MINI
REVIEW

The basic RNA polymerase II transcriptional machinery

Roberto Weinmann

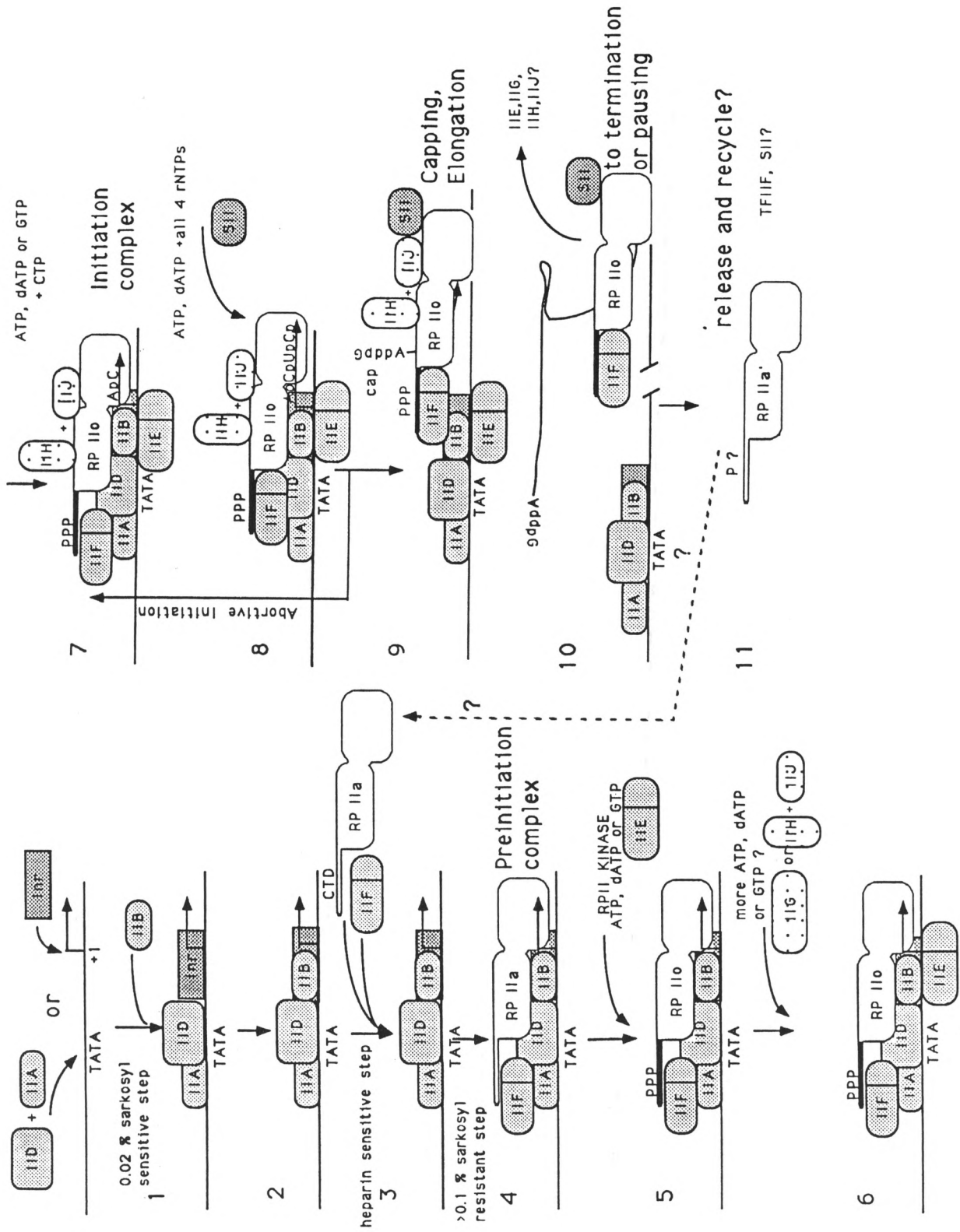The Wistar Institute of Anatomy and Biology, Philadelphia, Pennsylvania

All genes encoding proteins in eukaryotes are transcribed by RNA polymerase II. The first step in analyzing transcriptional regulation requires understanding the general mechanisms of RNA polymerase II-specific gene transcription. The basal promoter, a template containing a TATA box devoid of upstream regulatory sequences, has been used to identify and characterize the factors which, together with RNA polymerase II, govern transcription in mammalian systems: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIG, TFIIH, and TFIIJ. Interactions between regulatory transcription factors and basal elements of the transcriptional machinery affect the transcriptional rate in a positive or negative fashion. As these multiple proteins are purified, and their coding sequences are isolated, we come closer to reproducing these processes in vitro with pure components, and thus to elucidating the complex interactions among them.

P resent understanding of eukaryotic tran-
scriptional mechanisms derives in great part
from mammalian systems. In documenting cur-
rent knowledge in the field, this review will
emphasize mammalian transcription factors.
Yeast and Drosophila will be mentioned briefly
if complementary information is available, and
other organisms will be excluded entirely. Like-
wise, several controversial issues are beyond the
scope of this review. (For recent comprehensive
reviews, see Saltzman and Weinmann, 1989; Mer-
melstein et al., 1989; Sawadogo and Sentenac,
1990; and Stone et al., 1991.)

Figure 1 summarizes our current understand-
ing of transcription in mammalian cell systems.
It represents schematically the basic reactions
occurring during transcription initiation by RNA
polymerase II, together with the function of each
transcription factor. While detailed knowledge
of some of these processes has lagged because
of difficulties in obtaining purified factors, con-
siderable progress has been made recently
through the cloning of specific factors and an

abundance of cloned gene products in bacteria.
The availability of these proteins, free of other
contaminant polypeptides, allows us to assign
specific functions unequivocally. At the same
time, novel requirements are revealed each time
new, recombinant, or purified transcription
factors — the latter free of other mammalian
proteins — are used in reconstituted in vitro tran-
scription systems (see for example Dynlacht et
al., 1991; Meisterernst et al., 1991). Some of the
genes encoding transcription factors described
below are now cloned, and active, bacterially
expressed versions of their proteins can be easily
obtained in the lab in large amounts, while oth-
ers must still be purified from cell extracts.

The earlier fractionation of transcriptionally
competent HeLa cell extracts after phospho-
cellulose chromatography (Matsui et al., 1980)
resulted in fractions eluted at 0.1 M KCl, 0.3 M
KCl, 0.5 M KCl, and 1.0 M KCl. In transcription
reconstitution experiments, only fractions elut-
ing at 0.1, 0.5, and 1 M KCl were required (Mat-
sui et al., 1980). The fraction eluting at 1 M KCl

Initiation complex

ATP, dATP or GTP + CTP

ATP, dATP + all 4 rNTPs

SII

Capping, Elongation

IIE, IIG, IIH, IIJ?

to termination or pausing

release and recycle?

TFIIF, SII?

PPP IIF IIA IID TATA IIB IIE RP IIo IIH + IIJ ApC

7

Abortive Initiation

PPP IIF IIA IID TATA IIB IIE RP IIo CpUpCp IIH + IIJ

8

cap AppdG PPP IIF IIA IID TATA IIB IIE RP IIo IIH + IIJ + SII

9

AppdG P? IIF TATA IID IIB RP IIo SII

10

IIF IIA IID TATA IIB RP IIa'

11

Preinitiation complex

IID + IIA or

TATA +1

Inr

0.02 % sarkosyl sensitive step

1

IIA IID TATA Inr IIB

2

heparin sensitive step

CTD IIF IIA IID TATA Inr IIB

3

>0.1 % sarkosyl resistant step

IIA IID TATA IIB RP IIa IIF

4

RPII KINASE ATP, dATP or GTP IIE

IIF IIA IID PPP TATA IIB RP IIo

5

more ATP, dATP or GTP ? IIG + IIH + IIJ or

IIF IIA IID PPP TATA IIB RP IIo IIE

6

contained two activities, a poly-ADP-ribopoly-merase binding at DNA nicks (Slattery et al., 1983) and TFIID, while the fraction eluting at 0.5 M KCl contained TFIIB and TFIIE (Reinberg and Roeder, 1987).

The process of transcription initiation begins by the binding of protein TFIID to the template at the TATA box (step 1). The TATA box is an important promoter sequence element present in a large number of mammalian genes; it is located approximately 30 nucleotides upstream of the initiation start site. Binding of the IID protein is the earliest event in the process of transcription initiation and determines template commitment (Davison et al., 1983; Fire et al., 1984; Reinberg et al., 1987). When binding to the TATA box by a partially purified TFIID fraction precedes the assembly of chromatin on those same DNA fragments, the corresponding promoter is active in an in vitro transcription system (Workman and Roeder, 1987); in contrast, when chromatin assembly precedes TFIID binding, the promoter is inactive. Thus, the presence of the TATA box binding protein (TBP) is an essential determinant for fulfilling the potential for gene modulation by other transcription factors. Housekeeping genes and some viral genes (adenovirus E2e and IVa2) have AT-rich sequences which are distinct from typical TATA boxes. However, the protein fraction containing factor TFIID is still required for transcription of the adenovirus IVa2 gene and artificial SP1 TATA-less constructs in vitro (Carcamo et al., 1989; Pugh and Tjian, 1991). In the case of IVa2, the functional TATA box appears to be located downstream of the transcription start site, and the direction of transcription may be determined by the initiator (Smale and Baltimore, 1989; Carcamo et al., 1990, 1991). The initiator element, whose presence was suggested by the effect of point mutants around the start site (Lee et al., 1988), was demonstrated by the experiments of Smale and Baltimore (1989) on the terminal

deoxynucleotidyltransferase promoter, and confirmed by the experiments of Carcamo et al. (1989, 1990, 1991). The initiator element will be discussed more fully below.

The TATA box binding protein (TBP)

TFIID, purified from the fraction eluting with 1 M salt from phosphocellulose, participates in the initial steps of template commitment. This protein's instability has been a major roadblock for further purification of other factors. The differential thermal sensitivity of TFIID (Nakajima et al., 1988) and the fact that a yeast protein fraction can complement the human TFIID-depleted extracts (Buratowski et al., 1988) have allowed several groups to purify the yeast TFIID polypeptide and clone the yeast gene encoding for it. The cloned gene predicts a protein of 27 kDa for yeast (Cavallini et al., 1989; Hahn et al., 1989; Horikoshi et al., 1989; Schmidt et al., 1989; Eisenmann et al., 1989; Fikes et al., 1990). Using the sequence of the yeast TFIID gene, the human TBP gene was cloned (Kao et al., 1990; Peterson et al., 1990; Hoffman et al., 1990), and the open reading frame predicts a protein of 37 kDa. TBP contain a very highly conserved carboxy-terminal 160-amino acid domain with tandem repeats and homology to bacterial σ factor (Horikoshi et al., 1989; Hoffman et al., 1990). Limited homology to bacterial σ has also been found in RAP 30 (the small subunit of TFIIF), the fourth subunit of yeast RNA polymerase II, and mitochondrial RNA polymerase (reviewed by Jaehning, 1991). A single gene appears to encode for the human, Drosophila, and yeast TBP, but two TBP genes are present in Arabidopsis (Gash et al., 1990; Hoey et al., 1990; Hoffman et al., 1990). The human TBP contains an amino-terminal domain with unique features, such as a run of 38 glutamines in a row. Although the yeast and human proteins appear to function interchangeably in a mammalian in vitro tran-

---

Figure 1. Schematic representation of transcription initiation on the adenovirus 2 major late promoter (Ad2 MLP). The diagram summarizes data available as of October 10, 1991, and contains many modifications which are not fully described or attributed in the text, since not all of the data have been published. The location and role of some required factors, such as BTF-3 (Zheng et al., 1990), TFIIG (Sumimoto et al., 1990), or TFIIH (Flores et al., 1991), are less well characterized. The inverted G in step 9 indicates the 5'-5' CAP dinucleotide formation, catalyzed by a guanosyltransferase that may also be associated with the transcription complex (Reinberg et al., 1987). Interactions with elongation factor SII or TFIIS (Rappaport et al., 1987, 1988) and the abortive initiation process (Luse and Jacob, 1987; Rougvie and Lis, 1988) or pausing (Resnekov and Aloni, 1989) are not discussed in this review. They do, however, point to areas of further study, and to the fate of transcription factors and RNA polymerase in recycling.

scription system, the human TFIID is unable to substitute for the yeast TFIID in vivo. A region in the carboxy-terminal domain (CTD) of the yeast protein seems to be essential for this species-specific restriction (Gill and Tjian, 1991; McCormack et al., 1991). Binding specificity to modified TATA boxes (TGTAAA) correlates to amino acid changes in the second tandem repeat of the conserved domain (Strubin and Struhl, 1992). The purified yeast TBP induces DNA bending when bound to DNA (Horikoshi et al., 1992). Activation by upstream regulatory factors may require additional proteins (Peterson et al., 1990; Pugh and Tjian, 1990; Kelleher et al., 1990; Flanagan et al., 1991; Lewin, 1990). Most importantly, the yeast and mammalian TBP overproduced in bacteria are active in the in vitro transcriptional systems, insuring a steady source of this important reagent.

TBP-associated factors

Although in in vitro transcription experiments recombinant TBP is able to support transcription from basal promoter in a highly purified system, no response of upstream regulatory factors, such as SP1, could be detected, unless the less pure mammalian cell TFIID fraction was used (Peterson et al., 1990; Pugh and Tjian, 1990). The nomenclature we follow, in accord with these authors, is to call the multicomponent complex purified from cells TFIID, and to call the recombinant 37-kDa protein TBP, for TATA box binding protein. Purification or immunoprecipitation of TBP from cell extracts has revealed the presence of tightly associated proteins important for transactivation. They form what has been called the TFIID fraction or complex (Inostroza et al., 1991; Dynlacht et al., 1991). Some or all of the six cellular proteins coprecipitating with TBP-specific antibodies are required for activation by upstream factors (Dynlacht et al., 1991). Other proteins, like TFIID repressor (Dr), co-purifying with the TBP, might be negative regulators, since they displace the TBP-IIA DNA complex to form TBP-Dr complexes with distinct gel mobilities (J. Inostroza, F, Mermelstein, I. Ha, W. S. Lane, and D. Reinberg, unpublished data). A fraction containing upstream stimulatory activity (USA) responsible for stimulation by upstream SP1 and USF contains both positive and negative regulatory elements (Meisterernst et al., 1991) and may contain some of the same proteins found in the TFIID complex.

Transactivation

The exact mechanism by which TFIID-mediated transactivation occurs remains unclear. Promoter activity may be stimulated by direct interaction of acidic transactivator proteins with TFIID. The viral transactivator VP16 is able to interact with yeast TFIID (Stringer et al., 1990). Point mutations in VP16 that affect its transactivation potential also affect its interaction with TFIID (Ingles et al., 1991). Others have suggested that VP16 interaction with TFIID is weak, while interaction with TFIIB is strong (Lin and Green, 1991). Experiments in a yeast-reconstituted transcription system suggest that "mediator" proteins are required to detect gal4-VP16 effects (Kelleher et al., 1990). We and others have recently shown that human and yeast TFIIDs are able to interact directly with adenovirus E1A, a potent pleiotropic transactivator (Horikoshi et al., 1991; Lee et al., 1991). These interactions occur between the first repeat and the basic domain of TBP and the Zn^{++} finger of E1A (Lee et al., 1991; N. Horikoshi, U. Graeven, and R. Weinmann, unpublished data). In contrast to VP16 (Lin and Green, 1991), the interactions between E1A and TBP are strong, while those between E1A and TFIIB are weak (I. Ha, N. Horikoshi, D. Reinberg, and R. Weinmann, unpublished data). Recent studies have revealed negative regulators involved in basal and activated transcription. For example, in the case of the retinoblastoma protein whose association with transcription factor E2F represses its DNA-binding activity, dissociation by E1A of E2F/Rb complexes results in DNA binding and possibly transcriptional activation (Bagchi et al., 1991; Hiebert et al., 1992). A 19-kDa repressor protein (Dr) bound to TBP has recently been purified and its encoding sequences isolated (Inostroza and Reinberg, 1991). Properties of the negative component of USA (NC1; Meisterernst et al., 1991) appear similar to the negative regulator Dr. By analogy, dissociation of this Dr transcriptional inhibitor from TFIID by E1A may correlate with E1A-mediated activation via TATA boxes (see discussion in Horikoshi et al., 1991).

Transcription initiation factor (Inr or TFII I)

The TATA motif is not the only promoter element required for the assembly of transcription competent complexes. Natural promoters without apparent TATA motifs or mutations which alter the sequence reduce but do not completely

obliterate transcription. The sequences around the start site and their importance as promoter elements were first pointed out for the terminal deoxynucleotidyl-transferase initiator (Inr; Smale and Baltimore, 1989). Similar motifs with a common $^{CTCA}_{GAGT}$ sequence (where the A is the transcriptional start) have been described for the adenovirus major late and IVa2 promoters (Carcamo et al., 1991).

The initiator protein bound to DNA can substitute for the TATA box as nucleation center for the transcription initiation complex (Carcamo et al., 1991). The DNA-bound initiator element is directly and weakly recognized by RNA polymerase II. This process is greatly enhanced by transcription factors IID, IIB, and IIF, which selectively stimulate correct start site utilization (Carcamo et al., 1991). It appears that strong initiators are required with weak TATA boxes, while strong TATA boxes tolerate weak initiators. Thus, cooperative interactions between the two elements appear important (Lee et al., 1988; Carcamo et al., 1991). Moreover, the initiator motif alone is able to nucleate all the other transcription factors around itself in a complex containing IID, IIB, RNA polymerase II, and IIF in the absence of a TATA box (Carcamo et el., 1991). This explains the TBP requirement of TATA-less promoters, since TBP can be recruited to the promoter by the other proteins of the initiation complex (Pugh and Tjian, 1991). The gene encoding the sequence for the transcription factor YY-1, a Krüppel-like 55-kDa polypeptide with four Zn^{++} fingers, was recently cloned by several laboratories (Shi et al., 1991; Hariharan et al., 1991; Park and Atchison, 1991; Flanagan et al., 1991). This protein is able to function in an in vitro transcription system as an initiator, either alone or in the presence of transcription factor SP1 (Seto et al., 1991). A novel 120-kDa helix-loop-helix initiator protein, which binds at the initiation site of the adenovirus major late promoter and is able to interact functionally with the upstream factor USF, has been described by Roy et al. (1991) as TFII I.

Transcription factor IIA (TFIIA)

The visualization of the sequential steps in transcription complex assembly was greatly aided by the development of a gel mobility shift assay in which sequential binding of the transcription factors could be readily followed (Buratowski et al., 1989). The presence of TFIIA is required for formation of gel-shifted complexes between TFIID and promoter DNA, but this requirement can be overcome in the presence of excess Mg^{++} (Buratowski et al., 1989; Horikoshi et al., 1989). Some report transcription factor IIA as a 43-kDa polypeptide (Egly et al., 1984), or as a 13- and 17-kDa protein complex (Samuels and Sharp, 1986); others have purified a TFIIA 35-kDa active protein from wheat (Burke et al., 1990). The protein purified from calf thymus has full activity in the in vitro system (Samuels and Sharp, 1986). Using eight chromatographic steps to purify this TFIIA factor, it appears that it contains two activities (TFIIA and TFIIJ; Cortes et al., 1992). The TFIIA activity was further purified by bacterially produced TDB-affinity chromatography and contains polypeptides of 35, 19, and 14 kDa, corresponding to the previously described TFIIA activity (Samuels and Sharp, 1986; Cortes et al., 1992). It stimulates transcription but is not required when TBP from recombinant sources is used in reconstitution experiments. In contrast, TFIIJ is required and apparently acts at a later stage in the initiation process (step 6; Cortes et al., 1992). A more complete understanding of the role of the different forms of TFIIA and TFIIJ will have to await characterization and cloning of the encoding genes, and further study of the proteins. The requirement of TFIIA for the formation of the TBP promoter complex appears controversial, and depends on the source of TBP and binding conditions.

Transcription factor IIB (TFIIB)

After the TFIID–TFIIA complex is formed on promoter DNA, the factor TFIIB enters into the transcription complex (step 3). TFIIA stimulates the formation of the DAB complex significantly if D and B alone are provided but does not affect later stages (Maldonado et al., 1990). The TFIIB 33-kDa protein appears distinct from the recently reported BTF-3 (Zheng et al., 1990). The gene encoding TFIIB has been cloned by Ha et al. (1991) and Malik et al. (1991). It contains two 76 amino acid repeats and exhibits some homology with bacterial σ factors. Evidence for interaction of VP16-derived acidic activators with TFIIB has been recently presented (Lin and Green, 1991; Lin et al., 1991). This interaction appears to be stronger than the VP16-TBP interaction, in contrast to what occurs with E1A (see above).

Transcription factor IIF (TFIIF)

The use of more purified factors in transcription reconstitution has permitted the identification of other activities in the phosphocellulose fraction eluting at 0.5 M KCl: TFIIF and TFIIE (Reinberg and Roeder, 1987). The two-subunit transcription factor TFIIF is probably the same as RAP 74–30 kDa, a bimolecular complex binding to RNA polymerase II (Sopta et al., 1985; Flores et al., 1988; Burton et al., 1988), and also possibly similar to βγ (Conaway et al., 1989) and FC (Kitajima et al., 1990). The native protein sediments as a dimer of 220 kDa, containing two subunits of 30 kDa and two subunits of 78 kDa (Flores et al., 1990). The latter is phosphorylated in vitro by casein kinase II (O. Flores, R. Weinmann, and D. Reinberg, unpublished data). The role of TFIIF is to mediate the entry of RNA polymerase IIa into the complex of TFIID, TFIIB, and promoter DNA (Flores et al., 1991; step 3). Moreover, it appears that the small subunit of TFIIF is sufficient to insure entry of RNA polymerase II into the transcriptional complex (Flores et al., 1991). It is interesting to note that the sequence homology found between the 30-kDa subunit of TFIIF and the Escherichia coli σ 70 protein (Sopta et al., 1989) may thus extend to their functional role in transcription (Flores et al., 1991). Although some reports suggested that RAP 74 may be a helicase, no ATP-dependent helicase activity or ATPase has been detected in extensively purified human TFIIF (Flores et al., 1990). The final word will come from recombinant proteins. The small subunit of this two-subunit protein has been cloned by Sopta et al. (1989). The large subunit has recently been cloned by Aso et al. (1992), and by Finkelstein et al. (1992). The recombinant large and small TFIIF subunits associate and stimulate transcriptional activity of depleted extract, but lack ATPase or helicase activities (Aso et al., 1992; Finkelstein et al., 1992).

Transcription factor IIE (TFIIE)

TFIIE, also purified from phosphocellulose fraction eluting at 0.5 M KCl, can be separated from TFIIF on DEAE-5PW columns (Flores et al., 1990). TFIIE is also composed of two subunits, one 56 kDa and one 34 kDa, and behaves in sizing columns as a 200-kDa tetramer (Inostroza et al.,

1991). The similar molecular weight of the polypeptides of the transcription factors TFIIE and TFIIF has generated a great deal of confusion. A factor purified from rat liver with activities equivalent to TFIIE is ε (Conaway et al., 1991b). The two subunits of the TFIIE protein have been recently cloned (M. G. Peterson et al., 1991; Ohkuma et al., 1991; Sumimoto et al., 1991). The 34-kDa subunit contains a consensus nucleotide binding site, and the 56-kDa contains a pocket with a sequence similar to all kinases, in addition to a Zn finger domain possibly involved in DNA binding (Peterson et al., 1991). Although an ATPase activity was found associated with TFIIE, it is apparently absent in TFIIE preparations purified to homogeneity (Inostroza et al., 1991). The purification of this factor was greatly facilitated by the finding that it leaks to the cytoplasmic fraction (S100) during extract preparation.

Other transcription factors

The recently described TFIIG protein fraction (Sumimoto et al., 1990) is separated from the TFIID fraction on further purification. Because different laboratories use complementary fractions with varying degrees of purification, the exact nature of this factor will require purification to homogeneity to ascertain whether it is indeed a new factor or part of others previously described.

Recent purification results of Flores and Reinberg (1991) and Cortes et al. (1992) suggest that the fraction described as TFIIG may be composed of other proteins. After chromatography on SP-Sepharose, S-Sepharose, and phenyl-Superose, two activities, called TFIIJ and TFIIH, were separated (Flores and Reinberg, 1991). TFIIH could also be purified from the TFIIF fraction, while TFIIJ also co-purified through seven steps with TFIIA (Flores and Reinberg, 1991). TFIIH activity co-purified with polypeptides of 33, 40, 42, 60, and 90 kDa, while TFIIJ has not yet been purified to homogeneity (Flores and Reinberg, 1991). Both factors enter the transcription complex after DBPolFE are assembled on the promoter (steps 6 and 7 in the diagram). TFIIJ was required only when bacterially produced TBP was used, but was dispensable with TFIID purified from cells (Cortes et al., 1991). The small subunit of TFIIF or TFIIH does not cross-react

with anti-BTF3 antibodies (Flores and Reinberg, 1990). BTF3 is a 30-kDa polypeptide which binds to RNA polymerase II and is required in the transcription system from which it was purified and cloned (Zheng et al., 1990). Transcription factor BTF2 is also composed of 5 subunits of 90, 60, 43, 41, and 35 kDa and appears homologous to TFIIH and the rat liver δ (Gerard et al., 1991). The kinase involved in conversion of RNA polymerase IIa to IIo in the transcription complex appears to be associated with TFIIH (D. Reinberg, personal communication). A δ factor isolated from rat liver consists of at least 6 polypeptides and has a native molecular mass of 390 kDa (Conaway et al., 1989, 1991b). It is interesting to note that rat liver δ also contains polypeptides of 95, 85, 68, 46, 43, 38, and 35 kDa (Conaway et al., 1991b), which may be related to TFIIH. The native molecular weight of this δ complex and its polypeptide composition also appear to be related to the TFIID complex found in Drosophila (Dynlacht et al., 1991). This liver transcription factor has an ATPase activity associated with TATA box DNA binding (Conaway et al., 1989). The exact relationship between BTF2, the δ multisubunit complex, TFIIG, TFIIH, and TFIIJ remains to be determined.

RNA polymerase II

This enzyme, well characterized and purified by different laboratories, is at the moment the reagent which even within the next few years will continue to be purified from cells or tissues. For recent reviews see Sawadogo and Sentenac (1990), Corden (1990), Woychik and Young (1990), and Young (1991). This is a large enzyme with 10 subunits, the largest of which is 215 kDa in size. Most subunits of the yeast RNA polymerase II enzyme have been cloned by the laboratories of Sentenac and Young (see Sawadogo and Sentenac, 1990; Kolodziej et al., 1990; Young, 1991). Some subunits of the human enzyme have also been cloned recently (Saltzman and Weinmann, 1989; Pati and Weissman, 1989, 1990). Although most of the subunits may be cloned in the next few years, the reassembly of active enzyme from these components seems to present a difficult problem. Two-dimensional crystals of yeast RNA polymerase II are powerful analytical tools that have been used to obtain structural information at a resolution of 16Å (Edwards et al., 1990; Darst

et al., 1991). Structural analysis has revealed a 25Å-diameter channel through which the DNA may be threaded. This major 25Å-wide groove then bifurcates into a narrow groove, where single-stranded nucleic acids may be interacting (nascent RNA; Darst et al., 1991). A finger of protein projecting outward from the structure may represent the CTD (Darst et al., 1991).

Phosphorylation of the CTD of RNA polymerase II

Three forms of RNA polymerase II are normally purified from tissues: IIo, IIa, and IIb. These differ in the apparent size of the largest subunit, which migrates in gels as 240 kDa (IIo), 215 kDa (IIa), and 180 kDa (IIb). From the results of Corden et al. (1985), it is clear that IIb is a proteolyzed form of IIa, lacking the CTD multiheptapeptide repeated sequence of the large subunit. A unique feature of the 215 kDa polypeptide is the presence at the CTD of a heptapeptide which is repeated 52 times in the mouse enzyme (Ahearn et al., 1987) and 27 times in yeast, with variable numbers in other organisms (reviewed in Corden, 1990). The IIo form of RNA polymerase differs from IIa in the phosphorylation of this multiheptapeptide repeated sequence and is the only form found to crosslink with RNA in transcription complexes (Bartholomew et al., 1986). Although the IIb proteolyzed form has been discounted as a contributor to transcription in vivo, the transcriptional activity of IIb in an in vitro system (Zehring et al., 1988; Kim and Dahmus, 1989) suggests that this problem has to be reexamined. Pure RNA polymerase can be efficiently obtained by using nuclear pellets that are byproducts of the preparation of nuclear extracts (Lu et al., 1991). In the in vitro transcription system, reconstituted with highly purified factors devoid of endogenous phosphatase or kinase activities, the IIa form is 8-10 times more active in specific initiation than the IIo form, and forms stable complexes with promoter DNA, TFIID, TFIIB, and TFIIF (Lu et al., 1991).

The interconversion of the two forms apparently occurs in the transition from initiation to elongation (Laybourn and Dahmus, 1989, 1990; Payne et al., 1989; Lu et al., 1991; step 5 in the diagram) and is accompanied by phosphorylation of the CTD. The kinase responsible for phosphorylation remains to be identified.

We have established an ATP-hydrolysis requirement (Bunick et al., 1982) and a 5,6-diCl-1-β-D ribofuranosyl-benzimidazole (DRB)-sensitive process (Zandomeni et al., 1983) before formation of the first dinucleotide bond (step 6). The ATP hydrolysis suggested involvement of a kinase, a DNA-dependent gyrase, or an ATPase (Bunick et al., 1982), as did the identification of DRB sensitivity at an early transcriptional step (Zandomeni et al., 1983). We have identified casein kinase II as a DRB-sensitive kinase (Zandomeni and Weinmann, 1984; Zandomeni et al., 1986), although only a single phosphorylation site for this enzyme is present in the unique region of the CTD (Corden, 1990). It is important to establish which kinase(s) is responsible for the phosphorylation of the large subunit associated with the enzyme in the transcription complexes. One of the questions that remains unanswered is the nature of the ATP-hydrolysis requirement at transcription initiation (Bunick et al., 1982).

The steps in the formation of the *initiation* complexes have been defined by their sensitivities to sarkosyl, heparin, and salt, and the challenge by second templates (see Saltzman and Weinmann, 1989, for references). Clarification of the order of this process was recently provided by the analysis of the assembly intermediates of these complexes in a native gel system (Buratowski et al., 1989). The phosphorylation step required to pass from the preinitiation to the initiation complex results in the conversion of the RpIIa to RpIIo. The CTD domain of RpII contains 52 heptamer repeats, and it has been reported that from 15 to 200 phosphates are added to the CTD of RPIIa (Zhang and Corden, 1991; Arias et al., 1991). This amount of incorporated phosphate may suffice to produce the mobility shift to IIo and may explain, at least in part, the ATP requirement at initiation (steps 4–5). Since RNA polymerase IIa, IIo, and IIb have similar ATP hydrolysis requirements for transcription initiation, in addition to the phosphorylation of the CTD, other ATP-requiring steps may be involved (Laybourn and Dahmus, 1990). Several kinases have been implicated in this step of the initiation process, and it is not yet clear which one is involved. We have provided evidence for casein kinase II or a DRB-sensitive kinase (Zandomeni et al., 1986), while others suggest that a cell-cycle related kinase is involved in this process (Cisek and Corden, 1989). Interestingly, these latter kinases are also sensitive to DRB (Cisek and Corden, 1991). Kinases purified from Aspergillus (Stone and Reinberg, 1992) and yeast (Lee and Greenleaf, 1989) appear as single-subunit of 57 kDa or three-subunit enzymes of 58 kDa, 38 kDa, and 32 kDa. The large subunit of the yeast CTD kinase (CTK1) has been cloned and is required for normal growth of yeast (Lee and Greenleaf, 1991). One of the mammalian kinases recently purified contains polypeptides of 58 and 34 kDa. Its small subunit is related to the CDC2-CDC28 cell cycle genes of yeast, and the activity of the enzyme is sensitive to DRB (Cisek and Corden, 1989, 1991). A variant enzyme contains subunits of 62 kDa (equivalent to cyclin B) and 34 kDa (cdc2), implying that it is a major cell-cycle specific kinase (Cisek and Corden, 1991). The enzyme purified from yeast appears to be quite distinct in substrate specificity, amino acid sequence, and chromatographic properties (Lee and Greenleaf, 1989). An enzyme which is enhanced under heat-shock conditions and able to phosphorylate a CTD peptide appears to be more similar to the yeast-Aspergillus enzyme (Legagneux et al., 1990). A template-associated kinase has been implicated in CTD phosphorylation, but the enzyme has not been purified yet (Arias et al., 1991).

The substrate for the RpII-kinase able to efficiently convert the RNA polymerase IIa to IIo in the transcription complex can be either ATP, dATP, or GTP (Lu et al., 1991). DRB acts as a mixed-type inhibitor of casein kinase II (which can also use ATP, GTP, or dATP as substrates) and of in vivo and in vitro transcription at similar DRB concentrations (Zandomeni et al., 1986; Zandomeni, 1989). Thus, the kinase involved in phosphorylation of RpII would also be sensitive to DRB and, if different from CKII, sequence-related in the active site. Finally, phosphorylation by one kinase may be required to induce changes that allow phosphorylation by a second kinase. The difficulty here is that these enzymes act catalytically rather than stoichiometrically, like the transcription factors described above. This problem will be resolved when more purified components or recombinant proteins are used in transcription reconstitution. Phosphorylation of the CTD and utilization of up to 200 molecules of ATP per initiation event do not exclude the possibility that other ATP-dependent processes, such as unwinding, or other ATPase and kinase activities, are associated with the transcription initiation process.

Future directions

The availability of an increasing number of mammalian transcription factors from recombinant sources will greatly facilitate the purification, characterization, and cloning of the remaining transcription factors. Although studies have focused on the adenovirus major late promoter as a template and the HeLa cell as a source of proteins, application to other promoters and systems will permit fuller understanding of the principles of promoter selection, the mechanics of transcription initiation, and the effects of regulators on this process.

Other steps in the transcriptional process, such as elongation (which can be regulated by multiple factors, including TFIIS or SII and TFIIF), may play an important role in the control of gene expression. Moreover, it seems that some elements of the transcriptional machinery may be reutilized directly, because they remain bound to the promoter after initiation of RNA synthesis, like TFIID or TFIIA. Others, like the RpIIo, have to be dephosphorylated so that the RpIIa can re-enter the transcription process, while the fate of such factors as IIE, IIG, IIH, and IIJ is unknown. Little is known about how this recycling may occur and which enzymes are involved. The tools provided by studying the basal transcriptional machinery will also help to clarify the multiple ways in which promoter activity is regulated, as well as the multiple sites of regulation used by different effectors. Finally, we have addressed only those proteins directly interacting with naked DNA. In higher eukaryotes, cellular DNA is organized into chromatin structures, and these are also important determinants of gene expression (for a recent review, see Felsenfeld, 1992).

References

J. M. Ahearn, M. S. Bartolomei, M. L. West, L. J. Cisek, and J. L. Corden (1987), J Biol Chem 265, 425–437.

J. A. Arias, S. R. Peterson, and W. S. Dynan (1991), J Biol Chem 266, 8055–8061.

T. Aso, H. A. Vasavada, T. Kawaguchi, F. J. Germino, S. Ganguly, S. Kitajima, S. Weissman, and Y. Yasukochi (1992), Nature 355, 461–464.

S. Bagchi, R. Weinmann, and P. Raychaudhuri (1991), Cell 65, 1063–1072.

B. Bartholomew, M. E. Dahmus, and C. F. Meares (1986), J Biol Chem 261, 14226–14231.

D. Bunick, R. Zandomeni, S. Ackerman, and R. Weinmann (1982), Cell 29, 877–886.

S. Buratowski, S. Hahn, P. A. Sharp, and L. Guarente, (1988), Nature 334, 37–42.

S. Buratowski, S. Hahn, L. Guarente, and P. A. Sharp (1989), Cell 56, 549–561.

C. Burke, Y. Xu-Bu, L. Marchitell, E. A. Davis, and S. Ackerman (1990), Nucl Acids Res 18, 3611–3620.

Z. F. Burton, M. Killeen, M. Sopta, L. G. Ortolan, and J. Greenblatt (1988), Mol Cell Biol 8, 1602–1613.

J. Carcamo, S. Lobos, A. Merino, L. Buckbinder, R. Weinmann, V. Natarajan, and D. Reinberg (1989), J Biol Chem 264, 7704–7714.

J. Carcamo, E. Maldonado, P. Cortes, M.-H. Ahn, I. Ha, Y. Kasai, J. Flint, and D. Reinberg (1990), Genes Dev 4, 1611–1622.

J. Carcamo, L. Buckbinder, and D. Reinberg (1991), Proc Natl Acad Sci USA 88, 8052–8056.

B. Cavallini, I. Fans, H. Mathes, J. M. Chipoulet, J. M. Egly, and P. Chambon (1989), Proc Natl Acad Sci USA 86, 9803–9807.

L. J. Cisek and J. L. Corden (1989), Nature 339, 679–684.

L. J. Cisek and J. L. Corden (1991), Methods Enzymol 200, 301–324.

J. W. Conaway and R. C. Conaway (1989), J Biol Chem 264, 2357–2362.

J. W. Conaway, J. P. Hanley, K. P. Garrett, and R. C. Conaway (1991a), J Biol Chem 266, 7804–7811.

R. C. Conaway, K. P. Garrett, J. P. Hanley, and J. W. Conaway (1991b), Proc Natl Acad Sci USA 88, 6205–6209.

J. Corden, D. L. Cadena, J. M. Ahearn, and M. E. Dahmus (1985), Proc Natl Acad Sci USA 82, 7934–7938.

J. L. Corden (1990), Trends Biochem Sci 15, 387–389.

P. Cortes, O. Flores, and D. Reinberg (1992), Mol Cell Biol 12, 413–421.

S. A. Darst, A. M. Edwards, E. W. Kubalek, and R. D. Kornberg (1991), Cell 66, 121–128.

B. L. Davison, J. M. Egly, E. R. Mulvihill, and P. Chambon (1983), Nature 301, 680–686.

B. D. Dynlacht, T. Hoey, and R. Tjian (1991), Cell 66, 563–576.

A. M. Edwards, S. A. Darst, W. J. Feaver, N. E. Thompson, R. R. Burgess, and R. D. Kornberg (1990), Proc Natl Acad Sci USA 87, 2122–2126.

J. M. Egly, N. G. Miyamoto, V. Moncollin, and P. Chambon (1984), EMBO J 3, 2363–2371.

D. M. Eisenmann, C. Dollard, and R. Winston (1989), Cell 58, 1183–1191.

G. Felsenfeld (1992), Nature 355, 219–224.

J. D. Fikes, D. M. Becker, F. Winston, and L. Guarente (1990), Nature 346, 291–294.

A. Finkelstein, C. F. Kostrub, J. Li, D. P. Chavez, B. Q. Wang, S. M. Fang, J. Greenblatt, and Z. F. Burton (1992), Nature 355, 464–467.

A. Fire, S. Samuels, and P. A. Sharp (1984), J Biol Chem 259, 2509–2516.

P. M. Flanagan, R. J. Kelleher, III, M. H. Sayre, H. Tschochner, and R. D. Kornberg (1991), Nature 350, 436–438.

O. Flores, I. Ha, and D. Reinberg (1990), J Biol Chem 265, 5629–5634.

O. Flores, H. Lu, M. Killen, J. Greenblatt, Z. F. Burton, and D. Reinberg (1991), Proc Natl Acad Sci USA 88, 9999–10003.

O. Flores, E. Maldonado, Z. Burton, J. Greenblatt, and D. Reinberg (1988), J Biol Chem 263, 10812–10816.

O. Flores, E. Maldonado, and D. Reinberg (1989), J Biol Chem 264, 8913–8921.

A. Gash, A. Hoffman, M. Horikoshi, R. G. Roeder, and N.-H. Chua (1990), Nature 346, 390–394.

M. Gerard, L. Fischer, V. Moncollin, J.-M. Chipoulet, P. Chambon, and J.-M. Egly (1991), J Biol Chem 226, 20940–20945.

G. Gill and R. Tjian (1991), Cell 65, 333–340.

P. D. Gregor, M. Sawadogo, and R. G. Roeder (1990), Genes Dev 4, 1730–1740.

I. Ha, W. J. Lane, and D. Reinberg (1991), Nature 352, 689–695.

S. Hahn, S. Buratowski, P. A. Sharp, and L. Guarente (1989), Cell 58, 1173–1181.

N. Hariharan, D. E. Kelley, and R. P. Perry (1991), Proc Natl Acad Sci USA 88, 9799–9803.

S. W. Hiebert, S. P. Chellappan, J. M. Horowitz, and J. R. Nevins (1992), Genes Dev 6, 177–185.

T. Hoey, B. D. Dynlacht, M. G. Peterson, B. F. Pugh, and R. Tjian (1990), Cell 61, 1179–1186.

A. Hoffman, E. Sinn, T. Yamamoto, J. Wang, A. Roy, M. Horikoshi, and R. G. Roeder (1990), Nature 346, 387–390.

M. Horikoshi, T. Hai, Y.-S Lin, M. R. Green, and R. G. Roeder (1988), Cell 54, 1033–1042.

M. Horikoshi, C. K. Wang, H. Fuji, J. A. Cromlish, P. A. Weil, and R. G. Roeder (1989), Nature 341, 299–303.

N. Horikoshi, K. Maguire, A. Kralli, E. Maldonado, D. Reinberg, and R. Weinmann (1991), Proc Natl Acad Sci USA 88, 5124–5128.

M. Horikoshi, C. Beruccioli, R. Takeda, J. Wang, T. Yamamoto, and R. G. Roeder (1992), Proc Natl Acad Sci USA 89, 1060–1064.

C. J. Ingles, M. Shales, W. D. Cress, S. J. Triezenberg, and J. Greenblatt (1991), Nature 351, 588–590.

J. Inostroza, O. Flores, and D. Reinberg (1991), J Biol Chem 266, 9304–9308.

J. A. Jaehning (1991), Science 253, 859.

C. C. Kao, P. M. Lieberman, M. C. Schmidt, Q. Zhou, R. Pei, and A. J. Berk (1990), Science 248, 1646–1650.

R. J. Kelleher, III, P. M. Flanagan, and R. D. Kornberg (1990), Cell 61, 1209–1215.

W. Y. Kim and M. E. Dahmus (1989), J Biol Chem 264, 3169–3176.

S. Kitajima, Y. Tanaka, T. Kawaguchi, T. Nagaska,

S. M. Weissman, and Y. Yasukochi (1990), Nucl Acids Res 18, 4843–4849.

P. A. Kolodziej, N. Woychik, S.-M. Liao, and R. A. Young (1990), Mol Cell Biol 10, 1915–1920.

P. J. Laybourn and M. E. Dahmus (1989), J Biol Chem 264, 6693–6698.

P. J. Laybourn and M. E. Dahmus (1990), J Biol Chem 265, 13165–13173.

J. M. Lee and A. L. Greenleaf (1989), Proc Natl Acad Sci USA 86, 3624–3628.

J. M. Lee and A. L. Greenleaf (1991), Gene Expr 1, 149–167.

R. F. Lee, M. F. Concino, and R. Weinmann (1988), Virology 165, 51–56.

W. S. Lee, C. C. Kao, G. O. Bryant, X. Liu, and A. J. Berk (1991), Cell 67, 365–376.

V. Legagneux, M. Morange, and O. Besaude (1990), Eur J Biochem 193, 121–126.

B. Lewin (1990), Cell 61, 1161–1164.

Y.-S. Lin and M. R. Green (1991), Cell 64, 971–981.

Y.-S. Lin, I. Ha, E. Maldonado, D. Reinberg, and M. R. Green (1991), Nature 353, 569–571.

H. Lu, O. Flores, R. Weinmann, and D. Reinberg (1991), Proc Natl Acad Sci USA 88, 10004–10008.

D. S. Luse and G. A. Jacob (1987), J Biol Chem 262, 14990–14997.

E. Maldonado, I. Ha, P. Cortes, L. Weis, and D. Reinberg (1990), Mol Cell Biol 10, 6335–6347.

S. Malik, K. Hisatake, H. Sumimoto, M. Horikoshi, and R. G. Roeder (1991), Proc Natl Acad Sci USA 88, 9553–9557.

T. Matsui, J. Segall, A. P. Weil, and R. G. Roeder (1980), J Biol Chem 255, 1192–1196.

B. P. McCormack, M. Strubin, A. S. Ponticelli, and S. K. Struhl (1991), Cell 65, 341–348.

M. Meisterernst, A. L. Roy, H. M. Liev, and R. G. Roeder (1991), Cell 66, 981–993.

F. H. Mermelstein, O. Flores, and D. Reinberg (1989), Biochim Biophys Acta 1009, 1–10.

N. Nakajima, M. Horikoshi, and R. Roeder (1988), Mol Cell Biol 8, 4028–4040.

Y. Ohkuma, H. Sumimoto, A. Hoffman, S. Shimasaki, M. Horikoshi, and R. G. Roeder (1991), Nature 354, 401–404.

K. Park and M. Atchison (1991), Proc Natl Acad Sci USA 88, 9804–9808.

U. K. Pati and S. M. Weissman (1989), J Biol Chem 264, 13114–13121.

U. K. Pati and S. M. Weissman (1990), J Biol Chem 265, 8400–8403.

J. M. Payne, P. J. Laybourn, and M. E. Dahmus (1989), J Biol Chem 264, 19621–19629.

M. G. Peterson, J. Inostroza, M. E. Maxon, O. Flores, A. Admon, D. Reinberg, and R. Tjian (1991), Nature 354, 369–373.

M. G. Peterson, N. Tanese, B. F. Pugh, and R. Tjian (1990), Science 248, 1625–1630.

B. F. Pugh and R. Tjian (1990), Cell 61, 1187–1197.

B. F. Pugh and R. Tjian (1991), Genes Dev 5, 1935–1945.

J. Rappaport, K. Cho, A. Saltzman, J. Prenger, M. Golomb, and R. Weinmann (1988), Mol Cell Biol 8, 3136–3142.

J. Rappaport, R. Zandomeni, D. Reinberg, and R. Weinmann (1987), J Biol Chem 262, 5227–5232.

D. Reinberg, M. Horikoshi, and R. G. Roeder (1987), J Biol Chem 262, 3322–3330.

D. Reinberg and R. G. Roeder (1987), J Biol Chem 262, 3331–3337.

O. Resnekov and Y. Aloni (1989), Proc Natl Acad Sci USA 86, 12–16.

A. E. Rougvie and J. T. Lis (1988), Cell 54, 795–804.

A. L. Roy, M. Meisterernst, P. Pognonec, and R. G. Roeder (1991), Nature 354, 245–248.

A. Saltzman and R. Weinmann (1989), FASEB J 3, 1723–1728.

M. Samuels and P. A. Sharp (1986), J Biol Chem 261, 2003–2013.

M. Sawadogo and R. G. Roeder (1985), Cell 43, 165–175.

M. Sawadogo and A. Sentenac (1990), Annu Rev Biochem 59, 711–754.

M. C. Schmidt, C. C. Kao, R. Pei, and A. J. Berk (1989), Proc Natl Acad Sci USA 86, 7785–7789.

E. Seto, Y. Shi, and T. Shenk (1991), Nature 354, 241–245.

Y. Shi, E. Seto, L. Chang, and T. Shenk (1991), Cell 67, 377–388.

E. Slattery, J. D. Dignam, T. Matsui, and R. G. Roeder (1983), J Biol Chem 258, 5955–5959.

S. T. Smale and D. Baltimore (1989), Cell 57, 103–113.

M. Sopta, Z. F. Burton, and J. Greenblatt (1989), Nature 341, 410–414.

M. Sopta, R. W. Carthew, and J. Greenblatt (1985), J Biol Chem 260, 10353–10360.

A. Stevens and M. K. Maupin (1989), Biochem Biophys Res Commun 159, 508–515.

N. Stone, O. Flores, and D. Reinberg (1991), Pharm Technol 15, 36–46 and 128.

N. Stone and D. Reinberg (1992), J Biol Chem 267, 6353–6360.

F. F. Stringer, C. J. Ingles, and J. Greenblatt (1990), Nature 345, 783–785.

M. Strubin and K. Struhl (1992), Cell 68, 721–730.

H. Sumimoto, Y. Ohkuma, T. Yamamoto, M. Horikoshi, and R. G. Roeder (1990), Proc Natl Acad Sci USA 87, 9158–9162.

H. Sumimoto, Y. Ohkuma, E. Sinn, H. Kato, S. Shimasaki, M. Horikoshi, and R. G. Roeder (1991), Nature 354, 398–401.

M. W. Van Dyke, R. G. Roeder, and M. Sawadogo (1988), Science 241, 1335–1338.

J. L. Workman and R. G. Roeder (1987), Cell 51, 613–630.

N. A. Woychik and R. A. Young (1990), Trends Biochem Sci 15, 347–351.

R. A. Young (1991), Annu Rev Biochem 60, 689–715.

R. Zandomeni, D. Bunick, S. Ackerman, B. S. Mittleman, and R. Weinmann (1983), J Mol Biol 167, 561–574.

R. Zandomeni and R. Weinmann (1984), J Biol Chem 259, 14804–14811.

R. Zandomeni, M. C. Zandomeni, D. Shugar, and R. Weinmann (1986), J Biol Chem 261, 3414–3419.

J. Zhang and J. L. Corden (1991), J Biol Chem 266, 2290–2296.

X. M. Zheng, D. Black, P. Chambon, and J. M. Egly (1990), Nature 344, 556–559.

W. A. Zehring, J. M. Lee, J. R. Weeks, R. S. Jokerst, and A. L. Greenleaf (1988), Proc Natl Acad Sci USA 85, 3698–3702.