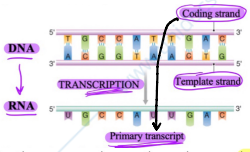
2. RNA polymerase II

1

RNA polymerase II is the enzyme that catalyses 5' to 3' RNA polymerisation during transcription.

$RNA_n + NTP + Mg^{+2} + DNA\ template \rightarrow RNA_{n+1} + PP_i$ (pyrophosphate)



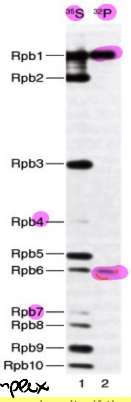
In the general reaction above, the magnesium ion and the DNA template are essential cofactors.

2

RNA polymerase II is highly processive (10^6 base pairs before dissociating) and its main functions are unwinding the DNA duplex, synthesising RNA and proofreading. The enzyme assembles into larger initiation and elongation complexes, capable of promoter recognition and response to regulatory signals.

4

The number of subunits of RNA polymerase II was obtained through epitope tagging, in which they attached a small foreign epitope to one of the yeast polymerase II subunits (Rpb3) by engineering its gene. Then they introduced this gene into yeast cells lacking a functional Rpb3 gene, labeled the cellular proteins with either ^{35}S (radioactive sulphur: sulphur is present only in proteins and not in nucleic acids) or ^{32}P (radioactive phosphorus), and used an antibody directed against the foreign epitope to precipitate the whole enzyme. After immunoprecipitation, they separated the labeled polypeptides of the precipitated protein by SDS-PAGE (SDS-PAGE (sodium dodecyl sulphate–polyacrylamide gel electrophoresis) is a discontinuous electrophoretic system commonly used as a method to separate proteins with molecular masses between 5 and 250 kDa) and detected them by autoradiography. This single-step purification method yielded essentially the pure RNA polymerase II with 10 apparent subunits (because Rpb9 and Rpb10 contain 2 subunits each). Band thickness depends on the stoichiometry/ratio of the subunits and on the amount of cysteine and methionine in the polypeptide: for example Rpb4 and Rpb7 are present with a ratio of 0.5.



5

The non-essential subunits are Rpb4 and Rpb9: cells can, in fact, grow without these subunits if they are cultivated in optimal conditions. Rpb1 and Rpb2 have the highest molecular weight. Rpb3 and Rpb11 somehow resemble the α subunit of the bacterial polymerase.

Five subunits (Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12) are found in all three yeast nuclear polymerases. We know little about the functions of these subunits, but the fact that they are found in all three polymerases suggests that they play roles fundamental to the transcription process.

There is a sort of similarity between the Rpb1 and Rpb2 subunits of RNA polymerase II and the β and β' subunits of bacterial polymerase. The rest have additional functions not found in bacterial enzymes.

The Rpb1 subunit of RNA polymerase II is the largest and contains what is called the CTD tail (carboxy-terminal domain): it typically consists of up to 52 repeats (in human, 26 in yeast) of the sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (heptamer). Not all the repeats follow the consensus and variations usually occur on the 7th amino acid. Other proteins can bind to the C-terminal domain of

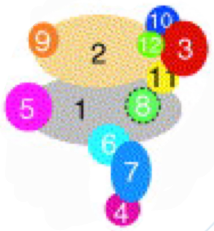
6

The most powerful and efficient way for determining the shape of a protein is x-ray crystallography, which can only be performed if all the polymerases under study are in the same state (no CTD, Rpb4 or Rpb7).

The structure of yeast polymerase II reveals a deep cleft that can accept a DNA template (formed between Rpb1 and Rpb2). The catalytic centre, containing a Mg^{+2} ion, lies at the bottom of the cleft. The other subunits are arranged around the main cleft. A second Mg^{+2} ion is present in low concentration, and presumably enters the enzyme bound to each substrate nucleotide.

The enzyme presents both types of secondary structures: α-helices (major component) and β-sheets.

The Rpb1, Rpb5 and Rpb9 subunits are mobile and form the jaws of the polymerase. The clamp, formed by the Rpb6, Rpb1 and Rpb2 subunits, is active during elongation and it covers the active site, bending the DNA: this increases stability and processivity of the enzyme. The small groove at the bottom of the clamp) represents an exit for the newly synthesised transcript. Rpb7 and Rpb4 close the clamp over the DNA when the RNA first exits the enzyme. The activity of the clamp can be regulated through phosphorylation. Linking Rpb1 and Rpb2, there is an α-helix named 'bridge helix', which is essential to guide both the catalytic reaction and translocation of RNA polymerase II.



Removal of one of the smallest subunits compromises the enzyme's stability and activity less than removing one of the largest ones.

The negative charges of the enzyme are present on the surface of the enzyme and the positive charges are present in the groove (in order for interaction with the DNA duplex to occur).

7

The DNA template initially enters the first chamber (jaw-lobe), which binds 15-20 base pairs without melting the duplex. The DNA melts when entering the second chamber: melting is due to the intervention of three protein loops, which are the rudder, the lid and the fork. The three loops form a strand-loop network, whose stability must drive the melting process.

RNA-DNA hybrid

The DNA template is bent with an angle of 90° when passing through the polymerase's groove, because the theoretical straight path contains a channel too small for the duplex to pass through and a wall at the end (formed by a portion of Rpb2).

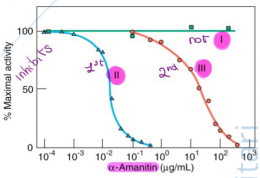
On the bottom of the enzyme, there is a pore (pore 1) which enters into a channel that communicates directly with the active site (it provides easy access to the active site). There is also pore 2.

~ 8-9 ribonucleotides of the newly synthesises RNA form a hybrid heteroduplex with the template DNA strand. The nucleotides enter the enzyme randomly, so most of them exit immediately. If the nucleotide is right, it remains in the active site long enough for the phosphate and hydroxyl group to come closer together and change the orientation of the bridge helix. This explains RNA polymerase II's high fidelity. In moving through the entry pore toward the active site of RNA polymerase II, an incoming nucleotide first encounters the E (entry) site, where it is inverted relative to its position in the A site, the active (or addition) site where phosphodiester bonds are formed. Another magnesium ion is present at the active site; the one already mentioned is permanently bound to the enzyme and the other one enters the active site complexed to the incoming nucleotide. The trigger loop of Rpb1 positions the substrate for incorporation and discriminates against improper nucleotides.

3

RNA polymerase II (transcribes mRNA and all non coding RNA) is not the only eukaryotic polymerase: we also have RNA polymerase I (mainly rRNA) and RNA polymerase III (mainly tRNA).

α-amanitin is a specific inhibitor for RNA polymerase II. An experiment, in fact, proved α-amanitin was found to have different effects on the three polymerases. At very low concentrations, it inhibits polymerase II completely while having no effect at all on polymerases I and III. At 1000-fold higher concentrations, the toxin also inhibits polymerase III from most eukaryotes. The inhibition then allowed to observe what products were no longer transcribed in order to understand which polymerase transcribes what RNA.

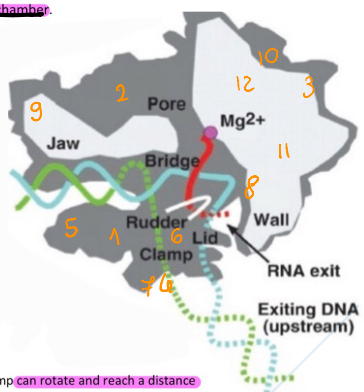


There are also particular loop structures made by Rpb2 and Rpb1 that with the wall will define the upper side of the chamber.

Two other important domains are the jaws and the clamp.

The jaws:
Are made by subunits 9 and 1 in one side and on the other by rpb5 and 1.
This domain has the action of catching the DNA and be sure that it will sits stably in the protein complex.

The clamp:
Is composed of subunits 6.
It is a sort of high hill, very important because can close over the DNA by changing its structure.
In the new conformation the hill is down to the DNA and it will close it. (it close the cleft)



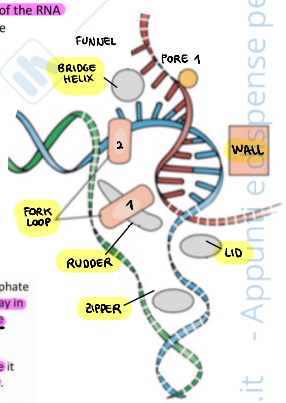
The clamp domain is very flexible. The position of the clamp can rotate and reach a distance that is equal to 30 Armstrong.

When the DNA enters the clamp, the clamp is still opened then when elongation of the RNA begins it closes onto the DNA, in this way the DNA is much more stable so that the dissociation from the enzyme is more difficult.

Melting the RNA-DNA hybrid:
Melting of the hybrid due to the intervention of three protein loops:

- **Rudder** ("ror") contacting DNA
- **Lid**, contacting RNA. A Phe side chain serves as wedge to maintain separation of the strands.
- **Fork loop 1** contacts base pairs **6** and **7**, limiting the strand separation.

The three loops form a strand-loop network, whose ability must drive the melting process.



A site

When entering the cleft, the incoming nucleotide will be orientated such that its phosphate will not be neat the OH of the RNA. Only if they are the right pair, there it will be a delay in the cleft that will permit to the nucleotide of rotate and then couple with the template base. **The rotation is enough to catalysis to occur.**

After the catalysis there is another important event. Now that the incoming nucleotide it has been added to the transcript, the cleft is full, and another nucleotide cannot enter.

For the next cycle to occur the cleft must be **regenerated**. (the template must be relative to the protein, if you block the movement, you block the action of the transcript)

Translocation (is coupled with the catalytic reaction) moves the DNA near the protein. Conformational changes of the protein generate movements.

The **bridge helix triggers the movement of DNA.** The bridge will bend so that it will push the hetero duplex and then it will return to the previous form forming a new cleft. The translocation is due to the bridge that will push the nucleic acid structure.

Rpb4 and 7 form a **dimer** that interacts with the 10 subunits complex through a small interface between Rpb7 and Rpb1 just at the base of the clamp, on the side where there is the narrow groove.

Rpb4 and 7 interact with the complex at the base of the clamp.

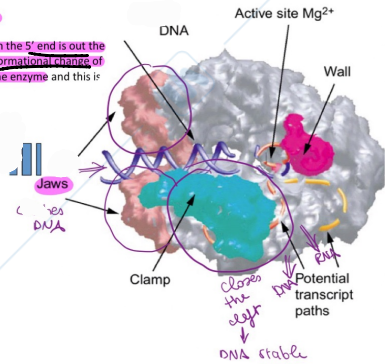
Rpb4 and 7 are important for the **regulation of the elongation phase.**

When exiting the transcript is very sensible to the Rpb4-7 dimer. When the 5' end is out the dimer sense the presence of the transcript and they will trigger a conformational change of the clamp that will close over the DNA increasing the processivity of the enzyme and this is when the elongation phase starts.

RNA exit:
The exit groove has an appropriate length and localization for binding a region of RNA 10-20 nt from the active site.

RNA in the groove at the base of the clamp could explain the great stability of transcribing complexes.

The first nucleotide of the transcript will be out at around 19 bp, that is a length too small for the RNA to be used and called. So until the RNA don't become 25-30 bp long because some part of it is still inside the narrow cleft.



8

Elongation is not a continuous phase: it can be interrupted and, sometimes, when proofreading, the polymerase has to backtrack (abortive initiation/transcription). Backtracking is irreversible because the catalytic cycle is interrupted. TFIIS (transcription elongation factor II S) promotes the elongation of arrested RNA polymerase II by stimulating the inherent RNA cleavage activity of RNA polymerase II.

TFIIS has two domains linked together through a flexible amino acid chain (linker). Domain III is rich in acidic amino acids and can reach the catalytic site through pore 1. RNA polymerase II alone has a nuclease activity, but it isn't very strong in absence of TFIIS. Cleavage can be of two types: one uses a water molecule, the other uses a pyrophosphate (the result is equivalent: both molecules basically revert the synthesis reaction of the polymerase).

3. Eukaryotic promoters of class II

Transcription is divided into three main phases: initiation, elongation and termination. Promoters are regions, usually upstream from the TSS, that control transcription initiation and determine the point at which it starts.

There are two types of class II promoters: CpG islands (about 300-500 bp long) and TATA containing promoters. CpG islands are areas rich in CG content representing 67% of promoters. They do not have a fixed initiation site (bidirectional promoter). TATA containing promoters have one initiation site and are unidirectional: divergent genes close together with TATA containing promoters need a TATA box each.

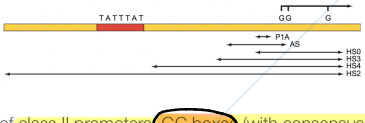
Class II promoters can be considered as having two parts: the core promoter and the proximal promoter. The core promoter attracts general transcription factors and RNA polymerase II at a basal level and sets the transcription start site and direction of transcription. It consists of elements lying within about 37 bp of the transcription start site, on either side. The proximal promoter helps attract general transcription factors and RNA polymerase and includes promoter elements that can extend from about 37 bp up to 250 bp upstream of the transcription start site. Elements of the proximal promoter are also sometimes called upstream promoter elements (UAS, upstream activating sequence).

The core promoter contains the TATA box, with consensus sequence TATA(A/T)A(A/T). It is homologous to the -10 element in bacteria (but it's 25-30 bp upstream from the TSS), rich in T and A. The function of the TATA box is better understood by studying transcription in its absence.

In this experiment, Christophe Benoist and Pierre Chambon (1981) performed a deletion mutagenesis study of the SV40 early promoter. The assays they used for promoter activity were primer extension (isolation of RNA, hybridisation of a complementary DNA primer, which is radioactively labelled, extension of the primer to the end of the RNA using reverse transcriptase and denaturation of the RNA/DNA hybrid) and S1 mapping. The products were labeled DNA fragments whose lengths tell us where transcription starts and whose abundance tells us how active the promoter is.

The P1A, AS, HS0, HS3, and HS4 mutants, created by deleting progressively more of the DNA downstream of the TATA box, including the initiation site, simply shortened the S1 signal by an amount equal to the number of base pairs removed by the deletion. Such a shift is what we would predict if the TATA box positions transcription initiation approximately 25 to 30 bp downstream of the last base of the TATA box. The gel of the H2 deletions shows that removing the TATA box caused transcription to initiate at a wide variety of sites, while not decreasing the efficiency of transcription. If anything, the darkness of the S1 signals suggests an increase in transcription. Thus, it appears that the TATA box is involved in positioning the start of transcription and its strength.

This conclusion was reinforced by systematically deleting DNA between the TATA box and the initiation site of the SV40 early gene and locating the start of transcription in the resulting shortened DNAs by S1 mapping. Transcription of the wild-type gene begins at three different guanosines, clustered 27-34 bp downstream of the first T of the TATA box. As more and more of the DNA between the TATA box and these initiation sites was removed, transcription started at other bases, usually purines, that were about 30 bp downstream of the first T of the TATA box. In other words, the distance between the TATA box and the TSSs remained constant, with little regard to the exact sequence at these initiation sites.



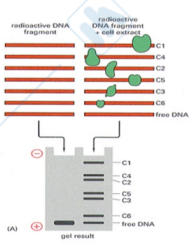
There are two upstream elements found in a variety of class II promoters: GC boxes (with consensus GGGCGG and CGCCCC) and CCAAT boxes (with consensus GGCCAATCT). A specific transcription factor called Sp1 binds to the GC boxes and stimulates transcription. The CCAAT box must also bind a transcription factor (the CCAAT-binding transcription factor [CTF], among others) to exert its stimulatory influence.

Promoters are different from enhancers: they both stimulate transcription, but differ in two important respects: enhancers are position- and orientation- independent. GC boxes are orientation-independent, but they do not have the position independence of classical enhancers. Chambon discovered the first enhancer in the 59-flanking region of the SV40 early gene. This DNA region had been noticed before because it contains many copies of a 72-bp sequence, called the 72-bp repeat. When deletion mutations were made in this region, they observed profoundly depressed transcription in vivo. This behaviour suggested that the 72-bp repeats constituted another upstream promoter element. However, they also discovered that the 72-bp repeats still stimulated transcription even if they were inverted or moved all the way around to the opposite side of the circular SV40 genome, over 2 kb away from the promoter. Thus, such orientation- and position-independent DNA elements were called enhancers to distinguish them from promoter elements.

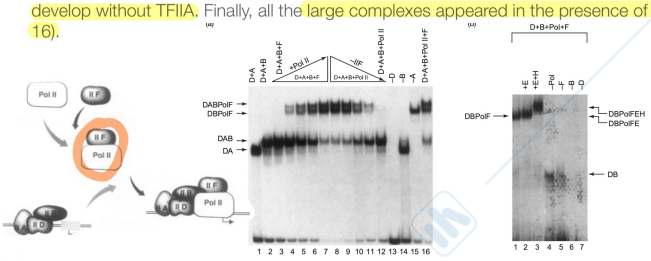
The function of transcription factors is to recruit RNA polymerase II to the promoter, given it is unable to bind it alone, and consequently bind the TSS. There are different types of transcription factors: general transcription factors and upstream transcription factors.

The general transcription factors combine with RNA polymerase to form a preinitiation complex that is competent to initiate transcription as soon as nucleotides are available. This tight binding involves formation of an open promoter complex in which the DNA at the transcription start site has melted to allow the polymerase to read it. There are six general transcription factors named TFIIA, TFIIB (single polypeptide, like TFIIA), TFIID (10-12 subunits), TFIIE, TFIIF, and TFIIH.

An electrophoretic mobility shift assay (EMSA) is a common affinity electrophoresis technique used to study protein–DNA or protein–RNA interactions, in this case, the general transcription factor's interaction with DNA. This procedure can determine if a protein or mixture of proteins is capable of binding to a given DNA or RNA sequence, and can sometimes indicate if more than one protein molecule is involved in the binding complex. Gel shift assays are often performed in vitro concurrently with DNase footprinting, primer extension, and promoter-probe experiments when studying transcription initiation, DNA gang replication, DNA repair or RNA processing and maturation, as well as pre-mRNA splicing. It is an electrophoretic separation of a protein–DNA or protein–RNA mixture on a polyacrylamide or agarose gel for a short period (about 1.5-2 hr for a 15- to 20-cm gel). The speed at which different molecules (and combinations thereof) move through the gel is determined by their size and charge, and to a lesser extent, their shape. The control lane (DNA probe without protein present) will contain a single band corresponding to the unbound DNA or RNA fragment. However, assuming that the protein is capable of binding to the fragment, the lane with a protein that binds present will contain another band that represents the larger, less mobile complex of nucleic acid probe bound to protein which is 'shifted' up on the gel (since it has moved more slowly).

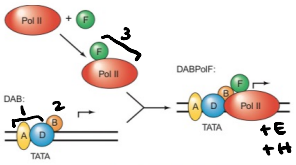


By observing the EMSA gels, we were able to conclude that TFIID is the protein complex that binds to DNA first, given without it no other protein was able to associate to the free DNA. Lane 1 shows the DA complex, formed with TFIID and A. Lane 2 demonstrates that adding TFIIB caused a new complex, DAB, to form. Lane 3 contained TFIID, A, B, and F, but it looks identical to lane 2. Thus, TFIIF did not seem to bind in the absence of polymerase II. Lanes 4–7 show what happened when the investigators added more and more polymerase II in addition to the four transcription factors: more and more of the large complexes, DABPolF and DBPolF, appeared. Lanes 8–11 contained less and less TFIIF, and we see less and less of the large complexes. Finally, lane 12 shows that essentially no DABPolF or DBPolF complexes formed when TFIIF was absent. Thus, TFIIF appears to bring polymerase II to the complex. The lanes on the right show what happened when one factor at a time was removed. In lane 13, without TFIID, no complexes formed at all. Lane 14 shows that the DA complex, but no others, formed in the absence of TFIIB. Lane 15 demonstrates that DBPolF could still develop without TFIIA. Finally, all the large complexes appeared in the presence of all the factors (lane 16).

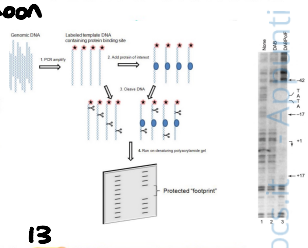


In the gel on the right, they started with the DBPolF complex (lacking TFIIA, lane 1) assembled on a labeled DNA containing the adenovirus major late promoter. Next, they added TFIIE, then TFIIH, in turn, and performed gel mobility shift assays. With each new transcription factor, the complex grew larger and its mobility decreased further. Lanes 4–7 show the result of leaving out various factors, denoted at the top of each lane. At best, only the DB complex forms. At worst, in the absence of TFIID, no complex at all forms.

Thus, the order of addition of the general transcription factors (and RNA polymerase) to the preinitiation complex in vitro is as follows: TFIID (or TFIIA + TFIID), TFIIB, TFIIF + polymerase II, TFIIE, TFIIH (the participation of TFIIA seems to be optional in vitro).



To study where the complex binds on the DNA, a DNase I footprinting assay was performed. DNase I footprinting assay is an in vitro method to identify the specific site of DNA binding proteins. It not only finds the target protein that binds to specific DNA, but also identifies which sequence the target protein is bound to. This technique can be used to study interactions between proteins and DNA both outside and within cells. The protein binds to the DNA fragment, protecting the binding site from cleavage by the DNase I. The fragments of the DNA molecule are left after cleavage, thus its sequence can be determined. On the autoradiogram of the polyacrylamide electrophoresis gel, there is no radio-labelled band corresponding to the site of protein binding. Each sequence differs from the previous one by one nucleotide, to it is a sort of sequencing (the missing bases are those corresponding to the sequence that the protein covers). Footprinting assay is specific, provides accurate positioning, and is widely used. On the right are the results provided by DNase I footprinting assay of the DABPolF complex. The DAB complex occupies the bases going from -17 to -42 and covers the TATA box. When TFIIF is present and the polymerase can bind to the complex, it occupies

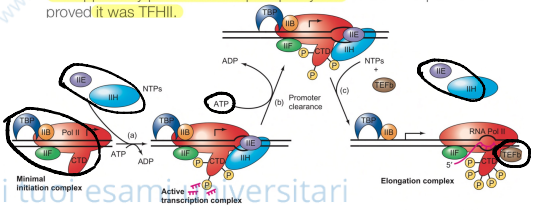


One subunit of TFIID is the TATA binding protein (TBP), which is always present together with the biggest subunit. It is the most conserved protein among the complex. The protein is almost symmetrical, with two subunits having the same secondary structure. The major groove of DNA is the one usually contacted by proteins, but the β sheet of the TBP contacts the minor groove. It causes a sort of widening of the minor groove, helped by the presence of the Ts and As. The TBP is involved in DNA melting (double strand separation) by bending the DNA with an 80° angle (the AT-rich sequence to which it binds facilitates melting).

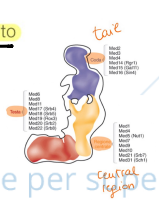
Footprinting studies (figure below, on the left) have indicated that the TAFs (TBP associated factors) attached to the TBP extend the binding of TFIID well beyond the TATA box in some promoters. In particular, TBP seemed to protect the 20 bp or so around the TATA box in some promoters, but TFIID protected a region extending to position 135, well beyond the TSS. This suggested that the TAFs in TFIID were contacting the initiator and downstream elements in these promoters. To investigate this phenomenon in more detail, TBP and TFIID abilities to transcribe DNAs bearing two different classes of promoters in vitro were tested. The first class (the adenovirus E1B and E4 promoters) contained a TATA box, but no initiator or downstream promoter element (DPE). The second class (the adenovirus major late promoter and the *Drosophila* heat shock protein [*hsp70*] promoter) contained a TATA box, an initiator, and a downstream promoter element. We can see that TBP and TFIID sponsored transcription equally well from the promoters that contained only the TATA box (compare lanes 1 and 2 and lanes 3 and 4). But TFIID had a decided advantage in sponsoring transcription from the promoters that also had an initiator and downstream promoter element (compare lanes 5 and 6 and lanes 7 and 8). Thus, TAFs apparently help TBP facilitate transcription from promoters with initiators and downstream promoter elements.

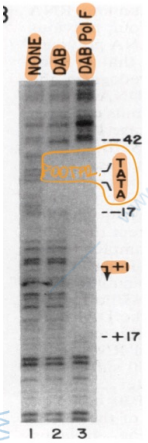
The last general transcription factor to join the preinitiation complex is TFIIH. It appears to play two major roles in transcription initiation: one is to phosphorylate the CTD of RNA polymerase II, the other is to unwind DNA at the TSS to create the transcription bubble and help clear the way for the polymerase (helicase function).

Reinberg demonstrated that TFIIH was a good candidate for the protein kinase that catalyses this CTD phosphorylation. First, he showed that the purified transcription factors, by themselves, are capable of phosphorylating the CTD of polymerase II. The evidence came from a gel mobility shift assay. Lanes 1–6 demonstrate that adding ATP had no effect on the mobility of the DAB, DABPolF, or DABPolFE complexes. On the other hand, after TFIIH was added to form the DABPolFEH complex, ATP produced a change to lower mobility. One possibility is that one of the transcription factors in the complex had phosphorylated the polymerase. Indeed, when the polymerase was isolated from the lower mobility complex, it proved to be the phosphorylated form, polymerase IIO. But polymerase IIA had been added to the complex in the first place, so one of the transcription factors had apparently performed the phosphorylation. Further experiments proved it was TFHII.



Another huge protein complex, made of over 20 polypeptides, is mediator, which can also be considered a general transcription factor because it is part of most class II preinitiation complexes. Unlike the other general transcription factors, Mediator is not required for initiation per se, but it is required for activated transcription. It can affect both default transcriptional rate and polymerase regulation. Recent studies confirm the hypothesis that Mediator marks genes for RNA polymerase I binding, which subsequently activates the preinitiation complex.





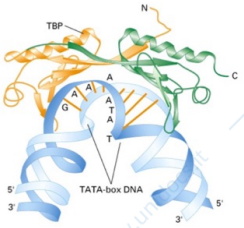
In the lane 1 there is the free DNA, we see there are different cuts, different intensities. When we incubate DNA with TFIID, A and B we see a footprint right on he location of the TATA box, when including pol II and TFIIF we maintain the footprint at the level of the TATA box. The pol II will set downstream the TATA box, we see that from the cuts.

TFIID binds to the TATA box while pol II downstream of it.

This model explain even why the TATA box is always at the same distance from the transcript start site.

TFIID contains 13 TAFs, in addition to TBP. Most of these TAFs are evolutionarily conserved in the eukaryotes. The TAFs serve several functions, but two obvious ones are interacting with core promoter elements and interacting with activators. TAF1 and TAF2 help TFIID bind to the initiator and DPEs of promoters and therefore can enable TBP to bind to TATA-less promoters that contain such elements. TAF1 and TAF4 help TFIID interact with Sp1 that is bound to GC boxes upstream of the transcription start site. These TAFs therefore ensure that TBP can bind to TATA-less promoters that have GC boxes. Different combinations of TAFs are apparently required to respond to various activators, at least in higher eukaryotes. TAF1 also has two enzymatic activities. It is a histone acetyltransferase and a protein kinase.

The TBP is always represented as a half moon, because of its shape we can understand that it will be the subunits that will bind to the TATA box.



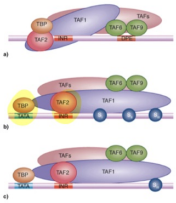
The DNA will bend, around 80°, the bend corresponds to the TATA box. Because of this bend the two strands will tend to separate.

Let's make a hypothesis:
When TAFs are present, all together, there is an increase in the transcription level only at certain promoter. (Way to regulate transcription)

Now testing with DNA footprint assay:
Two lanes with TBP alone, one with TBP attached with two big subunits, one line (the last one) in which there is a weight marker.

In the first lane there is free DNA, when we add TBP we can observe that there is the protection of the TATA box. When add the 2 subunits we see the same observation with the addition of some other protections downstream of the TATA box, this shows that the binding of the DNA is also made by some of the TAFs, the biggest one.

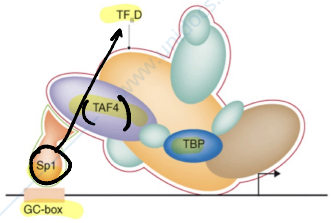
One more detail about the technique is that up the TATA box and downstream of it there is an enhancer of the cleavage, some time you see this kind of effect, because of the **distortion of DNA** caused by the binding protein, that distortion are more attacked by the nucleases.



TAF1 and 2 can also recognize the proximal promoter in the case in which the promoter is constituted by the TATA box.

Always the TBP will be with TFIIB that will recruit pol II.

What is the case of when the promoter is composed of CpG islands?



In this case the transcription factor that recognizes the promoter is called **Sp1**. Sp1 bind specifically to GC boxes. This protein is composed of 2 domains, the second is called **trans activating domain** which makes the activation that binds the Sp1 to the TFIID, precisely into the TAF4.

Summary:

- The recognition of the promoter is made from **basal transcription factors** TFIID.
- Sp1 is an **activating transcription factor**

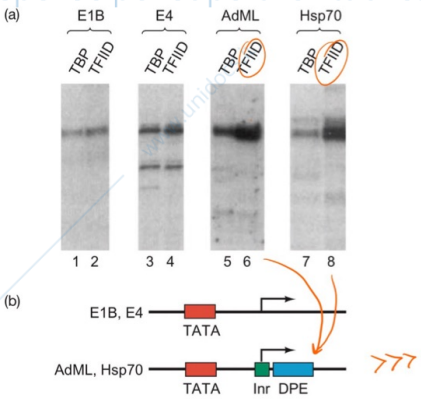


Figure 11.8 Activities of TBP and TFIID on four different promoters. **(a)** Experimental results. Tjian and colleagues tested a reconstituted *Drosophila* transcription system containing either TBP or TFIID (indicated at top) on templates bearing four different promoters (also as indicated at top). The promoters were of two types diagrammed in panel **(b).** The first type, represented by the adenovirus E1B and E4 promoters, contained a TATA box (red). The second type, represented by the adenovirus major late promoter (AdML) and the *Drosophila hsp70* promoter, contained a TATA box plus an initiator (Inr, green) and a DPE (blue). After transcription in vitro, Tjian and coworkers assayed the RNA products by primer extension (top). The autoradiographs show that TBP and TFIID fostered transcription equally well from the first type of promoter (TATA box only), but that TFIID worked much better than TBP in supporting transcription from the second type of promoter (TATA box plus Inr plus DPE). (*Source:* Verrijzer, C.P., J.-L. Chen, K. Yokomari, and R. Tijan, Promoter recognition by TAFs. *Cell* 81 (30 June 1995) p. 1116, f. 1. Reprinted with permission of Elsevier Science.)

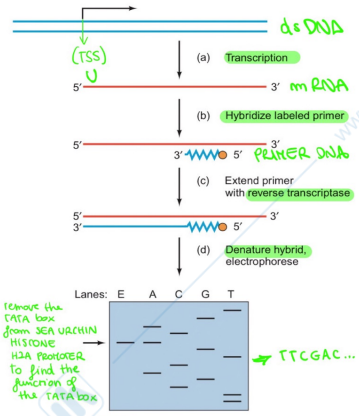
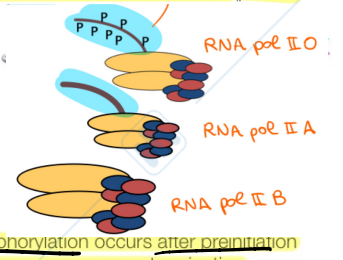


Figure 5.30 Primer extension. **(a)** Transcription occurs naturally within the cell, so begin by harvesting cellular RNA. **(b)** Knowing the sequence of at least part of the transcript, synthesize and label a DNA oligonucleotide that is complementary to a region not too far from the suspected 5′-end, then hybridize this oligonucleotide to the transcript. It should hybridize specifically to this transcript and to no others. **(c)** Use reverse transcriptase to extend the primer by synthesizing DNA complementary to the transcript, up to its 5′-end. If the primer itself is not labeled, or if it is desirable to introduce extra label into the extended primer, labeled nucleotides can be included in this step. **(d)** Denature the hybrid and electrophorese the labeled, extended primer (experimental lane, E). In separate lanes (lanes A, C, G, and T) run sequencing reactions, performed with the same primer and a DNA from the transcribed region, as markers. In principle, this can indicate the transcription start site to the exact base. In this case, the extended primer (arrow) coelectrophoreses with a DNA fragment in the sequencing A lane. Because the same primer was used in the primer extension reaction and in all the sequencing reactions, this shows that the 5′-end of this transcript corresponds to the middle A (underlined) in the sequence TTCGACTGACAGT.

The CTD tail is unique for RNA polymerase II and has an essential function in vivo. Different eukaryotic promoters show different dependence on the CTD.

Tyrosine, serine and threonine are the reversibly phosphorylatable amino acids on the CTD (proline cannot be phosphorylated).

- RNA polymerase IIO is the fully phosphorylated polymerase;
- RNA polymerase IIA is the fully dephosphorylated polymerase;
- RNA polymerase IIB is the polymerase with no CTD tail.]



The phosphorylation level changes during transcription: phosphorylation occurs after preinitiation complex assembly, and dephosphorylation occurs on free polymerase or upon termination.

The CTD has a role in recruitment of RNA polymerase II to promoters during initiation and a role in promoter clearance: CTD phosphorylation by TFIIH (i) disrupts interactions and RNA polymerase II is freed from the preinitiation complex and (ii) creates novel interactions with elongation factors.

There are different proteins that CTD can bind, including SRBs (suppressors of RNA polymerase B), general transcription factors and several proteins involved in pre-mRNA processing.

Kinases and phosphatases can have different functions based on wether RNA polymerase II is bound to DNA or not. Kinases have a repressor effect on the free RNA polymerase II and an activating effect if it is bound to DNA. Vice versa for phosphatases.

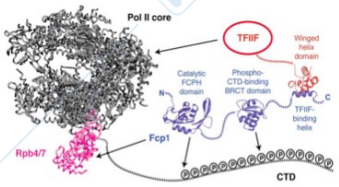
Four of the CTD kinases are members of the cyclin-dependent kinase (CDK)/cyclin family, whose members consist of a catalytic subunit bound to a regulatory cyclin subunit.

The main kinases are Cdk7, Cdk8 and Cdk9:

- Cdk7 has an activating effect, because it acts on RNA polymerase II when bound to DNA. It phosphorylates Ser5;
- Cdk8 has a repression effect (cyclin C) and also phosphorylates Ser5;
- Cdk9 is a component of pTEFb (positive transcription elongation factor) and it phosphorylates Ser2.

There are other kinases phosphorylating the CTD, but these are the main ones.

Ser5 phosphorylation is detected mainly at promoter regions (initiation) and Ser2 phosphorylation is seen only in coding regions (elongation).



The first CTD phosphatase characterised was FCP1: it is necessary for CTD dephosphorylation in vivo (Ser2). FCP1 presumably helps to recycle RNAP II at the end of the transcription cycle by converting RNAP IIO into IIA for another round of transcription. It is recruited through Rpb4/Rpb7.

Other phosphatases specific for Ser5 are:

- SCPs: are a family of small CTD phosphatases that preferentially catalyse dephosphorylation of Ser5. Expression of SCP1 inhibits activated transcription from a number of promoters. SCP1 may play a role in transition from initiation/capping to processive transcript elongation;
- Ssu72 is a component of the yeast cleavage/polyadenylation factor (CPF) complex and a CTD phosphatase with specificity for Ser5. Ssu72 may have a dual role in transcription: in recycling of RNAP II and in transcription termination.

The prolines in the CDT can be either in cis or trans conformation: Pin1 is an isomerase capable of changing the proline's conformation.