a similar circular architecture in which the protruding
domain is missing. Immunolabelling experiments showed that the CDK7 subunit is found in the protruding density, subunit p44 is within the ring substructure at the basis of the CDK7-containing protein density, and the two helicases XPB and XPD are also part of the annular domain, on either side of p44. The human and yeast studies are not fully compatible, because the ring densities are of different sizes, and the yeast core TFIIH has a large non-ring density area in spite of lacking the kinase complex. In addition, the assignment of Rad3p in the yeast structure to one of the two ellipsoidal densities is in clear disagreement with the position of its homologue, XPD, within the ring in the human holoenzyme as determined by antibody labelling. These differences may be a consequence of a dramatic change in architecture in the absence of the second helicase p44, which has recently rendered low-resolution structures of Taq and yeast TFIIH (red, blue and green) as fitted by Chang and Kornberg (Chang and Kornberg, 2000). The location of different subunits by antibody labeling is indicated in yellow for the human TFIIH. Assignment of the subunits in the yeast core TFIIH is based on existing biochemical and genetic evidence. Courtesy of Patrick Schultz (CNRS) and W. Chang and R. Kornberg (Stanford University). Reprinted from Cell, 98, Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. and Darst, S. A. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. 811-824. Copyright (1999), with permission from Elsevier Science.

**Fig. 6.** Structures of human holo-TFIIH (yellow) and yeast core TFIIH (red, blue and green) as fitted by Chang and Kornberg (Chang and Kornberg, 2000). The location of different subunits by antibody labeling is indicated in yellow for the human TFIIH. Assignment of the subunits in the yeast core TFIIH is based on existing biochemical and genetic evidence. Courtesy of Patrick Schultz (CNRS) and W. Chang and R. Kornberg (Stanford University). Reprinted from Cell, 98, Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. and Darst, S. A. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. 811-824. Copyright (1999), with permission from Elsevier Science.

**REFERENCES**


Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. 


