This is to certify that the

dissertation entitled

Functions of Transcription Factor IIF
in Initiation and Elongation by RNA Polymerase II

presented by

Chun-hsiang Chang

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Biochemistry

Major professor

Date 6/9/95

FUNCTIONS OF TRANSCRIPTION FACTOR IIF IN INITIATION AND ELONGATION BY RNA POLYMERASE II

BY

Chun-hsiang Chang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1995

ABSTRACT

FUNCTIONS OF TRANSCRIPTION FACTOR IIF IN INITIATION AND ELONGATION BY RNA POLYMERASE II

BY

Chun-hsiang Chang

Understanding RNA polymerase II-dependent transcription requires characterization of basal and regulatory factors and their participation in transitions between stages of the transcription cycle. This study focuses on functions of general transcription factor TFIIF and the transition from initiation to productive elongation. A role for the large TFIIF subunit (named RAP74 for RNA polymerase II-associated protein; 74 kilodaltons) in this transition is demonstrated. An assay for detection of the first stable transcription complexes initiated from the Adenovirus major late promoter has been developed and applied to understand functions of RAP74 in maintenance of short transcript stability.

Both RAP30 and RAP74 subunits of TFIIF are required for accurate transcription in vitro. Here we demonstrate that in some in vitro systems the RAP30 subunit is required for accurate initiation, but the RAP74 subunit is not. RAP74, however, is required for transition to a productive elongation complex. Thus, the two subunits of TFIIF have partially separable functions in initiation and early elongation of transcription.

To better understand the transition between initiation and elongation by RNA polymerase II, an assay system was designed in which the Adenovirus major late

promoter was immobilized on agarose beads. The first stable transcripts initiated on these immobilized templates were isolated and analyzed. A kinetic lag of 15 to 20 s was observed before stable transcripts appeared when nucleoside triphosphates were added to pre-formed transcription complexes. Depending on the nucleoside triphosphate that is limiting in concentration, RNA polymerase II pauses at various positions between nucleotide +11 to +20 of the RNA chain. Transcripts shorter than 11 nucleotides in length are not stably bound to the beads, and must be released as abortive transcripts. Short, paused RNAs are stable to washing with buffer and can be quantitatively chased to the runoff position on the template.

The stability of short RNA complexes is compromised if transcription is initiated in the presence of certain RAP74 mutants. Human RAP74 is a 517 amino acid protein. The C-terminal region of RAP74, including amino acids 409-517, contributes to short transcript stability. A 1-172 mutant, previously shown in another assay to be inactive for accurate transcription, is shown here to support accurate intiation. Transcripts initiated in the presence of this mutant are very unstable, compared to a 1-205 mutant. The role of RAP74 sequences in stabilizing short ternary complexes may partially explain the role of RAP74 in promoter escape.

In the final chapter of this thesis, preliminary experiments are shown in which a yeast protein Cdc73p is demonstrated to bind directly to RNA polymerase II. Cdc73p is of interest because from sequence analysis this protein appears to be a subunit of an alternate form of TFIIF. Cdc73p has similar structures to human RAP30, bacterial sigma factors, and bacterial delta factor. These similarities were used to predict the region of Cdc73p that might bind to RNA polymerase II. Deletion of 15 amino acids from this region severely inhibited polymerase binding.

Dedicated to my friends, my teachers, and my family.

ACKNOWLEDGMENTS

I would like to express my appreciation to my mentor, Dr. Zachary Burton, for his guidance ,support and encouragement through the years. I wish to thank members of my guidance committee, Drs. William Helferich, Rawle Hollingsworth, Arnold Revzin and William Smith for their advise and encouragement.

I would like to acknowledge the support of my colleagues, Dr. Ann Finkelstein, Dr. Wladyslaw Werel, David Chavez, Shimin Fang, Richard Fentzke, Jenny Jensen, Corwin Kostrub, Lei Lei, Augie Pioszak, Stephan Reymez, Bo Qing Wang, and Yong Wang. Especially, I want to thank Cory and Bo Qing for the recombinant RAP30 and RAP74 proteins that they have provided in my study.

Finally , I am grateful to my parents for everything they have given to me and to my wife for her understanding and unconditional support.

TABLE OF CONTENTS

LIST OF FIGURES

PAGE

Chapter IV

LIST OF ABBREVIATIONS

α-A	α-amanitin
A_{280}	absorbance at 280 nm
A_{600}	absorbance at 600 nm
aa	amino acid(s)
AdMLP	Adenovirus major late promoter
AN	AMPPNP
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
CTP	cytosine triphosphate
Da	Dalton
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
EGTA	ethyleneglycol-bis-(β-aminoethyl ether) N,N,N',N'-tetraacetic acid
FPLC	fast protein liquid chromatography
GTP	guanine triphosphate
HCA	hydrophobic cluster analysis
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
IPTG	isopropyl-β-D-thiolgalactopyranoside
kb	kilobase pairs
kDa	kilodaltons
LB	Luria-Bertani
mG	3'-O-methyl GTP
mRNA	messenger RNA

NE	HeLa nuclear extract
nt	nucleotides
NTP	nucleoside triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RNAP II	RNA polymerase II
RAP	RNA polymerase II associating proteins
RNA	ribonucleic acid
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
TAF	TBP associated factor
TBP	TATA box-binding protein
TFII	transcription factor of RNA polymerase II
tRNA	transfer RNA

Single letter abbreviations for the amino acids: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

CHAPTER I

INTRODUCTION

2
LITERATURE REVIEW

Regulation of transcription is a critical stage of differentiation and development in biological systems. A single form of RNA polymerase performs transcription of ribosomal RNA (rRNA), messenger RNA (mRNA) and transfer RNA (tRNA) in prokaryotes, whereas RNA polymerase I (RNAP I), RNA polymerase II (RNAP II) and RNA polymerase III (RNAP III), respectively, perform these functions in eukaryotes. An important mechanism for controlling protein levels in cells is to regulate synthesis of mRNA. Therefore, regulation of RNAP II is important to control growth and differentiation of eukaryotic cells. Understanding the mechanism of transcription by RNAP II is necessary to understand how this enzyme cooperates with regulators, coactivators and basal factors to synthesize mRNA.

Transcription in prokaryotes

RNA synthesis mechanisms and regulation in prokaryotes provide important models for understanding these processes in eukaryotes. The prokaryotic apparatus includes the RNA polymerase core enzyme and a sigma factor. This holoenzyme has the capacity to initiate RNA synthesis accurately from a promoter DNA sequence. Different holoenzymes are distinguished by their associated sigma factors and specifically transcribe genes under control of a specific set of promoters. For instance, most promoters in *E. coli* are recognized by σ^{70}. This initiation factor has an RNA polymerase binding domain and sequence-specific DNA-binding domains recognizing the -10 (TATAAT) and -35 (TTGACA) regions of promoters. When the sequence of the promoter does not match the consensus, DNA-binding activator proteins are needed in addition to σ^{70} for efficient initiation. σ^{70} first binds to RNA polymerase before holoenzyme associates with the promoter. Binding of σ^{70} releases RNA polymerase from non-specific sites on DNA, and also alters the conformation of σ^{70} to expose

hexamer DNA-binding domains (Gross et al., 1992; Drombowski et al., 1992). Other sigma factors recognize distinct classes of promoters.

Initiation of transcription can be roughly divided into specific binding, isomerization, abortive initiation and promoter escape. In the presence of its sigma factor, RNA polymerase binds tightly to promoter DNA to form the closed complex. The simultaneous interaction of RNAP with the -35 and -10 hexamers induces distortion of the DNA (Mecsas et al., 1991); RNAP binding alters the torsion between the -35 and -10 sequences altering the twist of the spacer sequence (Ayer et al., 1989). The closed complex isomerizes into an open complex by separating the template DNA strands. Strand separation may initiate from the release of the torsional stress accumulated within the spacer region. Abortive initiation occurs at many promoters. RNAP synthesizes numerous short transcripts before entering a productive elongation mode (reviewed by von Hippel et al., 1984; McClure, 1985). This process involves repeated re-initiation by RNAP without escape from the promoter.

RNAP has two RNA binding sites within its active site (Reynolds et al., 1992). A "loose" RNA binding site at the catalytic center where phosphodiester bonds are formed, and a "tight" RNA binding site about 10-12 nucleotides upstream of the catalytic center. Abortive products are formed when RNA is too short to fill the loose binding site (Surratt et al., 1991; Borukhov et al., 1993; Mustaev et al., 1993; Mustaev et al., 1994). Abortive initation occurs when the high affinity of holoenzyme for the promoter out competes productive elongation. A change in RNA polymerase conformation causes release of σ^{70} and allows promoter escape and productive elongation (von Hippel et al., 1992; Gill et al., 1991).

Elongation and termination of transcription are competitive processes. The probability of continuing or terminating a chain is dependent on the conformation of RNAP, interactions with RNA bound in the "tight" and "loose" sites, and interaction with the template. Termination is likely to require dissociation of RNA from the tight binding

site. The decision to elongate or terminate is influenced by several mechanisms: 1) modification of RNAP into a terminating or anti-terminating form by binding accessory transcription factors (e.g. λN and λQ proteins) ; 2) modification of RNA secondary structures by ribosomes or specific trans-acting factors (e.g. rho-mediated termination) ; and 3) direct binding of a trans-acting factor to DNA serving as a steric impediment to polymerase progress (e.g. lac repressor and Lex A protein) (reviewed by Spencer and Groudine, 1990; Greenblatt et al., 1993).

NusA protein stabilizes the elongation conformation of RNAP. NusA enhances pausing of RNAP at certain sites and is important for transcription termination at other sites. NusA also serves to couple certain bacteriophage anti-termination factors (λ N protein) to RNA polymerase when the operon contains an appropriate recognition element (a *nut* site in the case of λ N protein). Other *E. coli* elongation factors, NusB, NusE and NusG, also modulate termination and antitermination.

Chamberlin and coworkers observed hydrolytic cleavage of the RNA by RNAP near the 3' end of nascent transcripts (Surratt et al., 1991). For some arrested elongation complexes, hydrolytic cleavage may be necessary to resume elongation. GreA and GreB proteins promote RNA cleavage. GreB can be added to complexes after RNAP has become arrested. GreA must be added before the complexes reach the arrest site to promote hydrolysis (Borukhov et al., 1992; Borukhov et al., 1993). GreA recycles between RNAP molecules and helps RNAP read through transcriptional pause sites. The functions of GreA and GreB are very analogous to those of eukaryotic elongation factor TFIIS.

In summary, accurate promoter recognition, efficient elongation, and regulated termination require proteins that bind to the core catalytic component of RNAP. All these proteins are RNA polymerase associating proteins (RAPs) in prokaryotes. Studies of these factors provide models for understanding the mechanism and regulation of eukaryotic RNAP II.

Transcription by RNAP II

The constellation of accessory factors for accurate and regulated initiation by RNAP II is more complex than that for bacterial RNAP. These factors have been classified into four different categories (Matsui et al., 1980; Davidson et al., 1983; Samuels et al., 1982; reviewed by Drapkin et al., 1993). The first category consists of basal or general transcription factors, that specify accurate initiation and support a basal level of initiation by RNAP II (reviewed by Weinmann, 1992; Zawel and Reinberg, 1993; Conaway and Conaway, 1993; Buratowaki, 1994). The second group contains sequence-specific regulators including activators and repressors. These factors have DNA binding domains to recognize specific DNA sequences within promoters, enhancers, and silencers, and regulatory domains to activate or repress transcription through direct or indirect interaction with basal factors (reviewed by Ham et al., 1992; Tjian and Maniatis, 1994). These sequence-specific factors can influence initiation, elongation and termination of transcription (Connelly and Manley, 1989a; Yankulov et al., 1994). The third group termed adaptors, co-activators or mediators act as bridges between sequence-specific transcription factors (activators and repressors) and basal factors to relay regulatory signals to RNAP II (Dynlacht et al., 1991; Tanese et al., 1991; Merino et al., 1993; Auble and Hahn, 1993; Ge and Roeder, 1994; Kretzschmar et al., 1994). The fourth group consists of factors that influence elongation (Reinberg and Roeder, 1987; Flores et al., 1989; Reines et al., 1989; Bengal et al., 1991; Marshall and Price, 1992). Elongation factors may also be targets of activators or repressors and may interact with mediators to receive signals from regulators.

In order to initiate an RNA chain from a promoter DNA sequence accurately, RNA polymerase II requires general transcription factors (TFIIA, TFIIB TFIID, TFIIE, TFIIF, and TFIIH). Physiological elongation rates require additional factors (TFIIS and TFIIX). Activators and repressors can potentially interact with the basal machinery to

dictate the level of transcription during initiation and/or elongation. Potentially, termination and RNAP II recycling could be additional targets for transcriptional control. Because of the complexity of the transcription cycle, regulatory checkpoints may include: 1) promoter recognition; 2) assembly of a preinitiation complex; 3) open complex formation; 4) synthesis of the first phosphodiester bond; 5) transition from abortive initiation to productive elongation; 6) formation of a stable ternary complex; 7) elongation; 8) termination; and 9) recycling of polymerase.

General Transcription Factors

TFIID has a central function in transcription by RNAP II. The native TFIID is a multiprotein complex consisting of the TATA-box-binding protein (TBP) and at least 7 tightly bound proteins termed TBP-associated factors (TAF_{II}s) in human (Pugh and Tjian, 1991; Zhou et al., 1993), 8 TAF_{II}s in Drosophila (Dynlacht et al., 1991) and 9 TAF_{II}s in yeast (Reese et al., 1994). Biological and genetic evidence suggests that TBP participates in RNAP I, RNAP II and RNAP III transcription both in yeast and human (Dahlberg and Lund, 1991; Simmen et al., 1991; Lobo et al., 1991; Comai et al., 1992; Schultz et al., 1992; Sharp, 1992; Rigby, 1993; Hernandez, 1993). The SL1 complex for RNAP I transcription consists of TBP, TAF_I110, TAF_I63 and TAF_I48 (Comai et al., 1994). The TFIIIB complex for RNAP III transcription consists of TBP, $TAF_{III}70$ (TDS4/BRF) and $TAF_{III}90/170$ (Kassavetis et al., 1992; Taggart et al., 1992; Poon et al., 1994). The human RNAP II TFIID complex consists of TBP, TAF_{II}s of 250, 125, 95, 78, 50, 30 and 28 kDa (Zhou et al., 1993). Genes encoding most of the TAFs have been isolated (Dynlacht et al., 1993; Goodrich et al., 1993; Hisatake et al., 1993; Hoey et al., 1993; Kokubo et al., 1993, 1994; Ruppert et al., 1993; Weinzierl et al., 1993; Yokomori et al., 1993a,b). TAF_{II}s have been suggested to stabilize binding of TBP to the carboxy terminal domain (CTD) of RNAP II (Conaway et al., 1992) and TFIIB (Goodrich et al., 1993). TAFs may also mediate signals from sequence-specific factors (Pugh and Tjian,

1990; Dynlacht et al., 1991; Hoey et al., 1993; Gill et al., 1994; Goodrich et al., 1993; Serizawa et al., 1994a; Wang and Tjian, 1994; Xiao et al., 1994; Thut et al., 1995).

TBP is the only general transcription factor shown to have sequence-specific DNA-binding activity. Recombinant TBP can form a stable complex with DNA containing a TATA box. The size of TBP varies among species (yeast 27 kDa, *Arabidopsis* 22 kDa, *Drosophila* 39 kDa and human 38 kDa), because the N-terminal domain is highly divergent across species. The C-terminal 180 amino acids, however, are highly conserved (reviewed by Greenblatt, 1991a). Minor differences within the conserved C-terminal domain contribute to the species specificity of function rather than the highly divergent N-terminal sequence (Cormack et al., 1991; Gill and Tjian, 1991). Mutagenic study of TBP indicates that the conserved C-terminal domain can bind to the TATA box and support basal transcription (Horikoshi et al., 1990; Hoey et al., 1990; Peterson et al., 1990), and the divergent N-terminal domain may interact with co-activators (Pugh and Tjian, 1990). Even though human TBP fails to replace yeast TBP functionally *in vivo*, human TBP and yeast TBP can substitute for each other in basal transcription *in vitro* (Cormack et al., 1991; Gill and Tjian, 1991). TBP binds to the minor groove of DNA (Starr and Hawley, 1991; Lee et al., 1991). The concave inner side of TBP interacts with DNA via a curved antiparallel β-sheet, whereas the convex outer surface can interact with many different proteins. Binding of TBP to promoter DNA induces a 100^{o} bend in DNA observed by a gel mobility shift assay (Horikoshi et al., 1992) and by X-ray crystallography of a DNA/protein complex (Kim et al., 1993; Kim et al., 1993). Bending may bring regulatory DNA sequences closer to the promoter. TBP partially unwinds the DNA in the TATA box, but flanking sequences remain B-form DNA. In order to compensate for unwinding, the DNA assumes a partial superhelical twist. The biological function of this helix distortion is unknown.

TFIIA has been identified as a heterodimer of 32 and 13.5 kDa subunits in yeast (Ranish and Hahn, 1991) and a heterotrimer consisting of 34, 19, and 14 kDa subunits in

human (Ma et al., 1993; DeJong and Roeder, 1993). Yeast and human TFIIA function interchangeably in basal transcription (Ranish et al., 1992). Studies show that TFIIA functions at an early stage of pre-initiation complex formation (Reinberg et al., 1987; Flores et al., 1992). The binding of TFIID to DNA is stimulated by TFIIA through direct interaction with TBP (Buratowski et al., 1989; Maldonado et al., 1990; Lee et al., 1992; Yamamoto et al., 1992; Buratowski and Zhou, 1992). However, the TBP mutants that cannot interact with TFIIA, still perform basal transcription (Yamamoto et al., 1992; Buratowski and Zhou, 1992). Also in a reconstituted transcription system, TFIIA becomes dispensable for accurate transcription when TFIID is replaced by recombinant TBP (Cortes et al., 1992). Although TFIIA is not necessarily required for basal transcription, this factor may have an important role in activation and anti-repression processes through interactions with TAF co-activators. TFIIA has been suggested to overcome inhibitory effects of a negative regulator of basal transcription named Dr2 (Merino et al., 1993; Ma et al., 1993; Liberman and Berk, 1994; Yokomori et al., 1994; Ozer et al., 1994; Sun et al., 1994). TFIIA seems to have an essential role in more complex and regulated systems and becomes dispensable in highly purified systems.

TFIIB is a single polypeptide of 33 kDa in human (Ha et al., 1991), 34.5 kDa in *Drosophila* (Wampler and Kadonaga, 1992), and 35 kDa in yeast (Pinto et al., 1992). It has been characterized as having multiple functional domains which can directly interact with TBP, TAF$_{II}$40, RNAP II, RAP30, and acidic activators (Lin and Green, 1991; Lin et al., 1991; Tschochner et al., 1992; Goodrich et al., 1993; Roberts et al., 1993; Ha et al., 1993; Barberis et al., 1993; Buratowski and Zhou, 1993). These biochemical studies demonstrate that the N-terminus of RAP30 and IIB interact with each other directly and the C-terminal domain of TFIIB interacts with TBP and also RNAP II. Also TFIIB is critical to interactions between the initiation complex and upstream activators. Association of TFIIB with the pre-initiation complex, in some cases, may be the rate-limiting step in initiation. A genetic approach applied by Hampsey and coworkers shows

that mutations in yeast TFIIB (Sua7p) cause a shift in transcriptional start site selection *in vivo* (Pinto et al., 1992; 1994). Mutations in the largest subunit (Sua8p) of yeast RNAP II similarly alter initiation, suggesting that start site selection may involve interaction between the largest RNAP II subunit and TFIIB (Berroteran et al., 1994). The *sua7* and *sua8* mutations cause initiation from new start sites but maintain initiation from normal sites (Berroteran et al., 1994). These studies suggest that TFIIB plays critical roles in recruitment of RNAP II/TFIIF to the DA complex (a complex of promoter DNA, TFIIA and TFIID), communication to regulatory factors, and transcription start site selection.

TFIIF (RAP30/74) in humans consists of two subunits performing important functions in initiation, promoter escape, elongation, and polymerase recycling (Burton et al., 1986, 1987, 1988; Chang et al., 1993; Reinberg and Roeder, 1987; Flores et al., 1989; Bengal et al., 1991; Izban and Luse, 1992a). Recent work also indicates that TFIIF can be a target for transcriptional activators (Zhu et al., 1994; Joliot et al., 1995). RAP30 and RAP74 subunits can perform separable functions in initiation and very early elongation (see Chapter II; Chang et al., 1993). The apparent molecular weight of TFIIF is 220-280 kDa as determined by gel filtration studies, suggesting a heterotetrameric structure (Conaway and Conaway, 1989; Flores et al., 1990; Kitajima et al., 1990; Wang et al., 1994). TFIIF subunits were initially identified as polypeptides capable of binding to RNAP II (Sopta et al., 1985). Both subunits of TFIIF are essential for accurate transcription in HeLa-cell nuclear extracts, from which TFIIF was removed by immunoprecipitation with antibody directed against RAP30 (Burton et al., 1986, 1988). Anti-RAP30 antibody co-precipitates RAP74, indicating that RAP30 and RAP74 are normally associated in a complex (Burton et al., 1988). Binding to RNAP II prevents phosphorylation of RAP30 but not RAP74, suggesting that the RAP30 subunit interacts directly with RNAP II (McCracken and Greenblatt, 1991). However, both recombinant RAP30 and RAP74 can bind to RNAP II in a gel-mobility shift assay independently (Killeen and Greenblatt, 1992; Tyree et al., 1993; Wang and Burton, 1995). RAP74

inhibits transcription of RNAP II in a non-promoter transcription assay in the absence of RAP30, and the C-terminal region of RAP74 binds tightly to RNAP II (Wang and Burton, 1995). TFIIF inhibits nonspecific binding of RNAP II to free DNA (Conaway and Conaway, 1990; Killeen and Greenblatt, 1992) and RAP30 assists RNAP II to bind to the DAB complex (Flores et al., 1991; Killeen et al., 1992). However, recombinant RAP30 itself cannot release RNAP II from non promoter DNA once it has bound, although TFIIF has this capacity (Killeen et al., 1992; Killeen and Greenblatt, 1992).

TFIIF interacts with TFIIB through the RAP30 subunit (Ha et al., 1993), and also with serum response factor through the RAP74 subunit (Zhu et al., 1994; Joliot et al., 1995). Gel mobility shift assays suggest that RAP30/RNAP II can bind to the DAB complex (TFIID TFIIA TFIIB and promoter DNA). In gel shifts, the intensity of the DABPolF complex is dramatically increased by RAP74 (Flores et al., 1991). Therefore, RAP74 strongly increases the stability of polymerase association with the promoter, although this TFIIF subunit is not required for polymerase entry.

TFIIF isolated from human (RAP30/74), rat ($\beta\gamma$), and fly (Factor 5) contains two subunits, but the counterpart in yeast (factor g) consists of three subunits, Tfg1p/Ssu71p, Tfg2p, and Tfg3p/Anc1p/TAF30 (Henry et al., 1992; 1994). Studies of *S. cerevisiae* demonstrate that the yeast counterpart of RAP74 (Ssu71p) interacts with TFIIB (Sua7p) genetically (Sun and Hampsey, submitted). Mutations in RAP74 (Ssu71p) suppress both a cold-sensitive growth phenotype and the altered initiation pattern conferred by a TFIIB defect. The RAP30 counterpart in yeast is Tfg2p. From recent studies of RNAP II binding proteins in yeast, Cdc73p appears to be have a function related to those of Tfg2p and RAP30 (See Chapter IV of this thesis; Wade et al., 1995; Fang and Burton, 1995). Cdc73p may be part of an alternate TFIIF function in yeast. Tfg3p/Anc1p functions in stimulating transcriptional activity and can be dispensable for basal transcription (Henry et al., 1992). Tfg3p/Anc1p has been shown to be TAF30 of the yeast TFIID complex (Henry et al., 1994). Tfg3p/Anc1p has also been identified as the 30 kDa subunit of the

SWI/SNF complex, which alters chromatin structure to activate transcription (Cairns et al., 1994).

Since TFIIF is involved in many phases of the transcription cycle, regulators interacting with TFIIF might impact on promoter recognition, transcription complex assembly, initiation of RNA synthesis, promoter escape, and elongation.

TFIIE is a heterotetramer composed of two 34- and 56-kDa subunits (Ohkuma et al., 1990; Inostroza et al., 1991), that can stably associate with RNAP II in solution. TFIIE has been purified as one of the RAPs from an RNAP II affinity column (Buratowski et al., 1991; Flores et al., 1989). TFIIE binds to the form of RNAP II that is not phosphorylated on the CTD (RNAP IIA) through the 56-kDa TFIIE subunit (Maxon et al., 1994). TFIIE stimulates the CTD kinase activity of TFIIH promoting CTD phosphorylation (Serizawa et al., 1994b; Ohkuma and Roeder, 1994). TFIIE has also been found to modulate the helicase activity of TFIIH either negatively (Drapkin et al., 1994a) or positively (Serizawa et al., 1994b) through a direct interaction with the largest subunit of TFIIH. It has been suggested that TFIIE assembly during pre-initiation complex formation is after assembly of RNAP II/TFIIF but before subsequent recruitment of TFIIH (Flores et al., 1992), and indeed TFIIE interacts with TFIIH and both subunits of TFIIF (Maxon et al., 1994). Recent studies indicate that TFIIE along with TFIIH and ATP are not required for initiation, but instead are required for promoter escape, suggesting that they are necessary for conversion of an initiated complex to an elongation complex (Goodrich and Tjian, 1994).

TFIIH is a multisubunit transcription factor with polypeptides 34, 38, 41, 44, 50, 62, 80, and 89 kDa (Schaeffer et al., 1993; 1994). TFIIH has two catalytic activities, an ATPase/helicase activity involved in promoter escape (Goodrich and Tjian, 1994), and a kinase activity involved in CTD phosphorylation of RNAP II (Lu et al., 1992). TFIIH also phosphorylates TBP, the 56 kDa subunit of TFIIE and the RAP74 subunit of TFIIF (Ohkuma and Roeder, 1994). The p62 subunit of TFIIH interacts with the activation

domains of VP16 and p53 indicating that TFIIH is a target of regulators (Xiao et al., 1994). Transcriptional activity and CTD kinase co-fractionate, and anti-p62 antibody inhibits both transcriptional and kinase activities. None of the TFIIH subunits, however, has been shown to have intrinsic CTD kinase activity. The largest and the second largest subunits of TFIIH, p89 and p80, in human contain consensus ATPase/helicase motifs and have been shown to be identical to the XPB-ERCC3 and XPD-ERCC2 excision repair proteins (Schaeffer et al., 1993; Feaver et al., 1993). Therefore, TFIIH functions in both transcription and DNA repair. The yeast counterparts of these helicases (SSL2 and RAD3) are both essential for cell viability and transcriptional activity (reviewed by Drapkin and Reinberg, 1994a; 1994b); however, point mutations in the nucleotide binding domain of the second largest subunit (RAD3) are not lethal and result only in defects to nucleotide-excision repair (Feaver et al., 1993; Sung et al., 1988), whereas similar mutations in the largest subunit (SSL2) are lethal. Extracts derived from *ssl2* mutants grown under the non-permissive condition are transcriptionally inactive (Guzder et al., 1994). Recent evidence indicates that there are two forms of TFIIH, a holo-TFIIH involved in transcription and a repairasome responding in nucleotide-excision repair (Svesjstrup et al., 1995). Thus, TFIIH appears to play a critical role in ATP hydrolysis and open complex formation, CTD phosphorylation, promoter escape, and elongation of RNAP II when template DNA contains a damaged base.

RNA polymerase II and RAPs

RNA polymerase II (RNAP II) is the nuclear DNA–dependent RNA polymerase that synthesizes mRNA in eukaryotic cells. Several conserved features exist among RNAP II isolated from different species. RNAP II contains 10-12 subunits and the relevant subunit sequences are highly conserved between species. RNAP II has two large subunits homologous to the largest subunits of RNAP I, RNAP III, and the β' and β subunits in prokaryotes. The largest subunit of RNAP II contains a unique C-terminal domain (CTD), not present in other polymerases, composed of multiple heptapeptide

repeats with the consensus sequence YSPTSPS, 52 repeats in the mammalian enzyme. The largest RNAP II subunit contains a DNA binding site and the second largest subunit binds nucleotide substrates. Both subunits contribute to the active site for RNA synthesis. Finally, RNAP II contains three common subunits of 14-28 kDa also found in RNAP I and RNAP III, and other small subunits unique to RNAP II (reviewed by Young, 1991).

The CTD is multiply phosphorylated on the two SP serines within each heptapeptide repeat by a stably associated CTD kinase using either ATP or GTP as substrate. The highly phosphorylated form of RNAP II is referred to as the IIO form, and the dephosphorylated form as the IIA form (Cadena and Dahmus, 1987; Kim and Dahmus, 1989).

RNAP II binds to the DB complex in the RNAP IIA form. The dephosphorylated CTD binds to the TBP subunit of TFIID. Conversion of the CTD to the phosphorylated form in the pre-initiation complex reduces the affinity of the CTD for TBP (Dahmus and Kedinger, 1983; Laybourn and Dahmus, 1989; Usheva et al., 1992). Transition from a pre-initiation complex to an elongation complex, therefore, involves CTD phosphorylation and RNAP IIO is the primary elongation form of polymerase. Phosphorylation of the CTD, however, is not the only requirement for the transition from the preinitiation complex to the initiated complex. RNAP II from which the CTD has been removed (termed the IIB form) can accurately initiate in an ATP-dependent reaction (Zehring et al., 1988). Moreover, transcription initiation can be uncoupled from CTD phosphorylation using a protein kinase inhibitor (Serizawa et al., 1993). Therefore, Conaway and coworkers suggest that phosphorylation of the CTD is not an essential step in basal transcription or factors that establish the requirement for CTD phosphorylation are missing in some *in vitro* systems (Serizawa et al., 1993). Although phosphorylation of the CTD and the CTD itself are not required for the catalytic activity of RNAP II *in vitro*, the CTD is essential *in vivo* (Nonet et al., 1987; Allison et al., 1988; Zehring et al.,

1988; Bartolomei et al., 1988). Genetic and biochemical evidence suggests that the CTD plays a role in mediating transcription activation by upstream regulators (Allison and Ingles, 1989; Usheva et al., 1992). Chromosome fluorescence staining using antibody against the hyperphosphorylated or hypophosphorylated CTD suggests that polymerase is in the IIA form in stalled elongation complexes and IIO form in active elongation complex (Weeks et al., 1993; O'Brien et al., 1994).

RAP is an acronym for an RNA polymerase II associating protein, and these factors have been isolated using affinity methods in which RNAP II is immobilized along with bound factors (Sopta et al., 1985; Burton et al., 1986; Greenblatt, 1991b; Wade et al., 1995). RAPs have been identified as initiation factors (e.g. TFIIF (RAP30/74), TBP, TFIIB, TFIIE, TFIIH), elongation factors (e.g. RAP30/74, RAP38), and possible mediators (e.g. AF9/ENL, Anc1p). RAPs may bind directly or indirectly to RNAP II. Such factors are not identified as RNAP II subunits because of dissociation during purification, just as the *E. coli* sigma factor and NusA can be separated from "core" RNAP (Burgess et al., 1969; Greenblatt and Li, 1981). RAPs have generally been considered accessory factors rather than subunits of RNAP II because binding is sensitive to elevated salt concentration (0.3-0.5 M KCl) and in many cases RAPs can be exchanged from one RNAP II to another. Dissociable interaction between RAPs and RNAP II allows regulatory mechanisms in which these contacts are modified. TFIIF and TFIIH, for instance, appear to be targets for transcriptional regulators. RAP activity can also be controlled by phosphorylation or other covalent modification. Recently, holo enzyme forms of RNAP II have been isolated from yeast (Koleske and Young, 1994) or partially reconstituted using mammalian factors (Conaway et al., 1992). Such higher order assembly complexes can accurately initiate and respond to regulatory signals with minimal supplementation of factors.

TFIIF subunits RAP30/74 were first isolated using an RNAP II affinity column (Sopta et al., 1985; Burton et al., 1986). Recently, yeast RAPs were isolated using an

anti-CTD monoclonal antibody immobilized on a column to capture RNAP II and associated factors (Wade et al., 1995). This direct capture protocol may allow isolation of RAPs not detected using the original methods, because exchange of factors to immobilized RNAP II is not required.

Some transcription factors that bind RNAP II through the CTD might be lost by anti-CTD chromatography. This appears to be the case for TBP, TFIIH, and suppressor of RNA polymerase B proteins (Srbps) (Fisher et al., 1992; Koleske et al., 1992; Thompson et al., 1993). The yeast RAPs isolated from an anti-CTD column included known transcription factors TFIIB, TFIIF and TFIIS, as well as several proteins of unascribed transcriptional functions. As discussed above, yeast TFIIF consists of three subunits namely Ssu71p /Tfg1p, Tfg2p and Anc1p/Tfg3p. Most interestingly, a possible RAP30 homologue named Cdc73p was identified by protein sequence analysis (Wade et al., 1995).

In prokaryotes, RNA polymerase obtains its promoter specificity by contacting with different σ factors. The polymerase binding domains of human RAP30, yeast Tfg2p, and Cdc73p, and bacterial sigma factors show sequence similarity and conserved function (Wade et al., 1995; Fang and Burton, 1995). Limited homology to bacterial σ factor DNA binding domain has also been found in the fourth subunit of yeast RNAP II and mitochondrial RNA polymerase (reviewed by Jaehning,1991). The CDC73 gene was isolated as a suppressor of a deletion mutant in the STE2 gene which encodes the α-factor receptor. The cdc73-1 mutant enables the ste2 deletion mutant to mate at permissive temperature and arrests the cell cycle in G1 upon temperature upshift (Reed et al., 1988). Our work confirms that Cdc73p is a yeast RAP and is similar in sequence to human RAP30 (Wade et al., 1995; Chapter IV). Therefore, Cdc73p can potentially function to link the regulation of cell cycle and transcription. Cdc73p activity could be controlled by cell cycle signal(s) and may function as an alternative RAP30 in transcription of specific

genes, in much the same way that alternate σ-factors are responsible for transcription from alternate promoters in bacteria.

Another yeast RAP Anc1p/Tfg3p may function as a mediator or co-activator rather than as a basal factor. Anc1p is not essential for yeast viability or for accurate transcription, although transcription is stimulated by this factor (Henry et al., 1992). In addition to our identification of Anc1p as a RAP (Wade et al., 1995), Anc1p is also a subunit of TFIIF, a yeast TAF (Henry et al., 1994) and a component of the SWI/SNF complex (Cairns et al., 1994). SWI/SNF alters nucleosome structure and can activate transcription on nucleosomal DNA (Imbalzano et al., 1994).

Mechanism Of RNA Polymerase II Transcription

RNAP II and basal factors cooperate to form a complex with capacity to bind a promoter specifically and to initiate transcription accurately. This elaborate process involves multiple steps including: 1) template activation; 2) promoter recognition; 3) DNA strand separation; 4) abortive initiation; 5) formation of a stable ternary complex; 6) processive elongation; 7) termination of transcription and 8) cycling of RNAP II. The mechanism of each phase is discussed as follows.

1) Template Activation

Before a promoter can be recognized, chromosome structure must be altered to expose the DNA sequence. The alteration from an inactive template to an active one must be influenced by cellular factors. Studies on histone-mediated transcription repression indicate that formation of active transcription complexes on a promoter competes directly with the assembly of the DNA into nucleosomes (Workman et al., 1990; reviewed by Paranjape et al., 1994). Activators enhance transcription in two ways: 1) a process referred to as antirepression which relieves chromatin-mediated repression of transcription (Croston et al., 1991), and 2) a true activating process which facilitates formation of the transcription complex (Laybourn and Kadonaga, 1992).

2) Promoter Recognition

Accurate transcription is initiated from promoter sequences (Weil et al., 1979; Manley et al., 1980). Two working models have been established for preinitiation complex formation, namely a multiple-step assembly pathway (reviewed by Zawel and Reinberg, 1993) and a holoenzyme RNAP II complex pathway (reviewed by Serizawa et al., 1994; Koleske and Young, 1995).

The core promoter (minimal DNA sequence required for basal transcription) can be roughly classified into two categories: 1) a TATA-containing promoter; and 2) a TATA-less promoter. On a TATA-containing promoter, the interaction between DNA and TBP serves as a foundation for transcription complex assembly. Transcription of RNAP II on some promoters without the consensus TATAAA sequence still requires TFIID to function. The recruitment of TBP to a TATA-less promoter requires tethering to upstream activators (e.g. SP1 and TAF interaction) and/or an initiator (Inr) binding protein (e.g. TFIII and TAF interaction). Inr is the DNA sequence encompassing the transcription initiation site. Now it appears that most promoters contain an Inr, although the nucleotide sequence of this element is not highly conserved (reviewed by Weis and Reinberg, 1992). Several lines of evidence also suggest that TBP can bind to the -30 region relative to the transcription start site on a TATA-less promoter (Wiley et al., 1992; Zenzie-Gregory et al., 1993). In any case, either TATA or Inr motif alone can potentially direct transcription initiation (Myers et al., 1986; Smale et al., 1990; Nakatani et al., 1990). When these two motifs are present together, they can function cooperatively (Smale and Baltimore, 1989; Nakatani et al., 1990; Conaway et al., 1990).

(I) Multiple-step assembly process

In the ordered-assembly model (Figure 1), TATA-binding protein (TBP) first binds to the TATA motif of the promoter. The binding of TFIID to DNA is facilitated by TFIIA, and under certain conditions, TFIIA is not absolutely required for the transcriptional activity. TFIIB then binds to form a DAB complex (a complex consisting of TFIID, TFIIA, TFIIB and promoter DNA) (Buratowski et al., 1989; Maldonado et al.,

1990). RNA polymerase II associated with RAP30/74 (TFIIF) binds the DAB complex. TFIIF binding to RNAP II suppresses non-specific DNA binding by polymerase and promotes stable association of RNAP II with the promoter. Although both RAP30 and RAP74 associate with polymerase *in vivo*, RAP30 is necessary and sufficient to guide polymerase to the DAB complex (Killen and Greenblatt, 1992; Flores et al., 1991). The RAP74 subunit stabilizes this interaction. Consistent with these observations, our study demonstrates that RAP74 can function at a later stage of the transcription process than RAP30 (Chang et al., 1993; Chapter II). After formation of the DBPolF complex, TFIIE and TFIIH join and form the DABpolFEH complex. On linear DNA templates, TFIIE and TFIIH are necessary to fully convert the closed DNA complex to an open, transcriptionally competent complex. It has been demonstrated that activators can interact with several basal transcription factors; therefore, the binding affinity of proteins in each assembly step could be modulated by protein-protein interaction between regulators and basal factors.

(II) RNAP II holoenzyme complexes

Recent evidence suggests that aggregates of basal and regulatory factors can assemble on RNAP II to constitute holoenzyme forms. RNAP II holoenzymes mayperform specific functions within cells, such as regulated initiation or elongation of RNA chains. Holoenzyme forms have been isolated or reconstituted that can bind to the promoter with minimal supplementation of additional factors (Conaway et al., 1992; Koleske and Young, 1994; Kim, Y.-J. et al., 1994). Some of these complexes maintain their integrity during many steps of purification, and can support activated transcription. Other aggregates can be dissociated under mild conditions, but may represent important RNAP II forms that exist in vivo.

Many of the individual protein-protein contacts that contribute to holoenzyme formation have been identified. For instance, TFIIB interacts directly with TBP and RNAP II to recruit RNAP II/TFIIF (Barberis et al., 1993; Buratowski and Zhou, 1993).

Figure 1. Preinitiation complex formation pathways: a multiple-step assembly pathway and a holoenzyme assembly pathway.

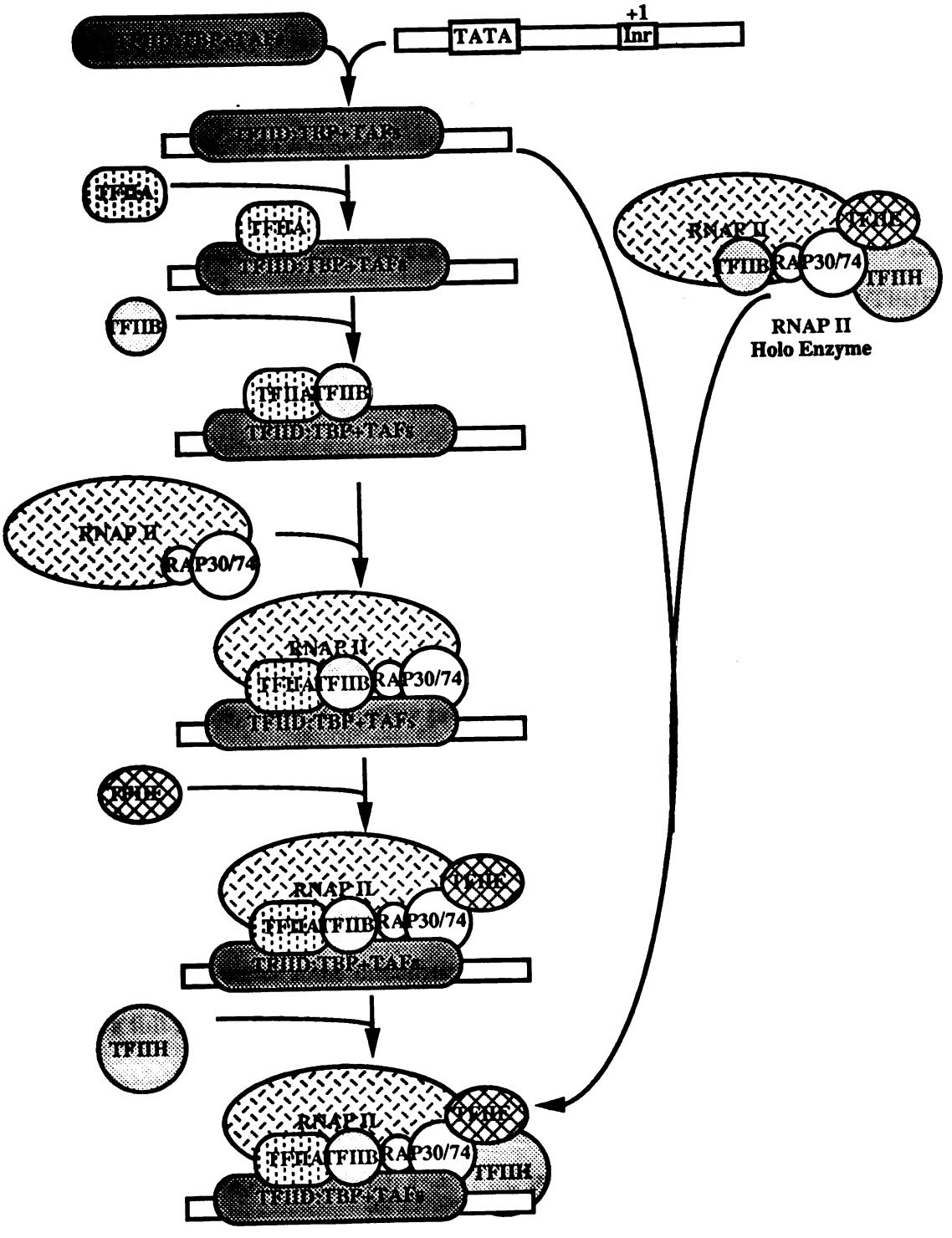

TATA
+1
Inr
TFIIA
TFIIB
TFIIH
RNAP II
RAP30/74
RNAP II
Holo Enzyme

Figure 1

The N-terminus of RAP30 and TFIIB interact (Ha et al., 1993). Furthermore, TBP and TFIIE can bind RNAP II through the CTD (Usheva et al., 1992; Maxon et al., 1994). Our laboratory has observed a number of RAPs associating with RNAP II isolated as a mixture of holoenzyme forms by anti-CTD affinity chromatography, including TFIIB, TFIIF, TFIIS, and Cdc73p (Wade et al., 1995).

Assembly mechanisms and holoenzyme forms indicate an essential role for TFIIF in establishment and maintenance of complexes. Since RAP30 interacts with TFIIB directly (Ha et al., 1993), and RAP74 increases the stability of the RNAP II/TFIIF complex (Garret et al., 1992), the overall stability of a holoenzyme likely depends on TFIIF, placing this factor at the core of the assembly process. Since TFIIF is part of both initiating and elongating polymerase, it is likely an important scaffold for construction of multiple holoenzyme forms for different stages of the transcription cycle.

Since the multi-step pathway was established according to the minimal contacts necessary for assembly in vitro, holoenzyme forms are expected to more closely resemble physiological forms of RNAP II. It may be that on some promoters, RNAP II binds in a particular holoenzyme form. On other promoters, particular contacts between factors may be regulated, and a mixture of holoenzyme and assembly mechanisms may be important for initiation. Elongating polymerase holoenzymes are likely to differ substantially from initiating forms.

(III) Minimal transcription systems

Interestingly, under some circumstances, RNAP II can accurately initiate in the absence of a full complement of basal transcription factors. For instance, on supercoiled templates, initiation can occur in the absence of TFIIE and TFIIH (Parvin and Sharp, 1993; Tyree et al., 1993). Negative supercoiling of the template removes the requirement for DNA strand separation by TFIIH, and TFIIE participates in this process (Goodrich and Tijan, 1994). In the case of a superhelical IgH promoter, TFIIF was dispensible for accurate transcription (Parvin and Sharp, 1993). Surprisingly, on the adeno-associated

virus (AAV) P5 promoter, TFIID is dispensible for accurate initiation. In this case YY1, an Inr-binding protein, TFIIB and RNAP II are sufficient to direct basal transcription from a supercoiled template (Usheva and Shenk, 1994).

Whether these minimal complexes have physiological relevance is not known. In highly purified transcription systems, TBP can substitute for TFIID, but in mammalian cell extracts TBP is assembled into TBP/TAF complexes. In the presence of TAFs, other co-activators and regulators, transcription is likely to require a more complete constellation of factors. Some promoters, on the other hand, may have slightly different requirements than those studied to date, and may initiate by different mechanisms using an alternate collection of factors. In Chapter IV of this thesis, a yeast protein named Cdc73p is identified that may be a component of an alternate form of TFIIF and may support an alternate transcriptional mechanism by RNAP II.

3) Open complex formation

DNA strands must be separated to expose the template for phosphodiester bond synthesis. In higher eukaryotes, initiation occurs 25 to 30 nucleotides downstream from the TATAAA sequence (Corden et al., 1980). DNA strand separation can be detected by modification of single-stranded DNA with reagents such as $KMnO_4$. Productive initiation requires hydrolysis of ATP at the β–γ bond position (Bunick et al., 1982; Sawadogo and Roeder, 1984). This is a unique requirement of RNAP II, since RNAP I, RNAP III, and bacterial RNAP do not require ATP hydrolysis for initiation. RNAPs of all sorts utilize ATP as a substrate for elongation, but as RNA chains are elongated, it is the α–β bond of ATP that is hydrolyzed. ATP analogs such as AMPPNP can substitute for ATP for initiation and elongation by most polymerases, but not for initiation by RNAP II. Full start site opening of the DNA helix requires ATP or ATP analogs with a hydrolyzable β–γ bond (Wang et al., 1992; Jiang and Gralla, 1995). At sufficiently high concentrations, GTP, CTP and UTP can be substituted for ATP as the substrate for β–γ bond hydrolysis (Wang et al., 1992; Jiang et al., 1994; Jiang and Gralla, 1995), but only

ATP has been shown to support this function for runoff transcription in vitro. Perhaps ATP hydrolysis supports more than one requirement in initiation of RNA chains by RNAP II.

A helicase activity has been proposed to carry out the DNA unwinding reaction in RNAP II initiation (Burton et al., 1988). Two subunits of TFIIH have been shown to include ATP-dependent helicase activities (Feaver et al., 1993; Serizawa et al., 1993; Schaeffer et al., 1993). Current evidence suggests that the largest TFIIH subunit is the helicase required for transcription, and the other helicase subunit is required for DNA repair (Guzder et al., 1994). A template already in the open conformation can be accurately transcribed without addition of ATP (Tantin and Carey, 1994). As mentioned above, highly supercoiled templates do not require TFIIH for initiation, nor do they require ATP hydrolysis.

4) Abortive initiation and promoter escape

RNAP II, similar to prokaryotic RNA polymerase, goes through an abortive phase of transcription in which short transcripts are synthesized and released before productive elongation occurs (reviewed by McClure, 1985). An open binary complex can initiate phosphodiester bond synthesis in the presence of all four nucleoside triphosphates. However, these newly synthesized RNAs are unable to associate with polymerase as a stable ternary complex until about 10 phosphodiester bonds have been formed (Chapter III; Chang and Burton, 1995). In our system this process requires about 15 s. Transcripts 5 to 10 nucleotides in length are not stably associated with polymerase and are released as abortive transcripts (Luse and Jacob, 1987; Luse, 1990; Jacob et al., 1991, 1994). Promoter escape, also called promoter clearance, is the transition between abortive initiation and productive elongation. One explanation for abortive initiation is that polymerase must overcome strong protein-protein and protein-DNA contacts holding it to the promoter. The molecular forces that drive promoter recognition and accurate initiation are in competition with those that drive elongation. Polymerase must make a

transition from being a sequence-specific DNA binding protein, with high selectivity for promoter sites, to an elongation enzyme that has little ability to discriminate between different sequences. Presumably, this transition requires conformational changes in polymerase and release of some initiation factors. For instance, escape from abortive initiation in bacterial systems involves the release of sigma factor and significant changes in polymerase conformation. Template contacts also change dramatically as promoter escape occurs (Straney and Crothers, 1987; Krummel and Chamberlin, 1989; Metzger et al., 1989).

The RAP74 subunit of TFIIF has important functions in promoter escape by RNAP II (Chapter II; Chang et al., 1993). Analysis of short transcripts formed in the presence of RAP74 mutants demonstrates that the stability of short ternary complexes is dependent on RAP74 (Chapter III). It is likely that the function of RAP74 in promoter escape relates to stabilization of short transcripts.

The helicase function of TFIIH has been proposed to drive promoter escape (Goodrich and Tjian, 1994; Maxon et al., 1994). Phosphorylation of the CTD reduces the affinity of RNAP II for TBP, and may be another feature of the transition between initiating and elongating enzyme (Usheva et al., 1992; Laybourn and Dahmus, 1989).

Interestingly, abortive initiation does not require TFIIH or ATP hydrolysis, although an accessible template strand is required (Goodrich and Tijan, 1994; Timmer, 1994). When open complex formation is monitored with the single-strand cleavage reagent o-phenanthroline-copper, TFIIH and ATP hydrolysis are found to be not required for sensitivity of the DNA. Apparently, the open complex is formed in two steps. The initial step requires assembly of the DABpolF complex. The open complex formed under this condition will support abortive initiation but not promoter escape and productive elongation. This open complex can be detected with o-phenanthroline-copper, but not with $KMnO_4$. To free polymerase from the promoter requires an ATP-dependent step in which the TFIIH transcription helicase makes a more extensive open complex.

Movement of TFIIH may be important to remodel protein-protein interactions that otherwise hold RNAP II at the promoter.

Recently Chamberlin and coworkers demonstrated that RNAP II has two RNA binding sites responsible for the association of RNA products with transcription complex (Johnson and Chamberlin, 1994; Chamberlin, 1995). Similar RNA binding sites have also been observed in *E coli* RNAP (Borukhov et al., 1993; Mustaev et al., 1994). A low affinity site is located at the catalytic center of the enzyme where phosphodiester bonds are synthesized and requires a chain length greater than about 10 nucleotides for occupancy. A higher affinity site is located 10-12 nucleotides distant from the catalytic center. Instability of short RNAs, therefore, may relate to the ability to fill in the low affinity site. Transcripts of less than about 10 nucleotides are weakly bound and subject to release as abortive transcripts.

5) Elongation

As each new phosphodiester bond is formed, polymerase makes a molecular decision whether to elongate or to terminate the chain. Polymerase can pause, and after traversing some kinetic barrier, resume elongation. Arrested complexes are distinct from paused complexes. These are stalled complexes that cannot be elongated without some event, such as nucleolytic cleavage of the transcript, to re-establish the association of the catalytic center with the template. Elongating, paused, arrested, and terminating complexes are characterized by binding of the transcript to polymerase, RNAP II conformation, and DNA contacts. Factors that regulate elongation and termination are expected to affect these processes through interactions with polymerase, transcript, and template (reviewed by Spencer and Groudine, 1990; Kerppola and Kane, 1991; Krumm et al., 1993; Greenblatt et al., 1993; Kane, 1994). .

Pausing is thought to be mediated through interaction with specific sequences within DNA or RNA. Pause sites can be recognized by RNAP II in the absence of other factors (Reines et al., 1987; Kerppola and Kane, 1988; Wiest et al., 1992). Differential

modification of the polymerase CTD by phosphorylation may alter interaction with pause sequences. Elongation factors TFIIF, TFIIS, and TFIIX increase RNAP II elongation efficiency perhaps by suppressing pausing (Reinberg and Roeder, 1987; Flores et al., 1989; Reines et al., 1989; Bengal et al., 1991; Izban and Luse, 1992a; Yankulov et al., 1994). These factors have distinct activities that depend on whether they are added before or after RNAP II reaches the pause site. TFIIF and TFIIX stimulate the rate of elongation and promote reading through pause sites. TFIIS exerts its influence on transcription after polymerase has paused.

To recover arrested complexes, endonucleolytic cleavage of the nascent transcript is necessary (Reines et al., 1992; Izban and Luse, 1992b, 1993; Wang and Hawley, 1993; Rudd et al., 1994; reviewed by Reines, 1994). Two distinct cleavage reactions have been identified. Pyrophosphorolysis is the reverse of the normal polymerization reaction, and hydrolytic cleavage of the transcript near the 3' end can also occur. RNAP II will catalyze these reactions in the absence of elongation factors indicating that these catalytic processes are intrinsic to polymerase. TFIIF stimulates the pyrophosphorolysis reaction, and TFIIS stimulates the rate of transcript hydrolysis (Reines et al., 1992; Wang and Hawley, 1993; Rudd et al., 1994).

Human immunodeficiency viruses (HIV) transacting protein Tat is a potent activator of transcription from the HIV-1 long terminal repeat (LTR) promoter (reviewed by Sharp and Marciniak, 1989; Cullen, 1990, 1993; Greenblatt et al., 1993). Tat stimulates initiation (Laspia et al., 1989; Southgate and Green, 1991), elongation, and processivity (Marciniak and Sharp, 1989; Laspia et al., 1989, 1990; Marciniak et al., 1990; Kato et al., 1992; Zhou and Sharp, 1995). Tat functions through binding to an RNA element termed TAR (Muesing et al., 1987; Hauber and Cullen, 1988). Roeder and coworkers observed that TFIIF increases the basal level of elongation but not Tat-activated stimulation, whereas TFIIS shows synergistic stimulation of elongation with Tat. Antiserum directed against the RAP74 subunit of TFIIF preferentially suppresses

the activated level of transcription exerted by Tat (Kato et al., 1992). These observations lead the authors to suggest that Tat may stimulate elongation through RAP74. Landick and coworkers observed that Tat augments the effect of TFIIF on RNAP II processivity and suggested that stimulation of transcription elongation by Tat occurs at least partially by recruitment of TFIIF to the elongating transcription complex(Meier et al., 1994). However, a direct protein-protein interaction between Tat and TFIIF has not yet been reported.

Price and coworkers reported that productive elongation complexes are derived from early paused elongation complexes by the action of a factor P-TEF (positive transcription elongation factor) to release an N-TEF (negative transcription elongation factor) (Marshall and Price, 1992). RNAP II complexes that are not appropriately modified are defective for processivity and terminate transcription. The identities of N- and P-TEF, however, have not yet been reported.

In vivo RNAP II molecules have been identified that are initiated but are stalled for elongation near the 5' ends of various genes (Rougvie and Lis, 1990; Bengal et al., 1991; Giardina et al., 1992; Krumm et al., 1992; Meulia et al., 1992; Strobl and Eick, 1992; Kash et al., 1993). For instance, on the *hsp70* and *hsp26* promoters of *Drosophila melanogaster*, RNAP II has been found to be initiated but stalled within the first 29 nucleotides of the mRNA chain prior to heat shock. After heat shock, polymerase molecules are found distributed throughout the *hsp70* gene, and the mature *hsp70* mRNA and protein are produced (Rougvie and Lis, 1990; Giardina et al., 1992). Even after the *hsp70* gene is activated by heat shock, the pause site for polymerase near the promoter is maintained and the single-stranded DNA region at the pause site can be detected with $KMnO_4$. The implication of these observations is that the heat shock transcription factor (HSF) stimulates polymerase to traverse the pause site and productively elongate the transcript through the gene, in addition to effects HSF may have on initiation. Pausing in

this case may poise RNAP II and the *hsp70* transcription unit to respond quickly to the heat shock signal.

The phosphorylation status of the CTD of stalled RNAP II molecules on *hsp70* and *hsp26* genes has been examined before and after heat shock. Hyperphosphorylated or nonphosphorylated RNAP II molecules were detected after ultraviolet cross linking, using specific antibodies directed against the different polymerase forms (Weeks et al., 1993). RNAP II in productive elongation complexes after heat shock has a phosphorylated CTD, while the CTD of the paused polymerase is mainly unphosphorylated (O'Brien et al., 1994). Since the primary elongation form of RNAP II is hyperphosphorylated, this result is consistent with the presumed function of CTD modification in the transition between initiating and elongating forms of polymerase. On heat shock genes, phosphorylation of the CTD occurs after initiation to drive polymerase into productive elongation.

6) Termination

The mechanism for RNA polymerase II to terminate a chain has not been clearly elucidated. 3'-end formation of mRNA transcripts involves transcript cleavage downstream of an AAUAAA processing and polyadenylation signal. Polymerase appears to lose processivity after processing occurs and to terminate at many sites beyond the polyadenylation signal. Termination by RNAP II at the 3' end of genes encoding poly(A) mRNAs is thought to require several distinct cis-acting elements including: 1) a functional poly(A) signal (Connelly and Manley, 1988; Laniox and Acheson, 1988; Logan et al., 1987; Whitelaw and Proudfoot, 1986); 2) a downstream transcriptional pause site (Logan et al., 1987); 3) a structural element causing a bend in the DNA helix (Kerppola and Kane, 1990); and 4) termination signal sites on the 5' flanking region of promoters which can be specifically recognized by trans-acting factors (Connelly and Manley, 1989a, 1989b; Meulia et al., 1992; Roberts et al., 1992). Perhaps CTD phosphorylation is decreased after polymerase traverses a polyA addition site.

7) Recycling of RNAP II

Polymerase must be in the dephosphorylated state to efficiently associate with a promoter. Since elongating polymerase is highly phosphorylated, a CTD phosphatase is likely to be important for polymerase recycling after termination. Chambers and Dhamus recently reported isolation of a CTD phosphatase (Chambers and Dhamus, 1994). CTD phosphatase functions by binding to a site on RNAP II other than the CTD and is strongly stimulated by the RAP74 subunit of TFIIF (Chambers et al., 1995).

Response of the general mechanism to activators

Each phase of the transcription cycle is a potential target for regulation. Basal factors are now thought to be required for initiation of all genes; activators and repressors dictate the rate at which the basal complex initiates transcription. Initiation can be stimulated by interaction between promoter-, silencer-, and enhancer-binding factors. Activation domains of regulators, for instance, have been shown to interact directly with TBP and TFIIB. In some cases, co-activators such as TAFs are important for transmission of these signals. In an ordered assembly model, a particular step in assembly could be affected by regulators and coactivators. In the holoenzyme model, these contacts could affect a rate-limiting step in initiation. Isolated RNAP II holoenzymes have been shown to respond to some activators including Gal4-VP16 and GCN4 (Kim, Y.-J. et al., 1994). Serum response factor regulates transcription through the RAP74 subunit of TFIIF (Zhu and Prewes, 1994; Joliot et al., 1995). Since RAP74 functions in initiation, promoter escape, and elongation, serum response factor may regulate each of these processes. The transcriptional activation domain of VP16 interacts with TFIIH (Xiao et al., 1994), and this contact may stimulate the initiation helicase to promote open complex formation (Jiang et al., 1994). Open complex formation is an important target for bacterial activators (reviewed by Gralla, 1990).

Overview

Understanding the mechanism and regulation of RNAP II requires description of the functions of factors in transitions between stages of the transcription cycle. Our laboratory has focused on functions of TFIIF. Cloning cDNAs encoding human RAP30 and RAP74 has facilitated these studies (Sopta et al., 1989; Finkelstein et al., 1992; Wang et al., 1993, 1994). An extract depleted of these factors has been used to assay TFIIF function in vitro (Burton et al., 1986, 1988; Finkelstein et al., 1992; Chang et al., 1993; Wang and Burton, 1995; Chang and Burton, 1995). Bacterial production of RAP30 and RAP74 proteins has provided active TFIIF for reconstitution studies. Deletion mutants of TFIIF subunits have been constructed to relate functional domains to protein sequence.

As described in chapter II, sarkosyl challenge and pulse-chase assays were used to distinguish the requirements of TFIIF subunits for initiation and elongation. RAP30 was essential to establish a sarkosyl-resistant complex, indicating that RAP30 is required for initiation. RAP74 was not required for creation of a sarkosyl-resistant complex, but surprisingly, strongly stimulated transcription when added after sarkosyl. The same result was observed with a pulse-chase protocol in which accurately initiated RNA is labelled during a short pulse, followed by chase with excess unlabelled nucleoside triphosphates. RAP30 was required to label the transcript during the pulse, but RAP74 was not. RAP74, however, was required to observe the runoff RNA. Therefore, RAP30 and RAP74 appear to have separable functions in this extract system. The RAP30 subunit supported all initiation functions of TFIIF. The RAP74 subunit was essential for early elongation of RNAP II transcription.

To follow up these observations, transcription was initiated on an immobilized DNA template so that the shortest stable transcripts could be identified. Short nascent RNA complexes were found to be initiated accurately from the Adenovirus major late promoter and paused 11-20 nucleotides from the initiating base. Synthesis of short RNAs requires ATP hydrolysis, and both RAP30 and RAP74. These complexes are paused and not arrested because they can be quantitatively chased to the runoff position. In testing

the stability of these complexes in the presence of RAP74 and RAP74 mutants, these complexes were very stable if initiated in the presence of RAP74. Deletion of RAP74 C-terminal sequences decreased the stability of these complexes. The ability of RAP74 to stabilize short transcripts may partially explain the requirement for this factor in promoter escape.

The final chapter of this dissertation describes preliminary experiments with a protein isolated from yeast that may be a component of an alternate form of TFIIF. The amino acid sequence of Cdc73p has been compared to that of human RAP30 and bacterial sigma factors by Shi Min Fang of our laboratory. This sequence comparison allowed us to predict a region of Cdc73p that might be involved in RNAP II binding. Cdc73p was found to bind directly and stoichiometrically to polymerase, and a mutant with a 15 amino acid deletion within the predicted polymerase binding site was severely inhibited for RNAP II binding.

REFERENCES

Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M. and Ingles, C. J. (1988). Mol. Cell. Biol. 8, 321-329.

Allison, L. A. and Ingles, C. J. (1989). Proc. Natl. Acad. Sci. USA 86, 2794-2798.

Auble, D. T. and Hahn, S. (1993). Genes Dev. 7, 844-856.

Arias, J. and Dynan, W.S. (1989) . J. Biol. Chem. 264, 3223-3229.

Ayer, D.G., Auble D.T., deHaseth, P. L. (1989). J. Mol. Biol. 207, 749-756.

Barberis, A., Muller, C. W., Harrison, S. C., and Ptashne, M. (1993). Proc. Natl. Acad. Sci. USA 90, 5628-5632.

Bartolomei, M. S., Halden, N. F., Cullen, C. R. and Corden, J. L. (1988). Mol. Cell. Biol. 8, 330-339.

Berroteran, R. W., Ware, D. E., and Hampsey, M. (1994). Mol. Cell. Biol. 14, 226-237.

Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991). Mol. Cell. Biol. 11, 1195-1206.

Borukhov, S., Polyakov, A.,Nikiforov, V., and Goldfarb, A. (1992). Proc. Natl. Acad. Sci. USA 88, 8899-8902.

Borukhov, S., Sagitov, V., and Goldfarb, A. (1993). Cell 72, 459-466.

Brinkmann, U., Mattes, R.E. and Buckel, P. (1989). Gene 85, 109-114.

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

Burgess, R. R., Travers, A. A., Dunn, J. J., and Bautz, E. F. K. (1969). Nature 221, 43-46.

Burton, Z. F., Ortolan, L. G. and Greenblatt, J. (1986). EMBO J. 5, 2923-2930.

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

Burton, Z. F., Sopta, M., Ortolan, L. G. and Greenblatt, J. (1987). In: RNA polymerase and the regulation og transcription. edited by W. S. Rezikoff et al. (Elsevier Science Publication Co., Inc.), pp 417-420.

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

Buratowski, S., Sopta, M., Greenblatt, J. and Sharp, P.A. (1991). Proc. Natl. Acad. Sci. USA 88, 7509-7513.

Buratowski, S. and Zhou, H. (1992). Science 255, 1130-1132.

Buratowski, S. and Zhou, H. (1993). Proc. Natl. Acad. Sci. 90, 5633-5637.

Buratowski, S. (1994). Cell 77, 1-3.

Cadena, D. L. and Dahmus, M. E. (1987). J. Biol. Chem. 262, 12468-12474.

Cairns, B. R., Kim, Y.-J., Sayer, M. H., Laurent, B. C., and Kornberg, R. (1994). Proc. Natl. Acad. Sci. USA 91, 1950-1954.

Carey, M. (1994). Nature 368, 402.

Chamberlin, M. J. (1995). Harvey Lect. 88, 1-21.

Chambers, S. R., and Damus, E. M., (1994). J. Biol. Chem. 269, 26243-26248.

Chambers, S. R., Wang, B. Q., Burton, Z. F., and Damus, E. M., (1994). J. Biol. Chem. in press.

Chang, C. H., Kostrub, C. F. and Burton, Z. F. (1993). J. Biol. Chem. 268, 20482-20489.

Chang, C. H., and Burton, Z. F. (1995). Manuscript in preparation.

Choy, H. E. and Adhya, S. (1993). Proc. Natl. Acad. Sci. USA 90, 472-476.

Comai, L., Zomerdijk, J. C. B., Beckmann, H., Zhou, S., Admon, A., and Tjian, R. (1994). Science 266,1966-1972.

Comai, L., Tanese, N., and Tjian, R. (1992). Cell, 68, 965-976.

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

Conaway, J.W. and Conaway, R. C. (1990). Science 248, 1550-1553.

Conaway, J. W., Travis, E. and Conaway, R. C. (1990). J. Biol. Chem. 265, 7564-7596.

Conaway, J. W., Bradsher, J. N. and Conaway, R. C. (1992). J. Biol. Chem. 267, 8464-8467.

Conaway, R. C. and Conaway, J. W. (1993). Annu. Rev. Biochem. 62, 161-190.

Connelly, S., and Manely, J.L. (1988). Genes Dev. 2, 440-452.

Connelly, S., and Manely, J.L. (1989a). Cell 57.561-571.

Connelly, S., and Manely, J.L. (1989b). Mol. Cell. Biol. 9, 5254-5259.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Science 209, 1405-1414.

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

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

Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. and Kadonaga, J. T. (1991). Science 251, 643-649.

Cullen, B. R. (1990). Cell 63, 655-657.

Cullen, B. R. (1993). Cell 73, 417-420.

Dahmus, M. E. and Kedinger, C. (1983). J. Biol. Chem. 258, 2303-2307.

Dahlberg,J . E. and Lund, E. (1991). Science 254, 1462-1463.

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

DeJong, J. and Roeder, R. G. (1993). Genes Dev. 7, 2220-2234.

Drapkin, R., Merino, A., and Reinberg, D. (1993). Curr. Opin. Cell Biol. 5, 469-476.

Drapkin, R., Reardon, J., Ansari, A., Huang, J.-C., Zawel, L., Ahn, K., Sancar, A. and Reinberg, D. (1994a). Nature 368, 769-772.

Drapkin, R. and Reinberg, D. (1994a). Nature 369, 523-524.

Drapkin, R. and Reinberg, D. (1994b). TIBS. 19, 504-508.

Drombowski, A.J., Walter, W.A., Record, M.T. Jr., Siegele, D.A. and Gross, C.A. (1992) Cell 70, 501-512.

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

Dynlacht, B. D., Weinzierl, R., Admon, A. and Tjian, R. (1993). Nature 363, 176-179.

Fang, S. M., and Burton, Z. F. (1995). manuscript in preparation.

Feaver, W. J., Svejstrup, J. Q., Bardwell, L., Bardwell, J., Buratowski, S., Gulyas, K. D., Donahue, T. F., Friedberg, E. C. and Kornberg, R. D. (1993). Cell 75, 1379-1387.

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

Fisher, L., Gerard, M., Chalut, C., Lutz, Y., Humbert, S., Kanno, M., Chambon, P., and Egly, J. M. (1992). Science 257, 1392-1395.

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

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

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

Flores, O., Lu, H. and Reinberg, D. (1992). J. Biol. Chem. 267, 2786- .

Ge, H. and Roeder, R. G. (1994). Cell 78, 513-523.

Giardina, C., Perez-Riba, M. and Lis, J.T. (1992). Genes and Develop. 6, 2190-2200.

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

Gill, G., Pascal, E., Tseng, Z. and Tjian, R. (1994). Proc. Natl. Acad. Sci. USA 91, 192-196.

Gill, S.C., Weitzel, S.E. and von Hippel, P.H. (1991) J. Mol. Biol. 220, 307-324.

Goodrich, J. A., Hoey, T., Thut, A., Admon, A. and Tjian, R. (1993). Cell 75, 519-530.

Goodrich, J. A. and Tjian, R. (1994). Cell 77, 145-156.

Gralla, J.D. (1990). Methods Enzymol. 185, 37-54.

Greenblatt, J. (1991a). Cell 66, 1067-1070.

Greenblatt, J. (1991b). Trends Biochem. Sci. 16, 408-411.

Greenblatt, J., and Li, J. (1981). Cell 24, 421-428.

Greenblatt, J., Justin, R. N., and Mason, S. W. (1993). Nature 364, 401-406.

Gross, C.A., Lonetto, M. and Losick, R. (1992) Transcriptional Regulation, S. L. McKnight and K.R. Yamamoto, eds., Cold Spring Harbor Laboratory Press, pp. 129-178.

Gruzder, S N., Sung, P., Bailly, V., Prakash, L., and Prakash,S. (1994). Nature 369, 578-581.

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

Ha, I., Roberts, S., Maldonado, E., Sun, X., Kim, L. U., Green, M. and Reinberg, D. (1993). Genes Dev. 7, 1021-1032.

Ham, J., Steger, G., and Yaniv, M. (1992). FEBS 307, 81-86.

Hauber, J., and Cullen, B. R. (1988). J. Virol. 62, 673-679.

Helmann, J. D. and Chamberlin, M. J. (1988). Annu. Rev. Biochem. 57, 839-872.

Henry, N. L., Sayre, M. H. and Kornberg, R. D. (1992). J. Biol. Chem. 267, 23388-23392.

Henry, N. L., Campbell, A. M., Feaver, W. J., Poon, D., Weil, P. A., and Kornberg, R. D. (1994). Genes Dev.8, 2868-2878.

Hernandez, N. (1993). Genes Dev. 7, 1291-1308.

Hisatake, K., Hasekawa, S., Takada, R., Nakatani, Y., Horikoshi, M. and Roeder, R. G. (1993). Nature 362, 179-181.

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

Hoey, T., Weinzierl, R. O. J., Gill, G., Chen, J., Dynlacht, B. D. and Tjian, R. (1993). Cell 72, 247-260.

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

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

Imbalzano, A. N., Kwon, H., Green, M. R., and Kingston, R. E. (1994). Nature 370, 481-485.

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

Izban, M.G. and Luse, D. S. (1992a). J. Biol. Chem. 267, 13647-13655.

Izban, M.G. and Luse, D. S. (1992b). Genes Dev. 6, 134201356,

Izban, M.G. and Luse, D. S. (1993). J. Biol. Chem. 268, 12874-12885.

Jacob, G. A., Luse, S. W., and Luse, D. S. (1991). J. Biol. Chem. 266, 22537-22544.

Jacob, G. A., Kitzmiller, J. A., and Luse, D. S. (1994). J. Biol. Chem. 269, 3655-3663.

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

Jiang, Y. and Gralla, J. D. (1995). J. Biol. Chem. 270, 1277-1281.

Jiang, Y., Triezenberg, S., and Gralla, J. D. (1994). J. Biol. Chem. 269, 5505-5508.

Johnson, T. L. and Chamberlin, M. J. (1994). Cell, 77.217-224.

Joliot, V, Demma, M., and Prywes, R. (1995). Nature 373, 632-635.

Kan, C. M. (1994). In: Transcription: Mechanisms and Regulation. edited by R. C. Conaway and J. W. Conaway. (Raven Press, Ltd., New York), pp 279-298.

Karlin, S. (1993). Proc. Natl. Acad. Sci. 90, 5593-5597

Kash, S. F., Innis, J. W., Jacckson, A. U., and Kellems, R. E. (1993). Molec. Cell. Biol. 13, 2718-2729.

Kassaveetis, G. A., Joazeiro, C., Pisano, M., Geiduschek, P., Colbert, T., Hahn, S., and Blanco, J. A. (1992). cell 71, 1055-1064.

Kato, H., Sumimoto, H., Pognonec, P., Chen, C., Rosen, C. A. and Roeder, R. G. (1992). Genes Dev. 6: 655-666.

Kerppola, T. K., and Kane, C. M. (1988). Mol. Cell. Biol. 8, 4389-4394.

Kerppola, T. K., and Kane, C. M. (1990). Biochemistry 29, 269-278.

Kerppola, T. K., and Kane, C. M. (1991). FASEB J. 5,2833-2842.

Killeen, M. and Greenblatt, J. (1992). Molec. Cell. Biol. 12, 30-37.

Killeen, M., Coulombe, B. and Greenblatt, J. (1992). J. Biol. Chem. 267, 9463-9466.

Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993). Nature 365, 520-52 .

Kim,Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993). Nature 365, 512- 519.

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

Kim Y.-J., Bjorklund, S., Li, Y., Sayre, M. H. and Kornberg, R. D. (1994). Cell 77, 599-608.

Kitajima, S., Tanaka, Y., Kawaguchi, T., Nagaoka, T., Weissman, S. and Yasukochi, Y. (1990). Nucleic Acids Res. 18, 4843-4849.

Kokubo, T., Gong, D.-W., Yamashita, S., Horikoshi, M., Roeder, R. G. and Nakatani, Y. (1993). Genes Dev. 7, 1033-1046.

Kokubo, T., Gong, D.-W., Wootton, J. C., Horikoshi, M., Roeder, R. G. and Nakatani, Y. (1994). Nature 367, 484-487.

Koleske, A. J., Buratowski, S., Nonet, M., and Young, R. A. (1992). Cell 69, 883-894.

Koleske, A. J., and Young, R. A. (1994). Nature 368, 466-469.

Koleske, A. J., and Young, R. A. (1995). TIBS 20, 113-116.

Kretzschmar, M., Kaise, K., Lottspeich, F. and Meisterernst, M. (1994). Cell 78, 525-534.

Krumm, A., Meulia, T., Brunvand, M. and Groudine, M. (1992). Genes and Develop. 6, 2201-2213.

Krumm, A., Meulia, T., and Groudine, M. (1993). Bioassay 15, 659-665.

Krummel, B., and Chamberlin, M. J. (1989). Biochemistry 28, 7829-7842.

Laniox, J., and Acheson, N.H. (1988). EMBO J. 7, 2515-2522

Laspia, M F., Rice, A. P., and Mathews, M. B. (1989). Cell 59, 283-292.

Laspia, M F., Rice, A. P., and Mathews, M. B. (1990). Genes Dev. 4, 2397-2408.

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

Laybourn, P. J. and Kadonaga, J. T. (1992). Science 257, 1682-1685.

Lee, D. K., DeJong, J., Hashimoto, S., Horikoshi, M., and Roeder, R. G. (1992). Mol. Cell. Biol. 12, 5189-5196.

Lee, D. K., Horikoshi, M. and Roeder, R. G. (1991). Cell 67, 1241-1250.

Lieberman, P. M., and Berk, A. J. (1994). Genes Dev. 8, 995-1006.

Lin, Y. S., Carey, M. F., Ptashne, M. and Green, M. R. (1991). Nature 353, 569-571.

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

Linn, S. C. and Luse, D. S. (1991). Mol. Cell. Biol. 11, 1508-1522.

Lobo, S. M., Lister, J., Sullivan, M. L., and Hernandez, N. (1991). Genes Dev.5, 1477-1489.

Logan, J., Falk-Pederson,,Dranell, J. E., and Shenk, T. (1987). Proc. Natl. Acad. Sci. 84, 8306-8310.

Lu, H., Zawel, L., Fisher, L., Egly, J- M., and Reinberg, D. (1992). Nature 358, 641-645.

Luse, D. S. (1990). In: The Eukaryotic nucleus: Molecular biochemistry and Macmolecular assemblies. edited by Caldwell, N. J. (Telford Press).

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

Ma, D., Watanabe, H., Mermelstein., Admon, A., Oguri, K., Sun, X., Wada, T., Imai, T., Shiroya, T., Reinberg, D. and Handa, H. (1993). Genes Dev. 7, 2246-2257.

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

Manley, J. L., Fire, A., Cano, A., Sharp, P. A. and Gefter, M. (1980). Proc. Natl. Acad. Sci. 77, 3855-3859.

Marciniak, R. A., Garcia-Blanco, M. A., and Sharp, P.A. (1990). Proc. Natl. Acad. Sci. USA 87, 3624-3628.

Marciniak, R. A., and Sharp, P.A. (1989). EMBO J. 10, 4189-4196.

Marshall, N. F., and Price, D. H. (1992). Mol. Cell. Biol. 12, 2078-2090.

Matsui, T., Segall, J., Weil, A. P. and Roeder, R. G. (1980). J. Biol. Chem. 255, 11992-11996.

Maxon, M. E., Goodrich, J. A. and Tjian, R. (1994). Genes Dev. 8, 515-524.

McClure,W.R. (1985). Annu. Rev. Biochem. 54, 171-204.

McCracken, S. and Greenblatt, J. (1991). Science 253, 900-902.

Mecsas, J., Cowing, D. W., and Gross, C.A. (1991). J.Mol. Biol. 220, 585-597.

Meier, T. I., Keene, R. G., DeVit, M. J., Wang, D. and Landick, R. (1994). J. Cell. Biochem. Supplement 18C, Wiley-Liss, Inc., New York. pp 63.

Metzger, W., Schickor, P., and Heumann, H. (1989). EMBO J. 8, 2745-2754.

Meulia, T., Krumn, A., Spencer, C., and Groudine, M. (1992). Mol. Cell. Biol. 12, 4590-4600.

Merino, A., Madden, K. R., Lane, W. S., Champoux, J. J. and Reinberg, D. (1993). Nature 365, 227-232.

Moore, C.L. and Sharp, P. A. (1985). Cell 41, 845-855.

Muesing, M. A., Smith, D. H., and Capon, D. J. (1987). Cell 48, 691-701.

Muson, L. M. and Reznikoff, W. S. (1981). Biochemistry, 20, 2081-2085.

Mustaev, A., Kashlev, M., Zaychikov, E., Grachev, M., and Goldfarb, A. (1993). J. Biol. Chem. 26819185-19187.

Mustaev, A., Zaychikov, E., Severinov, K.,Kashlev, M.,Polyakov, A.,Nikifokov, V., and Goldfarb, A. (1994). Proc. Natl. Acad. Sci. USA 91, 12036-12040.

Myers, R. M., Tilly, K. and Maniatis, T. (1986). Science 232, 613-618.

Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Besnard, F., Roeder, R. G. and Freese, E. (1990). Nature 348, 86-88.

O'Brien, T., Hardin, S. E., Greenleaf, A. L., and Lis, T. J. (1994). Nature 370, 75-77.

Nonet, M., Sweetser, D. and Young, R. A. (1987). Cell 50, 909-915.

Ohkuma, Y., Sumimoto, H., Horikoshi, M. and Roeder, R. G. (1990). Proc. Natl. Acad. Sci. USA 87, 9163-9167.

Ohkuma, Y. and Roeder R. G. (1994). Nature 368, 160-163.

Ozer, J., Moore, P. A., Bolden, A. H., Lee A.,, Rosen, C. A., (1993). Genes Dev. 8, 2324-2335.

Paranjape, S. M., Kamakaka, R. T. and Kadonaga, J. T. (1994). Annu. Rev. Biochem. 63, 265-297.

Parvin, J. D. and Sharp, P. A. (1993). Cell 73, 533-540.

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

Pinto, I., Ware, D. E. and Hampsey, M. (1992). Cell 68, 977-988.

Pinto, I., Wu, W- H., Na, J. G., and Hampsey, M. (1994). J. Biol. Chem. 269, 30569-30573.

Poon, D. and Weil, A. (1993). J. Biol. Chem. 268, 15325-15328.

Proudfoot, N. J. (1989). TIBS. 14, 105-110.

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

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

Ranish, J. A. and Hahn, S. (1991). J. Biol. Chem. 266, 19320-19327.

Ranish, J. A., Lane, W. S., and Hahn, S. (1992). Science 255-1127-1129.

Reed, S. I., Ferguson, J., Jahng, K. -Y. (1988). Cold Spring Harb.Symp. Quant. Biol. 53, 621-627.

Reese, J., Apone, L., Walker, S. S., Griffin, L. A., and Green, M. R. (1994). Nature 371, 523-527.

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

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

Reines, D., Camberlin, M. J. and Kane, C. M. (1989). J. Biol. Chem. 264, 10799-10809.

Reins, D. (1994). Transcription: Mechanisms and Regulation. edited by R. C. Conaway and J. W. Conaway. Raven Press, Ltd., New York. pp 263-278.

Reines, D., Ghanouni, P., Li, Q. Q., and Mote, J. (1992). J. Biol. Chem. 267, 15516-15522.

Reines, D., Wells, D., Camberlin, M. J. and Kane, C. M. (1987). Mol. Cell. Biol. 8, 4389-4394.

Reynolds, R., Bermudez-Cruz, R. M., and Chamberlin, M. J. (1992). J. Mol. Biol. 224, 31-51.

Roberts, S. G., Ha, I., Maldonado, E., Reinberg, D. and Green, M. R. (1993). Nature 363, 741-744.

Rigby, P. W. Cell 72, 7-10.

Roberts, S., Purton, T., and Bently, D.L. (1992). Genes Dev. 6, 11562-1574.

Rougvie, A.E. and Lis, J.T. (1990). Mol. Cell. Biol. 10, 6041-6045.

Rudd, M.D., Izban, M.G., and Luse, D.S. (1994). Proc. Natl. Acad. Sci. USA 91, 8057-8061.

Ruppert, S., Wang, E. D. and Tjian, R. (1993). Nature 362, 175-179.

Samuels, M. A., Fire, A. and Sharp, P. A. (1982). J. Biol. Chem. 257, 14419-14427.

Sawadogo, M. and Roeder, R. G. (1984). J. Biol. Chem. 259, 5321-3526.

Sawadogo, M. and Sentenac, A. (1990). Annu. Rev. Biochem. 56, 711-754.

Serizawa, H., Conaway, J. W. and Conaway, R. C. (1993). Nature 363, 371-374.

Serizawa, H., Conaway, J. W. and Conaway, R. C. (1994a). In: Transcription: Mechanisms and Regulation. edited by R. C. Conaway and J. W. Conaway. (Raven Press, Ltd., New York), pp 27-43.

Serizawa, H., Conaway, J. W. and Conaway, R. C. (1994b). J. Biol. Chem. 269, 20750-20756.

Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M.,Sarasin, A., Weeda, G., Hoeijimaker, J. H. J., and Egly, J.-M. (1994). EMBO J. 13, 2388-2392.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulin, W., Hoeijimaker, J. H. J., Chambon, P. and Egly, J.-M. (1993). Science 260, 58-63.

Schultz, M. C., Reeder, R. H., and Hahn, S. (1992). Cell 69, 697-702.

Sharp, P. A., and Marciniak, R. A. (1989). Cell 59, 229-230.

Sharp, P. A. (1992). Cell 68, 819-821.

Simmen, K. A., Bernues, J., Parry, H. D., Stunnenberg, H. G., Berkenstam, A., Cavallini, B., Egly, J -M., and Mattaj, I. W. (1991). EMBO j. 10, 1853-1862.

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

Smale, S. T., Schmidt, M. C., Berk, A. J. and Baltimore, D. (1990). Proc. Natl. Acad. Sci. USA 87, 5268-5272.

Sopta, M., Carthew, R. W. and Greenblatt, J. (1985). J. Biol. Chem. 260, 10353-10361.

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

Southgate, C, and Green, M. R. (1991). Genes Dev. 5,2496-2507.

Spencer, C.A. and Groudine, M. (1990) Oncogene 5, 777-785.

Starr, D. B. and Hawley, D. K. (1991). Cell 67, 1231-1240.

Straney, D. C., and Crother, D. M. (1987). J. Mol. Biol. 193, 267-278.

Strobl, L. J. and Eick, D. (1992). *EMBO J* 11, 3307-3314.

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

Sun, Z.-W. and Hampsey, M. (1994). Submitted to Nature.

Sun, X., Ma, D. Sheldon, M., Yeung, K., and Reinberg, D. (1994). Genes Dev, 8, 2336-2348.

Sung, P., Higgins, D., Prakash, L., and Prakash,S. (1988). EMBO J. 7, 3263-3269.

Surratt, C. K., Milan, S. C., and Chamberlin, M. J. (1991). Proc. Natl. Acad. Sci. USA 88, 7983-7987.

Svejstrup, J. Q., Wang, Z.,Feaver, W. J., Wu, X., Bushnell, D. A., Donahue, T. F., Friedberg, E. C., and Kornberg, R. D. (1995). Cell 80, 21-28.

Tanese, N., Pugh, B. F. and Tjian, R. (1991). Genes Dev. 5, 2212-2224.

Tantin, D. and Carey, M. (1994). J. Biol. Chem. 269, 17397-17400.

Taggart, A., Fisher, T. S., and Pugh, B. F. (1992). Cell 71-1015-1028.

Thompson, C. M., Koleske, A. J., Chao,D. M., and Young, R. A. (1993). Cell 73, 1361-1375.

Thut, C. J., Chen, J -L., Klemm, R., and Tjian, R. (1995). Science 267, 100-104.

Timmers, T. Th. M. (1994). EMBO J. 13, 391-399.

Tjian, R. and Maniatis, T. (1994). Cell 77, 5-8.

Tschochner, H., Sayre, M. H., Flanagan, P. M., Feaver, W. J. and Kornberg, R. D. (1992). Proc. Natl. Acad. Sci. USA 89, 11292-11296.

Tyree, C. M., Geoge, C. P., Lira-DeVito, L. M., Wampler, S. L., Dahmus, M. E., Zawel, L. and Kadonaga, J. T. (1993). Genes Dev. 7, 1254-1265.

Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D. and Aloni, Y. (1992). Cell 69, 871-881.

Usheva, A. and Shenk, T. (1994). Cell 76, 1115-1121.

von Hippel, P.H., Baer, D. G., Morgan, W. D., and McSwiggen, J. A. (1984). Annue. Rev. Biochem. 53, 389-446.

von Hippel, P.H., Yager, T.D. and Gill, S.C. (1992) Transcriptional Regulation, S.L. McKnight and K.R. Yamamoto, eds., Cold Spring Harbor Laboratory Press, pp. 179-202.

Wade, P. A., Werel, W., Fentzke, R. C., Thompson, N. E., Leykam, J. F., Burgess, R. R., Jaehning, J. A., and Burton, Z. F. (1995). Submit to Proc. Natl. Acad. Sci.

Wampler, S. L. and Kadonaga, J. T. (1992). Genes Dev. 6, 1542-1552.

Wang, B. Q., and Burton, Z. F. (1995). Submit to J. Biol. Chem.

Wang, B. Q., Kostrub, C. F., Finkelstein, A. and Burton, Z. F. (1993). Protein Expr. Purif. 4, 207-214.

Wang, B. Q., Lei, L. and Burton, Z. F. (1994). Protein Expr. Purif. 5, 476-485.

Wang, D. and Hawley, D. K. (1993). Proc. Natl. Acad. Sci. 90, 843-847.

Wang, E. H. and Tjian, R. (1994). Science 263, 811-814.

Wang, W., Carey, M. and Gralla, J. D. (1992). Science 255, 450-453.

Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M., and Greenleaf, A. L. (1993). Genes Dev. 7,2329-2344.

Weil, P. A., Luse, D. S., Segall, J. and Roeder, R. G. (1979). Cell 18, 469-484.

Weinmann, R. (1992). Gene Expression 2, 81-91.

Weis, L, and Reinberg, D. (1992). FASEB J. 6,3300-3309.

Weinzierl, R. O. J., Dynlacht, B. D. and Tjian, R. (1993). Cell 362, 511-517.

Whitelaw, E., and Proudfoot, N. (1986). EMBO J. 5, 2915-2922.

Wiely, S. R., Kraus, R.J., and Mertz, J. (1992). Proc. Natl. Acad. Sci. USA 89, 5814-5818.

Wiest, D. K., Wang, D., and Hawley, K. D. (1992). J. Biol. Chem. 267, 7733-7744.

Workman, J. L., Roeder, R. G. and Kingston R. E. (1990). EMBO J. 9, 1299-1308.

Xiao, H., Pearson, A., Coulombe, B., Truant, R., Zhang, S., Regier, J., Triezenberg, S. J., Reinberg, D., Flores, O., Ingles, C. J. and Greenblatt, J. (1994). Mol. Cell. Biol. 14, 7013-7024.

Yamamoto,T., Horikoshi, M.,Wang, J., Hasegawa, S., Weil, P. A., and Roder, R. G. (1992). Proc. Natl. Acad. Sci. USA 89, 2844-2848.

Yankulove, K., Blau, J., Purton, T., Roberts, S. and Bentley, D. L. (1994). Cell 77, 749-759.

Yokomori, K., Admon, A., Goodrich, J. A., Chen, J.-L. and Tjian, R. (1993a). Genes Dev. 7, 2235-2245.

Yokomori, K., Zeidler, M. P., Chen, J.-L., Verrijzer, C. P. Mlodzik, (1994). Genes Dev. 8, 2313-2323.

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

Zawel, L., Lu, H., Cisek, L.J., Corden, J.L., and Reinberg, D. (1993). Cold Spring Harb.Symp. Quant. Biol. 58, 187-198.

Zawel, L. and Reinberg, D. (1993). Prog. Nucl. Acids Res. 44, 67-108.

Zenzie-Gregory, B., Khachi, A., Garraway, I. P., and Smale, S. T. (1993). Mol. Cell. Biol. 13, 3841-3849.

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

Zhou, Q., and Sharp, P. A., (1995). EMBO J. 14, 321-328.

Zhou, Q., Boyer, T. G., Berk, A.J. (1993). Genes Dev. 7, 180-187.

Zhu, H., Joliot, V. and Prywes, R. (1994). J. Biol. Chem. 269, 3489-3497.

CHAPTER II

RAP30/74 (TFIIF) IS REQUIRED FOR PROMOTER ESCAPE BY RNA POLYMERASE II

ABSTRACT

RAP30 and RAP74 are subunits of the transcription factor called variously RAP30/74, TFIIF, βγ and FC. This factor is required for accurate transcription by RNA polymerase II, in addition to other basal transcription factors. Using recombinant human RAP30 and RAP74, the functions of these subunits have been tested separately during the initiation and elongation phases of transcription. RAP30 is required to form a sarkosyl-resistant complex at 0.25% sarkosyl, so RAP30 is required for initiation. RAP74, however, stimulates transcription when added after sarkosyl, indicating that RAP74 is dispensible for initiation. The same result is obtained using a pulse-chase protocol in which accurately initiated RNA is labeled during a short pulse, followed by a chase with excess unlabeled nucleoside triphosphates. RAP30 is required in order to label the transcript during the pulse, but RAP74 is not. RAP74 must be added during the chase, however, in order to obtain a short run-off transcript. The following conclusions can be drawn from these experiments: 1) RAP30 is an initiation factor; 2) RAP74 is not required for ATP hydrolysis in initiation, which precedes phosphodiester bond formation; 3) RAP74 is not required for template strand separation; 4) RAP74 is not required to initiate phosphodiester bond formation; and 5) RAP74 is required for very early elongation.

INTRODUCTION

To accurately express genetic information, RNA polymerases select promoter sequences with precision. Because of this requirement, polymerases interact with initiation factors that direct them into a tight association with the promoter. RNA polymerases must then make a transition from being a sequence-specific DNA binding protein, with high selectivity for promoter sites, to an elongation enzyme that has little ability to discriminate between different sequences. The transition between a pre-initiation complex, an initiated complex and an elongation complex involves multiple steps, including: 1) template binding in the presence of initiation factors; 2) separation of DNA template strands; 3) formation of phosphodiester bonds; 4) alterations in polymerase conformation; 5) dissociation of polymerase from sequence-specific DNA binding initiation factors; and 6) stabilization of an elongation conformation of RNA polymerase by elongation factors (reviewed in 1, 2).

In the Gram-negative eubacterium *E. coli*, mechanistic details of the transition from an initiated to an elongation complex have been elucidated. *E. coli* RNA polymerase initiates transcription in the presence of a sigma initiation factor. The primary sigma factor in *E. coli* is called σ^{70}; this initiation factor has an RNA polymerase binding domain and sequence-specific DNA-binding domains that separately recognize the -10 (TATAAT) and -35 (TTGACA) regions of promoters. σ^{70} first binds to RNA polymerase before associating with the promoter. Binding of sigma releases RNA polymerase from non-specific sites on DNA, and binding to polymerase alters σ^{70} conformation to expose DNA-binding domains (1, 3). In the presence of its sigma factor, RNA polymerase binds tightly to promoter DNA, first to form the closed complex, which is later opened by separation of the template DNA strands. At many promoters RNA polymerase then synthesizes numerous short transcripts before entering a productive elongation mode. This process is termed abortive initiation: a process that represents

repeated re-initiation by RNA polymerase without escape from the promoter. A change in RNA polymerase conformation is required that causes σ^{70} release, allowing polymerase escape (2, 4). This elongation conformation of RNA polymerase is stabilized by binding elongation factors such as NusA protein.

Despite evolutionary conservation of RNA polymerases and domains of some initiation factors, the initiation mechanism and the transition from an initiated complex to an elongation complex is less well understood in mammalian transcription. Human RNA polymerase II must interact with a number of general factors, including TFIID, TFIIB, TFIIF (RAP30/74), TFIIE, TFIIH and TFIIJ, in order to accurately initiate transcription from a promoter (reviewed in 5). TFIID binds to the TATAAA region of the promoter followed by association of TFIIB (DB complex). RNA polymerase II binds first to the RAP30 subunit of TFIIF (6). This binding suppresses non-specific DNA binding by RNA polymerase II (7, 8) and is required for stable association of RNA polymerase II with the promoter (6, 9). In keeping with its σ^{70}-like functions, human RAP30 is homologous to σ^{70} within its RNA polymerase-binding domain (10, 11). DNA recognition roles of σ^{70} appear to be a function of TFIID and TFIIB and may be reflected in a weak sequence similarity between the TATAAA binding domain of TFIID and the TATAAT binding domain of σ^{70} (12). After formation of the DBPolF complex (complex of template DNA, TFIID, TFIIB, RNA polymerase II and TFIIF), TFIIE, TFIIH and TFIIJ can be incorporated, forming the completely assembled pre-initiation complex (5). ATP hydrolysis is required by RNA polymerase II prior to template strand separation (13) and phosphodiester bond formation (14). After synthesis of the first phosphodiester bonds, RNA polymerase must escape the strong protein-protein and protein-DNA contacts holding it to the promoter. The mechanism by which this occurs is not known. Factors that remain bound at the promoter as polymerase exits, factors that dissociate from polymerase and the promoter, and factors that remain associated with polymerase during elongation have not been clearly determined.

One feature of the transition from a pre-initiation complex to an elongation complex has been proposed to involve covalent modification of RNA polymerase II (reviewed in 5, 15). At the COOH-terminal end of the largest subunit of RNA polymerase II (homolog of the ß' subunit of *E. coli* RNA polymerase) is an unusual carboxy terminal domain (CTD) which consists of 52 repeats of the consensus heptapeptide sequence YSPTSPS. This domain is multiply phosphorylated at the two SP sequences on serine. The dephosphorylated form of RNA polymerase II is referred to as the IIa form, and the multiply phosphorylated form as the IIo form. RNA polymerase II binds to the DB complex in the IIa form, at least in part because the dephosphorylated CTD binds to the TBP (TATA binding protein) subunit of TFIID (16). Within the pre-initiation complex, the CTD becomes phosphorylated by a stably associated CTD kinase (17). The phosphorylated CTD does not bind TBP, facilitating release of RNA polymerase II from the promoter (16).

Phosphorylation of the CTD does not account for the ATP requirement in initiation because: 1) accurate transcription using RNA polymerase II lacking a CTD requires ATP hydrolysis; 2) GTP will support CTD phosphorylation but not initiation; and 3) the ATP concentration requirement for initiation is significantly higher than that for conversion of IIa polymerase to the IIo form (16). The IIo form is the normal elongation form of RNA polymerase II.

In this report, we identify another feature of the transition from the pre-initiation complex to the initiated complex and elongation complex. This transition involves the activity of the heteromeric transcription factor RAP30/74. We demonstrate that the RAP30 subunit will support all of the initiation functions of this factor. The RAP74 subunit is, however, dispensible for initiation functions and instead is required for promoter escape by RNA polymerase II.

MATERIALS AND METHODS

DNA templates for transcription assays

Plasmid pSmaF contains the SmaI-F fragment of Adenovirus-2 subcloned into the SmaI site of pBR313 (18). When this plasmid is digested with SmaI, the accurately initiated transcript from the Adenovirus major late promoter (AdMLP) is 536 nucleotides in length. pSmaF digested with SmaI was used as the template for the experiments shown in Figures 1-4.

Plasmid pML was constructed by subcloning the AdMLP as an XhoI to HindIII fragment (coordinates -256 to +196 relative to the AdMLP cap site) between the XhoI and HindIII sites of the vector pBluescript II KS (+) (Stratagene). The resulting plasmid has a single SmaI site located at +217 relative to the AdMLP cap site. This template digested with SmaI was used for the experiments shown in Figures 5 and 6.

Extract systems for in vitro transcription

An extract derived from HeLa cell nuclei was prepared as described in Shapiro et al. (19). A RAP30/74 depleted extract was prepared as previously described (20, 21). Affinity-purified anti-RAP30 antibodies were mixed with the extract and then the solution was passed through a protein A-Sepharose column to remove RAP30/74 and antibodies from the solution. This antibody depletion was repeated a second time to remove any residual RAP30/74 from solution. This solution was then passed through two protein A-Sepharose columns to remove any remaining RAP30/74 and anti-RAP30 antibodies from the extract. The extract was then concentrated by centrifugation using a Centricon-10 microconcentrator (Amicon). The resulting extract is functionally depleted of both RAP30 and RAP74 (Figures 1-6). The extent of RAP30 depletion of the RAP30/74 depleted extract has previously been documented (20).

Recombinant RAP30 and RAP74

Methods for preparation of recombinant human RAP30 and RAP74 have been published elsewhere (22). RAP30 used in these experiments was purified to near homogeneity. RAP74 used in these experiments is a mixture of full length RAP74 and RAP74 fragments. RAP74 concentrations were estimated by Coomassie blue staining of gels and reflect the amount of the intact protein.

In vitro transcription assays

The methods for in vitro transcription have been published previously (23). Reactions were done at 30 ^{o}C. AdMLP DNA was at 60 μg/ml in all reactions. Pre-incubations were done in 20 μl. For Sarkosyl block and pulse-chase reactions, initiating nucleoside triphosphates were added in 2 μl. For sarkosyl block procedures, 2 μl of a 4 % sarkosyl solution was added to reactions, to a concentration of 0.33%. Elongating nucleoside triphosphates were added in 5 μl for all procedures. After addition of elongating nucleoside triphosphates, sarkosyl was diluted to 0.28 %. Detailed procedures for individual experiments are given in the figures and figure legends. α-amanitin was added to reactions to a final concentration of 1 μg/ml where indicated. Purification of RNA for electrophoresis has been described in detail previously (23). Transcripts were resolved in 6% polyacrylamide gels containing 50% (w/v) urea, and visualized by autoradiography. Quantitation of accurate transcription was done using a Molecular Dynamics Phosphorimager as described in (22).

RESULTS

RAP30 and RAP74 are required for accurate transcription from the adenovirus major late promoter

A RAP30/74-depleted extract (RAP30/74-DE) was prepared by immunodepletion using anti-RAP30 antibodies (20, 21). This extract was not active for accurate transcription unless it was supplemented with RAP30 and RAP74 (Figure 1). The template for runoff transcription is Adenovirus major late promoter (AdMLP) DNA digested with restriction endonuclease SmaI. The runoff transcript from the AdMLP is 536 nucleotides. In this report, RAP30 and RAP74 used in reconstitution assays were produced in *E. coli* using human cDNAs subcloned into bacteriophage T7 expression vectors (22). When RAP30 was omitted from the reconstituted system, no 536 transcript was observed (lanes 5, 7, 9 and 11). Similarly, omission of RAP74 from the reconstituted system resulted in loss of the 536 runoff transcript (lanes 5 and 6). Apparently both subunits of this factor are essential for initiation and/or elongation of this short runoff transcript.

The level of transcription observed in the fully reconstituted system is similar to that observed for the intact transcription system before immunodepletion (compare lane 12 with lanes 1-4). Addition of RAP30 and RAP74 to the intact system did not stimulate transcription substantially (lanes 2-4), indicating that these factors are not limiting in the untreated extract.

To test whether the requirement for RAP30 and RAP74 was during the initiation or elongation phases of transcription, RAP30 and RAP74 were added back to a RAP30/74 depleted extract in sarkosyl block and pulse-chase protocols (see below).

Figure 1. Both RAP30 and RAP74 are required for accurate transcription from the AdMLP. Additions to each reaction are indicated at the top of the figure. NE) Extract derived from HeLa cell nuclei; RAP30/74 DE) RAP30/74 depleted extract. 100 ng recombinant RAP30 was added where indicated. Amounts of RAP74 added into various reactions are in ng. The template for runoff transcription was pSmaF DNA digested with SmaI. The accurately intiated runoff transcript from the AdMLP is 536 nucleotides. The reaction protocol is shown at the bottom of the figure.

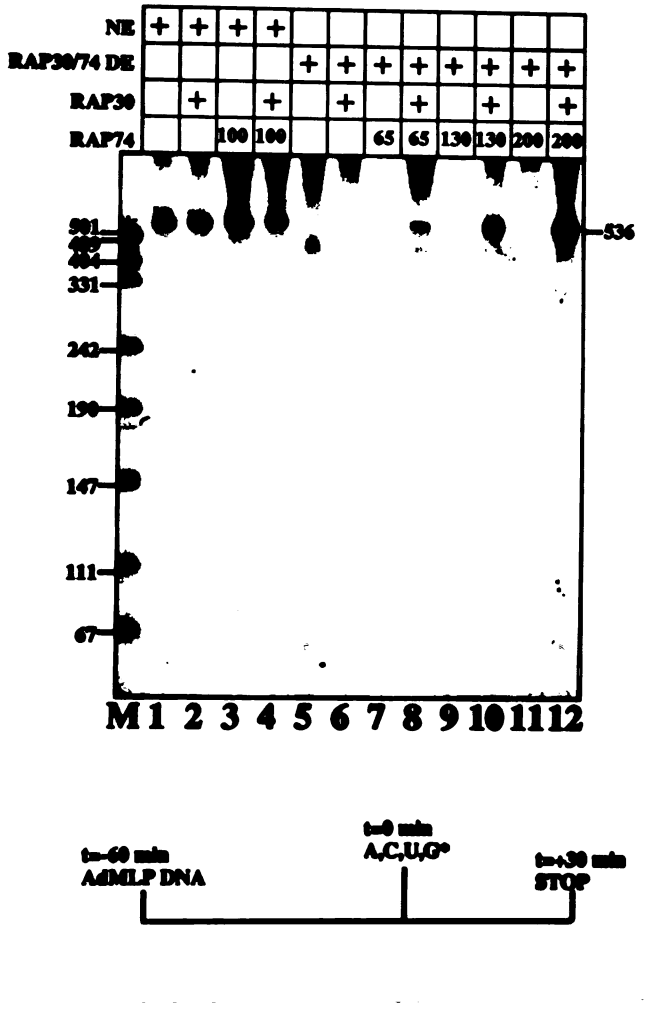

NE
RAP30/74 DE
RAP30
RAP74
100
100
65
65
130
130
200
200
536
331
242
190
147
111
67
M 1 2 3 4 5 6 7 8 9 10 11 12
t=-60 min
AdMLP DNA
t=0 min
A,C,U,G*
t=+30 min
STOP

Figure 1

RAP

sark

elim

sark

form

initi

bee

req

is n

fact

abs

sar

the

act

the

hig

pro

nu

nu

(la

5).

ac

tra

RAP30 and RAP74 are required for accurate transcription in the presence of sarkosyl

At concentrations higher than 0.25 %, the ionic detergent sarkosyl completely eliminates re-initiation by RNA polymerase II (24). The requirements for forming a sarkosyl-resistant complex are: 1) hydrolysis of the β–γ phosphate bond of ATP; and 2) formation of the first few phosphodiester bonds of the RNA chain (24, 25). Since initiation must precede formation of the sarkosyl-resistant complex, this detergent has been used to discriminate between initiation and elongation functions: a factor that is required for transcription prior to addition of sarkosyl is an initiation factor; a factor that is required for transcription but can function when added after sarkosyl is an elongation factor, since no new initiation can occur after addition of the detergent.

Because the requirements for transcription elongation might be different in the absence or presence of sarkosyl, we tested the requirement for RAP30 and RAP74 in a sarkosyl block protocol (Figure 2). When the transcript is elongated in the presence of the detergent, both RAP30 and RAP74 must be added to restore accurate transcriptional activity to a RAP30/74 depleted extract. The reaction scheme is shown at the bottom of the figure. The fully reconstituted system is shown in lanes 1 and 2. In the presence of high sarkosyl, RNA polymerase II pauses at position +186 from the adenovirus major late promoter (AdMLP) (24). The runoff transcript, which is not observed, would be 536 nucleotides in length. In the presence of 1 μg/ml α-amanitin, production of the 186 nucleotide transcript is abolished, showing that it was synthesized by RNA polymerase II (lane 3). In the absence of RAP30 and RAP74 no transcript was produced (lanes 4 and 5). Neither RAP30 (lane 6) nor RAP74 (lane 7) alone restored accurate transcriptional activity to the depleted extract. Therefore, both RAP30 and RAP74 are required for transcription in the presence of sarkosyl.

Figure 2. Both RAP30 and RAP74 are required for accurate transcription in the presence of sarkosyl. HK + glc) 10 μU hexokinase and 50 μM D-glucose; Amanitin) 1 μg/ml α-amanitin. The 186 nucleotide paused, accurately initiated transcript from the AdMLP is indicated. RAP30/74 depleted extract was incubated with AdMLP template. RAP30 (100 ng), RAP74 (200 ng) and other additions were made at t = -60 min. Initiating nucleotides: 100 μM ATP, CTP and UTP (t = -1 min). Sarkosyl was added to 0.33%. Elongating nucleotides: 600 μM ATP, CTP, UTP and 25 μM α^{32}P-GTP. Elongation was allowed to proceed for 40 min.

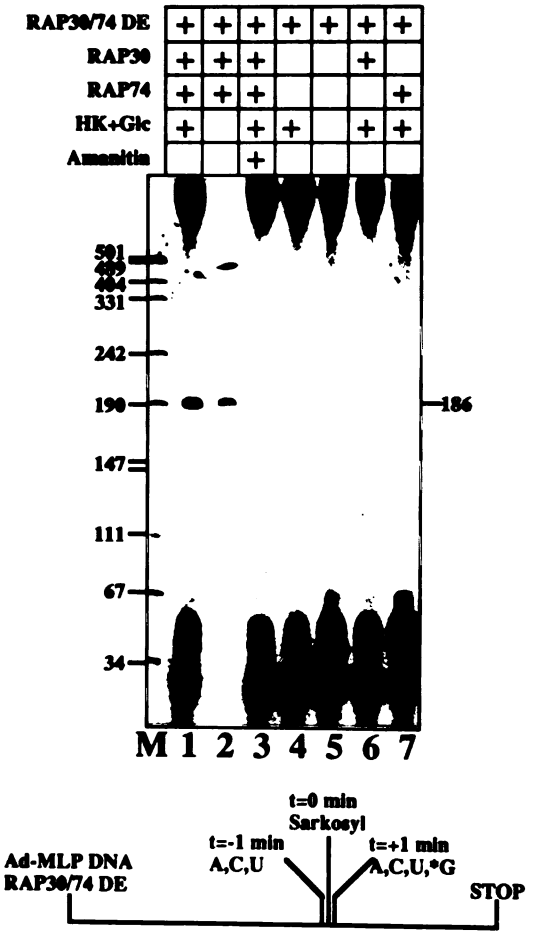

RAP30/74 DE
RAP30
RAP74
HK+Glc
Amanitin
501
404
331
242
190
147
111
67
34
186
M 1 2 3 4 5 6 7
t=0 min
Sarkosyl
t=-1 min
A,C,U
t=+1 min
A,C,U,*G
Ad-MLP DNA
RAP30/74 DE
STOP

Figure 2

glucos

confir

treatm

ATP f

ATP a

In the

added

Requ

at var

is sho

AdM

allow

MLP

ATP

radio

addit

scann

was

analy

kinet

previ

incor

Since initiation of transcription requires ATP hydrolysis, hexokinase and D-glucose were added to some of the reactions (lanes 1, 3, 4, 6 and 7) to deplete ATP and confine initiation events to a 1 min window prior to the addition of sarkosyl. This treatment renders the extract completely dependent on the addition of β–γ hydrolyzable ATP for initiation (data not shown). Without these additions, low levels of contaminating ATP and other nucleoside triphosphates might allow initiation during the pre-incubation. In the experiments shown in Figures 3 - 6, hexokinase and D-glucose (50 μM) have been added to the extract prior to the pre-incubation.

Requirement for RAP30/74 function before sarkosyl addition

In Figure 3A, recombinant RAP30/74 was added to a RAP30/74 depleted extract at various times before or after addition of sarkosyl (open squares). The reaction protocol is shown at the bottom of the figure. The RAP30/74 depleted extract was incubated with AdMLP template for 1 hr. At time = -1 min, 100 μM ATP, CTP and UTP were added to allow ATP hydrolysis and formation of the first 9 phosphodiester bonds from the Ad-MLP (pppACUCUCUUCCG). Sarkosyl is added at t = 0 min. At t = +1 min, 600 μM ATP, CTP and UTP are added along with 25 μM α-^{32}P GTP to allow elongation and radiolabeling of the transcript. RAP30/74 was added at the indicated times before or after addition of sarkosyl. The +186 AdMLP transcript was quantitated using a β-radiation scanner, and accurate transcription is reported in arbitrary units.

Strong stimulation of accurate transcription was only observed when RAP30/74 was added before addition of sarkosyl, so RAP30/74 is an initiation factor. Based on this analysis, either RAP30, RAP74 or the RAP30/74 complex is required for initiation. The kinetics of stimulation of transcription by recombinant RAP30/74 is very similar to that previously reported for human RAP30/74 (20). The recombinant proteins are efficiently incorporated into the pre-initiation complex in a rapid step, late in the assembly pathway.

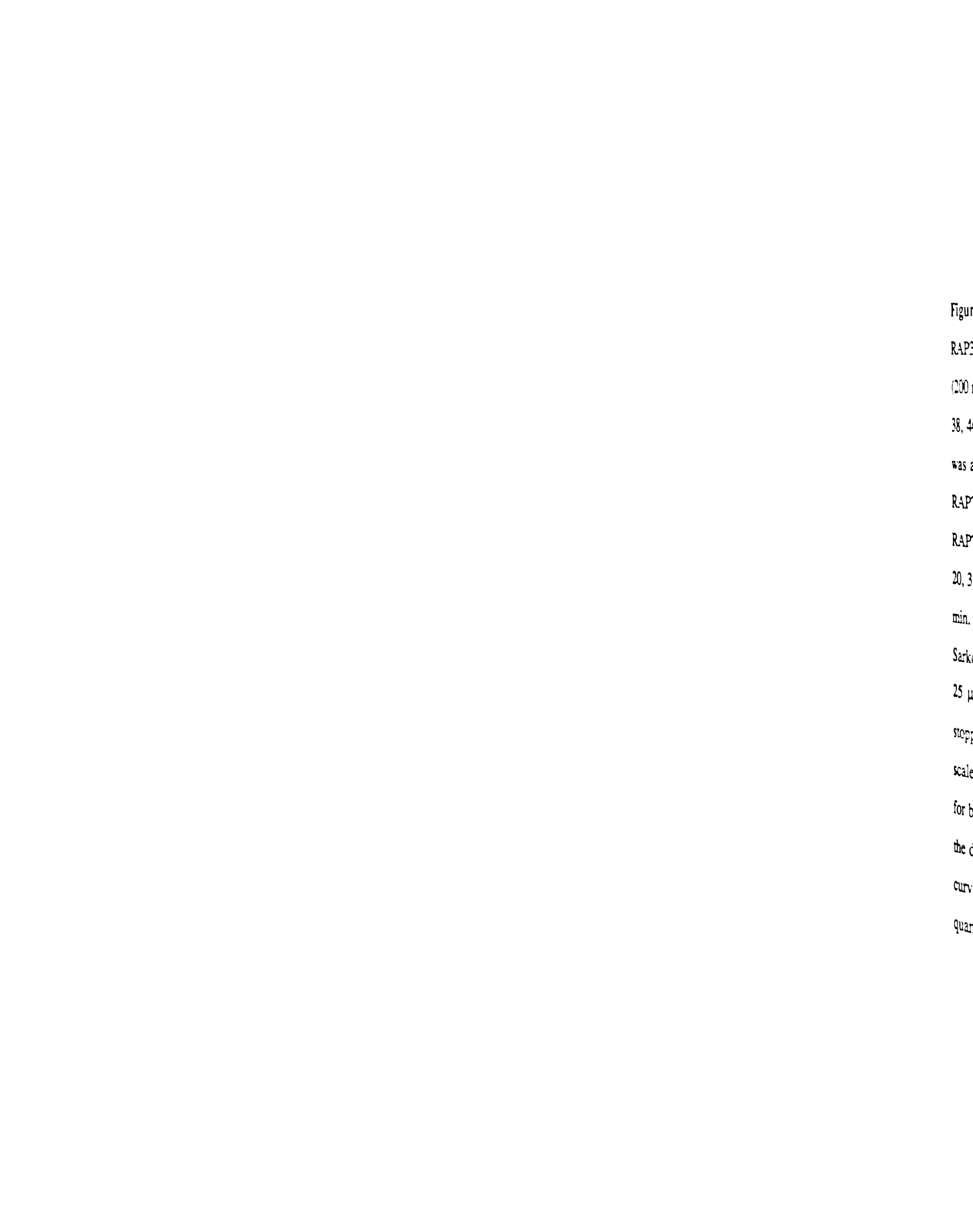


Figure 3. Kinetics of RAP30 and RAP74 function in transcription. Panel A) RAP30/74 and RAP30 are initiation factors. (open squares) RAP30 (100 ng) and RAP74 (200 ng) were added at the indicated times (t = -60, -40, -20, -10, -5, -2, -1, 0, 1, 2, 5, 10, 38, 40 min). (closed circles) RAP74 (200 ng) was added at -60 min; RAP30 (100 ng) was added at the indicated times (t = - 10, -5, -2, -1, 0, 1, 2, 5, 10, 40 min). Panel B) RAP74 is an elongation factor. (open circles) RAP30 (50 ng) was added at t = -60 min; RAP74 was added at the indicated times (t = -60, -40, -20, -10, -5, -2, -1, 0, 1, 2, 5, 10, 20, 38, 40 min). A RAP30/74 depleted extract was incubated with AdMLP DNA for 60 min. At t = -1 min, ATP, CTP and UTP were added to 100 μM to initiate transcription. Sarkosyl was added to 0.33 % at t = 0 min. At t = +1 min, 600 μM ATP, CTP, UTP and 25 μM α^{32}P-GTP were added to allow elongation of transcription. Reactions were stopped at t = +40 min. The data presented in B is the combination of two experiments scaled together by making the -10 min time points equivalent in transcriptional activity for both experiments. Apparent scatter in the data reflects a slightly different kinetics in the dip of transcriptional activity seen near the time of sarkosyl addition. The shapes of curves for both experiments were very similar. The paused 186 nucleotide transcript was quantitated as described in Materials and Methods.

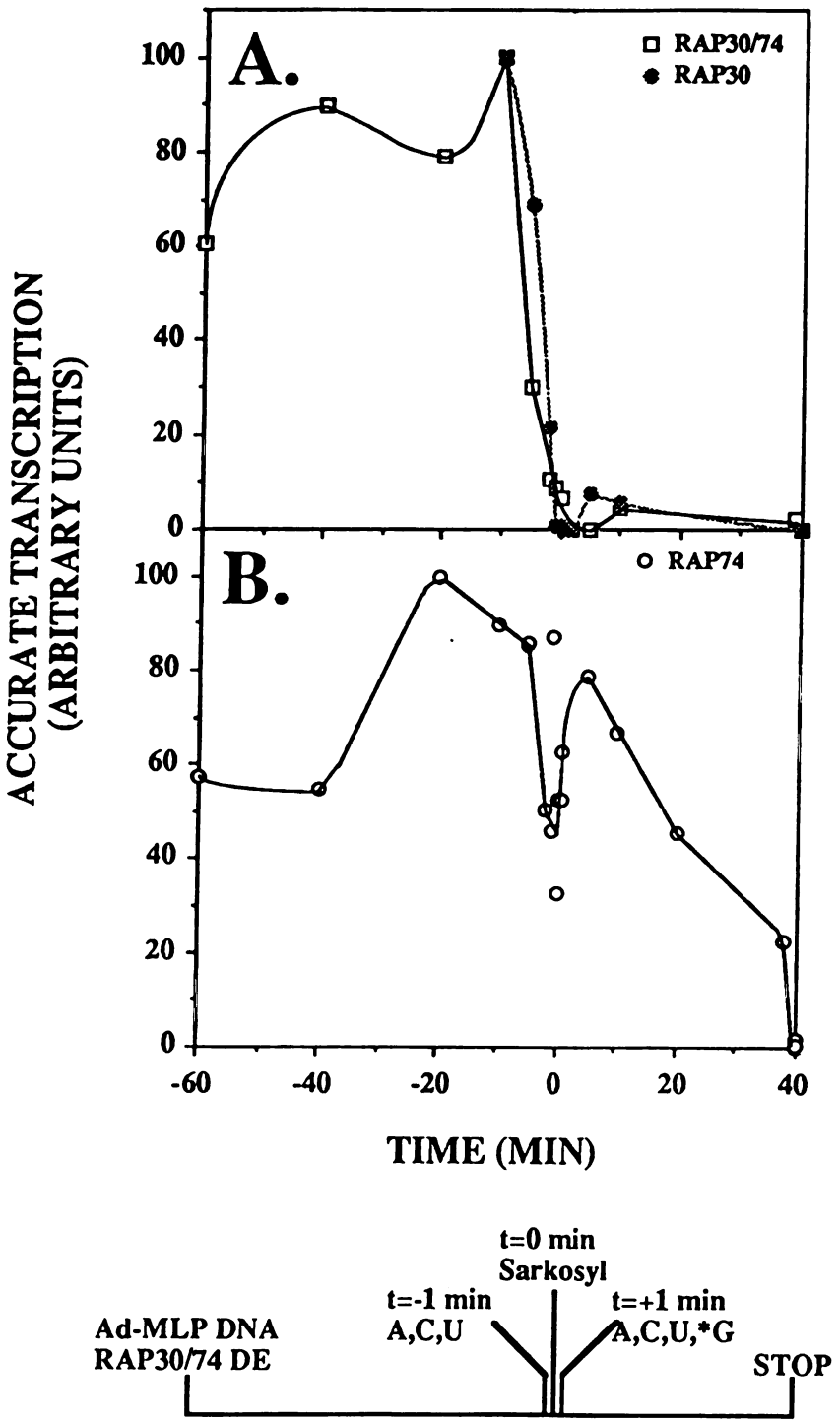

A.
RAP30/74
RAP30
B.
RAP74
ACCURATE TRANSCRIPTION
(ARBITRARY UNITS)
100
80
60
40
20
0
-60
-40
-20
0
20
40
TIME (MIN)
t=0 min
Sarkosyl
t=-1 min
A,C,U
t=+1 min
A,C,U,*G
Ad-MLP DNA
RAP30/74 DE
STOP

Figure 3

RAP3

t = -6

sarkos

prior

into t

kineti

previ

RAP

RAP

but

sarko

elong

mixe

time

time

obse

was

elon

poin

sarko

Acc

In Figure 3A, we show that RAP30 is an initiation factor (filled circles). RAP30/74 depleted extract, supplemented with RAP74, was mixed with template DNA at t = -60 min. RAP30 was added to the reaction at various times before or after addition of sarkosyl. Accurate transcription was strongly stimulated only when RAP30 was added prior to sarkosyl, so RAP30 is an initiation factor. The kinetics of RAP30 assimilation into the pre-initiation complex is indistinguishable from assembly of RAP30/74. The kinetics of assembly of recombinant RAP30/74 and RAP30 is indistinguishable from that previously published for human RAP30/74 (20).

RAP74 is dispensible for initiation but required for elongation

A more startling observation was made in the experiment shown in Figure 3B. RAP74 was required for accurate transcription, as previously shown in Figures 1 and 2, but RAP74 strongly stimulated transcription whether it was added before or after sarkosyl. By this analysis, RAP74 is not required for initiation; rather, RAP74 is an elongation factor.

In this experiment, RAP30/74 depleted extract, supplemented with RAP30, was mixed with template DNA and pre-incubated for 1 hr. RAP74 was added at various times before or after addition of sarkosyl (open circles). When RAP74 was added at the time of sarkosyl addition or after sarkosyl addition, significant accurate transcription was observed. If RAP74 was never added (t = +40 min), no transcription from the AdMLP was observed. Even when RAP74 was added at t = +38 min, leaving only 2 min for elongation, 20 % of the +186 transcript was observed relative to the t = -10 min time point. Apparently, RAP74 can function in early elongation even in the presence of sarkosyl.

Most surprisingly, however, RAP74 is dispensible for initiation functions. According to this analysis, RAP74 is an elongation factor that is absolutely required to

synthesize a short (186 nucleotide) transcript. RAP74 is, therefore, not required for: 1) pre-initiation complex formation; 2) ATP hydrolysis in initation; 3) template strand separation; or 4) first bond formation. It is only required after these processes are completed, and the sarkosyl-resistant complex is formed. RAP74 is, however, required to synthesize a short runoff transcript from the AdMLP. RAP74 appears to be required for promoter escape by RNA polymerase II.

The dip in transcriptional activity seen when RAP74 is added along with sarkosyl (t = 0 min) or one minute after sarkosyl (t = +1 min) has been reproducible in replicate experiments. Apparently, RAP74 activity is most sensitive to sarkosyl at the time of early transcript elongation. When added before initiation (t = -20 to t = -5 min), RAP74 may be protected from sarkosyl inhibition by incorporation into the pre-initiation complex (6). Also, RAP74 may have completed its function in promoter escape before addition of detergent. More puzzling is the increase in transcriptional activity when RAP74 is added just after sarkosyl (t = +2 to +10 min). Perhaps sarkosyl dissociates a factor that destabilizes the ternary complex (ie. TFIIS, see Discussion). Also, the effective concentration of sarkosyl may decrease with time after addition, as detergent binds proteins in the extract. Detergent concentration was slightly decreased by addition of nucleoside triphosphates at t = +1 min. The decrease in accurate transcription seen when RAP74 was added at t = +5 to +38 min probably reflects both instability of ternary complexes and shorter elongation times. Apparent scatter in data points around the time of addition of sarkosyl is due to combination of data from two separate experiments to produce the graph reported in Figure 3B. Scatter represents slightly different kinetics of the dip in transcriptional activity around the time of sarkosyl addition. In any case, all of the data collected in replicates of this experiment supports our major conclusion, that RAP74 is an elongation factor (see also Figure 4).

In Figure 4, the surprising result that RAP74 is dispensible for initiation functions is presented as gel data. RAP30/74 depleted extract, supplemented with RAP30, was

Figure 4. RAP74 is an elongation factor. The reaction protocol was the same as shown in Figure 3B. RAP30 (100 ng) was added at t = -60 min; RAP74 (200 ng) was added at t = -10 min, +10 min or omitted altogether, as indicated.

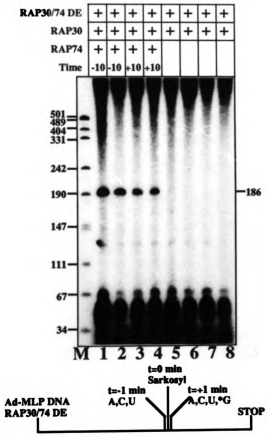

RAP30/74 DE
RAP30
RAP74
Time
-10 -10 +10 +10
501
489
404
331
242
190
147
111
67
34
186
M 1 2 3 4 5 6 7 8
t=0 min
Sarkosyl
t=-1 min
A,C,U
t=+1 min
A,C,U,*G
Ad-MLP DNA
RAP30/74 DE
STOP

Figure 4

mixed w

(lanes 1

min). I

transcri

after ad

was ob

Confir

in Fig

instan

absen

exper

inclu

pre-i

adde

figur

(625

ATP

labe

= 0

nuc

mir

cha

mixed with AdMLP template DNA. RAP74 was added to the extract at t = -10 min (lanes 1 and 2) or at t = +10 min (lanes 3 and 4), relative to addition of sarkosyl (t = 0 min). In lanes 5-8 RAP74 was never added to the reaction (t = +40 min). Accurate transcription was strongly stimulated when RAP74 was added to the reaction before or after addition of sarkosyl. In the absence of RAP74 addition, no accurate transcription was observed. Clearly, RAP74 is an elongation factor.

Confirmation of the sarkosyl results using a pulse-chase protocol

One possibility that we wished to investigate was whether the observations made in Figures 3 and 4 were somehow limited to transcription reactions done in sarkosyl. For instance, sarkosyl might stabilize a short ternary complex that would be unstable in the absence of detergent. RAP74, for instance, might appear to be an initiation factor in an experiment that did not include sarkosyl.

In Figure 5, RAP30 is shown to be an initiation factor in a protocol that does not include sarkosyl. The RAP30/74 depleted extract was mixed with template DNA and pre-incubated for 60 min. RAP74 was added to the reaction at t = -5 min. RAP30 was added to the reaction at t = -5 min, t = +5 min or omitted altogether, as indicated in the figure. Initiating nucleoside triphosphates ATP, UTP (100 μM each) and α-^{32}P CTP (625 nM) were added at t = -1 min. At t = 0 min elongating nucleoside triphosphates ATP, UTP, CTP and GTP (1 mM each) were added. The >1000 fold excess of unlabeled CTP added during the chase eliminated radiolabeling of transcripts initiated after t = 0 min. The accurately initiated runoff transcript from the AdMLP in this assay is 217 nucleotides.

Accurate transcription was strongly stimulated when RAP30 was added at t = -5 min, before the pulse and chase (lanes 3 and 4). Because of high backgrounds in pulse-chase experiments, critical points were done in duplicate. This 217 nucleotide transcript

Figur

added

min, a

min.

μCi).

is 217

Figure 5. Pulse-chase protocol: RAP30 is an initiation factor. RAP74 (200 ng) was added at t = -5 min, as indicated; RAP30 (100 ng) was added at t = - 5 min or at t = +5 min, as indicated. A RAP30/74 depleted extract was mixed with AdMLP DNA at t = -60 min. Pulse nucleoside triphosphates: 100 μM ATP, UTP and <1 μM α^{32}P-CTP (10 μCi). Chase: 1 mM ATP, CTP, UTP and GTP. The accurately initiated runoff transcript is 217 nucleotides.

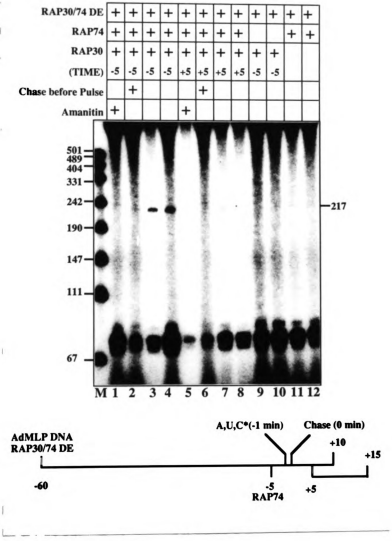

RAP30/74 DE
RAP74
RAP30
(TIME)
Chase before Pulse
Amanitin
501
489
404
331
242
190
147
111
67
217
M 1 2 3 4 5 6 7 8 9 10 11 12
AdMLP DNA
RAP30/74 DE
A,U,C*(-1 min)
Chase (0 min)
+10
+15
-60
-5
RAP74
+5

Figure 5

is a

and

sho

new

did

prio

trans

mad

initia

nucle

pulse

the in

supple

RAP7

+5 mi

before

RAP7

after s

transcr

transcr

6) abo

by add

is abolished in the presence of 1 μg/ml α-amanitin (lane 1). When the order of the pulse and chase was reversed (lane 2), this also abolished detection of the accurate transcript, showing that sufficient cold CTP was added during the chase to eliminate labeling of newly initiated transcripts. Addition of RAP30 at t = +5 min, after the pulse and chase, did not strongly stimulate transcription (lanes 7 and 8). Since RAP30 must be added prior to addition of chase nucleotides, in order to incorporate ^{32}P-CMP into the transcript, RAP30 must be an initiation factor. These results confirm the observations made in Figure 3A using a sarkosyl block protocol: RAP30 is required for accurate initiation of transcription.

Omission of RAP74 from the reaction (lanes 9 and 10) also abolished the 217 nucleotide transcript, showing that RAP74 is required for accurate transcription in the pulse-chase protocol.

In the experiment shown in Figure 6, RAP74 is tested for its requirement during the initiation and elongation phases of transcription. The RAP30/74 depleted extract, supplemented with RAP30, was mixed with template DNA and pre-incubated for 60 min. RAP74 was added to the reaction at either t = -5 min (before the pulse and chase) or at t = +5 min (after the chase). RAP74 strongly stimulated transcription whether it was added before the pulse (lanes 3 and 4) or after the chase (lanes 7 and 8). From this analysis, RAP74 is an elongation factor. RAP74 is required for accurate transcription, but only after synthesis of an accurately initiated RNA has occured.

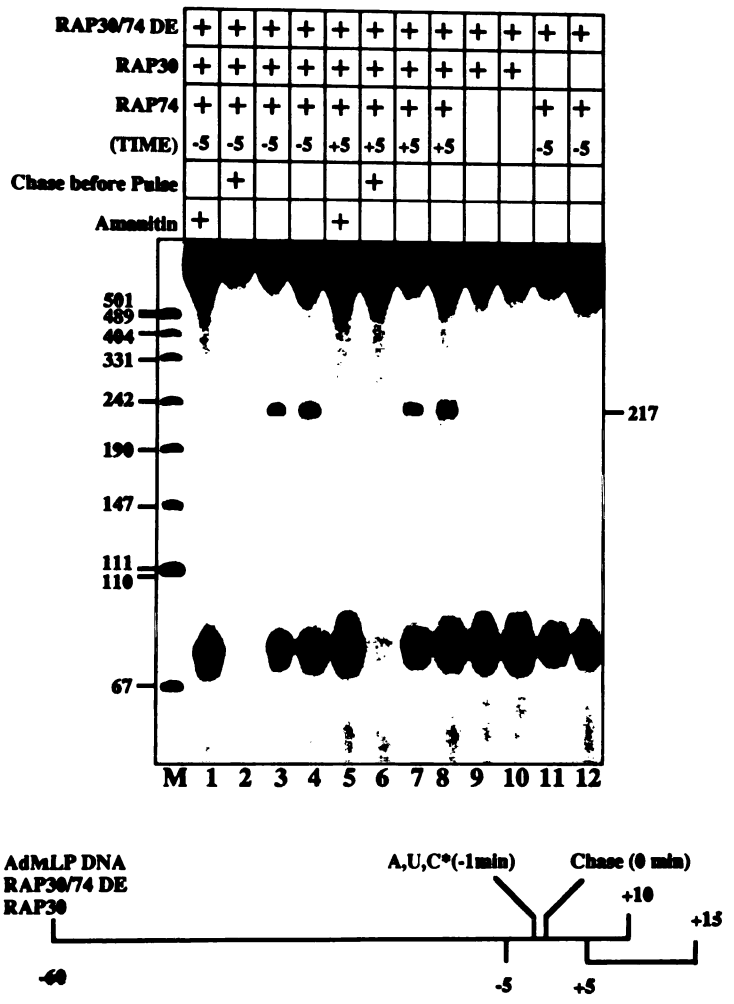
As expected, omission of RAP74 from the reaction abolished accurate transcription (lanes 9 and 10). Omission of RAP30 from the reaction abolished accurate transcription (lanes 11 and 12). Reversing the order of the pulse and chase (lanes 2 and 6) abolished visualization of the transcript, and the transcript was completely abolished by addition of 1 μg/ml α-amanitin.

Figu

adde

as in

Figure 6. Pulse-chase protocol: RAP74 is an elongation factor. RAP30 (100 ng) was added at t = -60 min, as indicated. RAP74 (200 ng) was added at t = -5 min or at +5 min, as indicated. All other reaction parameters were as described in Figure 5.


RAP30/74 DE
RAP30
RAP74
(TIME)
Chase before Pulse
Amanitin
501
489
404
331
242
190
147
111
110
67
217
M 1 2 3 4 5 6 7 8 9 10 11 12
AdMLP DNA
RAP30/74 DE
RAP30
A,U,C*(-1min)
Chase (0 min)
+10
+15
-60
-5
+5

Figur 6

DISC

inita

that

func

and

likel

Surp

elon

intia

pape

norm

bind

RAP

com

(Fig

the

poly

bonc

hydr

altho

less

com

phos

trans

DISCUSSION

We have separately analyzed the functions of RAP30 and RAP74 during the initation and early elongation phases of transcription by RNA polymerase II. We find that RAP30 is an essential initiation factor, but that RAP74 is dispensible for initiation functions and is only required for very early elongation of the transcript. Since RAP30 and RAP74 are tightly associated in a heteromeric complex (26-28), these factors most likely enter the pre-initiation complex together, bound to RNA polymerase II. Surprisingly, RAP30 and RAP74 have separate functions in accurate initiation and early elongation of RNA chains.

In Figure 7 we present a model for the function of RAP30 and RAP74 during intiation and elongation of transcription. This model is based on data presented in this paper and also the work of many others (reviewed in 5). The RAP30/74 complex normally enters the pre-initiation complex bound to RNA polymerase II. RAP30/74 binds to RNA polymerase II primarily through the RAP30 subunit (8, 11), and the RAP74 subunit is not required to recruit RNA polymerase II into the pre-initiation complex (6). Our observation that RAP30 is required for initiation of transcription (Figures 3A and 5) is consistent with evidence that RAP30 helps to bring polymerase to the promoter (6, 9). ATP hydrolysis is required for accurate initiation by RNA polymerase II (14, 25, 29, 30), and hydrolysis precedes formation of phosphodiester bonds (14). Sensitivity of transcription complexes to $KMnO_4$ indicates that ATP hydrolysis is required to separate template strands to form the open complex (13), although studies with o-phenanthroline copper to detect open complex formation seem less consistent with this conclusion (31). In any event, resistance of transcription complexes to sarkosyl requires both ATP hydrolysis and formation of the first few phosphodiester bonds (24, 25). In this paper we show that RAP74 is not required for transcription until after addition of sarkosyl (Figures 3B and 4), showing that RAP74 is

Figu
trans
polyr
is rec
requi
comp
requi
view

Figure 7. A model for RAP30/74 function in initiation and early elongation of transcription. RAP30/74 normally enters the pre-initiation complex bound to RNA polymerase II, but the RAP74 subunit is dispensible in this process. The RAP30 subunit is required to bring RNA polymerase II into the pre-initiation complex. RAP74 is not required for transcription until after: 1) ATP β–γ bond hydrolysis in initiation; 2) open complex formation; and 3) formation of the first few phosphodiester bonds. RAP74 is required for RNA polymerase II to escape from the promoter. This model represents our view based on the work of many investigators (see the text for details).

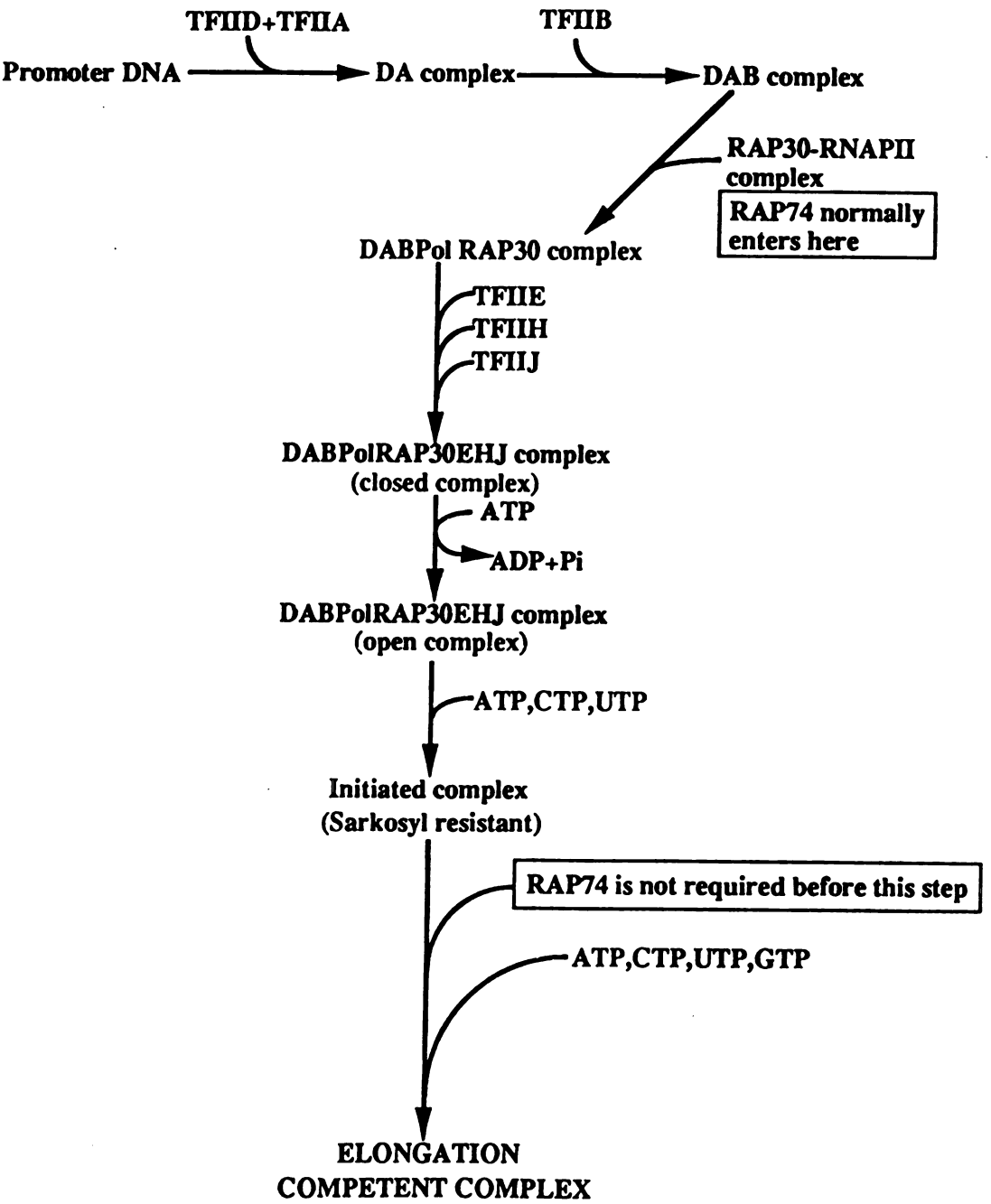

TFIID+TFIIA
TFIIB
Promoter DNA
DA complex
DAB complex
RAP30-RNAPII complex
RAP74 normally enters here
DABPol RAP30 complex
TFIIE
TFIIH
TFIIJ
DABPolRAP30EHJ complex (closed complex)
ATP
ADP+Pi
DABPolRAP30EHJ complex (open complex)
ATP,CTP,UTP
Initiated complex (Sarkosyl resistant)
RAP74 is not required before this step
ATP,CTP,UTP,GTP
ELONGATION COMPETENT COMPLEX

Figure 7

only

form

only

omis

posit

expe

phos

be a

nucl

pres

any

to t

sho

abs

ess

RA

to p

RA

the

of

no

hy

hig

tra

rep

or

only required after ATP hydrolysis, DNA strand separation and phosphodiester bond formation have occured. Using a pulse-chase protocol (Figure 6), we show that RAP74 is only required after formation of a short RNA chain (fewer than 11 nucleotides by omission of GTP) labeled by incorporation of ^{32}P-CMP (C can be incorporated into positions 2, 4, 6, 9 and 10 of the chain in the absence of GTP). The pulse-label experiment shows that RAP74 is dispensible for template strand separation and phosphodiester bond formation. After these steps in initiation have occured, RAP74 must be added to the reaction for RNA polymerase II to synthesize a short runoff (217 or 536 nucleotides) or paused (186 nucleotide) transcript. Inspection of transcripts formed in the presence of all 4 nucleoside triphosphates but in the absence of RAP74 does not reveal any short paused transcripts (see Figure 1). Attempts to observe short RNAs very close to the promoter on 20 % polyacrylamide gels have so far not been successful (data not shown). Apparently, RNA polymerase II is stalled very close to the promoter in the absence of RAP74. Our conclusion from these observations is that RAP74 has an essential function in promoter escape by RNA polymerase II. We think it likely that the RAP30 subunit of RAP30/74 participates in promoter escape, at least to help bind RAP74 to polymerase.

Although the RAP30/74-depleted extract is functionally depleted of RAP30 and RAP74 (Figures 1-6), we cannot be certain that we have removed all of the RAP74 from the extract. Residual RAP74 might be sufficient to support initiation but not elongation of transcription. RAP74 might, therefore, have an essential function in initiation that is not detected in these assays, and our conclusion that RAP74 is not required for ATP hydrolysis and intiation might be incorrect. If this alternate view were true, however, a higher concentration of RAP74 would be necessary for elongation than for initiation of transcription. Either RAP74 would have to cycle off of polymerase after initiation and be replaced by new RAP74 molecules with reduced polymerase affinity during elongation, or a larger number of RAP74 molecules would have to be required for elongation than for

initiat

suppo

this w

II.

σ^{70}

bind

poly

poly

RAI

bind

RA

DN

pol

pro

are

do

the

tra

po

tr

po

R

a

R

a

initiation. Such models appear to be too ornate for consideration without further supporting evidence. In any case, such models do not contradict the major conclusion of this work, that RAP74 has an essential function in promoter escape by RNA polymerase II.

The RAP30 subunit of RAP30/74 is homologous to the bacterial initiation factor σ^{70} (10, 11). The region of sequence similarity is found within the domain of σ^{70} that binds to bacterial RNA polymerase (11, 32). Human RAP30 binds to *E. coli* RNA polymerase and is displaced by binding of σ^{70}. σ^{70} can also bind to calf thymus RNA polymerase II (11). In addition to this conservation of RNA polymerase-binding sites in RAP30 and σ^{70}, these proteins also share some functional roles in transcription. σ^{70} binding to *E. coli* RNA polymerase suppresses non-specific DNA binding by polymerase. RAP30 (8) and RAP30/74 (9) have also been shown to suppress non-specific binding to DNA by mammalian polymerase. σ^{70} has the additional ability to dissociate RNA polymerase from non-specific sites on DNA to which it is bound. RAP30/74 has this property also (8, 9); but the RAP30 subunit by itself does not (8). Both RAP30 and σ^{70} are required factors for accurate initiation by their respective RNA polymerases. RAP30 does not appear to possess the sequence-specific promoter recognition properties of σ^{70}; these properties appear to be the function of other basal and promoter-specific transcription factors in the mammalian system.

Since the RAP30/74 complex may have an important role in release of RNA polymerase II from non-promoter DNA sites (8, 9), RAP30/74 may have a role in transcription termination. A role has also been defined for RAP30/74 in bringing RNA polymerase II to the promoter (6, 14). In this paper we show another function of RAP30/74 in promoter escape by RNA polymerase II. Other investigators have reported a role for RAP30/74 in later stages of transcript elongation (33, 34) and in elongation of RNA polymerase II through assembled nucleosomes (35). If RAP30/74 remains associated with RNA polymerase II throughout the transcription cycle, perhaps

RA

σ70

fact

RA

200

pol

elo

site

In a

of t

site

wh

RA

reg

pol

pol

co

elo

co

ini

fac

sti

ter

de

sa

RAP30/74 should be considered a polymerase subunit rather than an accessory factor. σ^{70} is considered to be an initiation factor for *E. coli* RNA polymerase because this factor is catalytic in the initiation process, cycling off polymerase during elongation. RAP30/74 is easily dissociable from RNA polymerase II at low salt concentrations (ie. 200 mM NaCl (36)), indicating that this factor may have a reversible interaction with polymerase at some point in the transcription cycle. Perhaps RAP30/74 and/or other elongation factors are dissociated from polymerase beyond splicing and polyadenylation sites, causing polymerase to lose processivity and ultimately to terminate transcription. In any case, RAP30/74 has been shown to have very specific functions at several stages of the transcription cycle.

Purified RNA polymerase II can synthesize long RNA chains from non-promoter sites when it initiates in the absence of accessory factors. In this paper, we show that when RNA polymerase II initiates transcription from a promoter in the absence of RAP74, the transcript is not elongated. Presumably factors present in cell extracts regulate polymerase escape from promoters. RAP74 may function to dissociate RNA polymerase II from strong DNA-protein and protein-protein interactions holding polymerse at the promoter. We suggest that RAP74 participates in early elongation to convert accurately initiated RNA polymerase II molecules from an initiated form to an elongation competent form. RAP74 might do this by stabilizing an elongation conformation of RNA polymerase II, and/or destabilizing interactions with other initiation factors. Factors that regulate promoter escape may include basal initiation factors and other proteins present in our HeLa cell extracts.

RNA polymerase II has recently been reported to have an exonuclease activity, stimulated by the elongation factor TFIIS (37-40), that might be expected to digest short ternary complexes, such as those formed in the absence of RAP74. Our data, however, demonstrates that such complexes can be stable for up to 38 min in the presence of sarkosyl (Figure 3B) and up to 5 min in the absence of sarkosyl (Figure 6). Sarkosyl

likely dissociates TFIIS from RNA polymerase II, and this may explain why transcription is paused at +186 in the presence of sarkosyl (24, 41).

In *Drosophila*, early elongation by RNA polymerase II appears to be regulated in part by a factor named P-TEF, which is involved in overcoming early termination of RNA chains (42). This previous report appears to be a distinct observation from that made in this paper, because P-TEF does not appear to be factor 5 (the RAP30/74 homolog in *Drosophila* (43)), and P-TEF functions further from the promoter (42) than the function defined for RAP74 in this paper. It may be that the human version of P-TEF or other factors cooperate with RAP74 in regulating promoter escape.

There is mounting *in vivo* evidence that transcription by RNA polymerase II can be regulated at very early elongation (reviewed in 44). In *Drosophila*, heat shock regulated genes have been shown to have a paused RNA polymerase very close to the promoter prior to gene induction by temperature upshift (45). Paused polymerase is also found just downstream of promoters for other *Drosophila* genes (45). The human c-myc gene also has a stalled RNA polymerase molecule proximal to its promoter (46). Based on the experiments in this paper, we suggest that stalled RNA polymerase II molecules are initiated in the absence of RAP74 (or another factor that regulates promoter escape). Promoter escape may be an important regulatory check point in control of gene expression and should be considered a potential target for regulation by activating and silencing transcription factors.

HIV-1 *tat* transactivator stimulates transcription through a *tar* RNA sequence, and regulation by *tat* influences both initiation and elongation of RNA chains (47-50). Regulation of elongation by *tat* may involve general elongation factors such as TFIIS and RAP30/74.

ACI

(G.

Exp

Wei

Jack

ACKNOWLEDGEMENTS

This research was supported by a grant from the National Institutes of Health (GM 40708). Z.B. is a member of the Michigan State University Agricultural Experiment Station. We thank Lee Kroos, Laurie Kaguni, Steven Triezenberg and Janek Werel for critical reading of this manuscript. We thank Judith Jaehning, Donal Luse, Jack Greenblatt and David Price for stimulating discussions and encouragement.

REFERENCES

1. Gross, C.A., Lonetto, M. and Losick, R. (1992) Transcriptional Regulation, S. L. McKnight and K.R. Yamamoto, eds., Cold Spring Harbor Laboratory Press, pp. 129-178.

2. von Hippel, P.H., Yager, T.D. and Gill, S.C. (1992) Transcriptional Regulation, S.L. McKnight and K.R. Yamamoto, eds., Cold Spring Harbor Laboratory Press, pp. 179-202.

3. Drombowski, A.J., Walter, W.A., Record, M.T. Jr., Siegele, D.A. and Gross, C.A. (1992) Cell 70, 501-512.

4. Gill, S.C., Weitzel, S.E. and von Hippel, P.H. (1991) J. Mol. Biol. 220, 307-324.

5. Zawel, L. and Reinberg, D. (1992) Prog. Nucleic Acids Res. Mol. Biol. 44, 67-108.

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

7. Conaway, J. W. and Conaway, R. C. (1990) Science 248, 1550-1552.

8. Killeen, M. and Greenblatt, J. (1992) Molec. Cell. Biol. 12, 30-37.

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

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

11. McCracken, S. and Greenblatt, J. (1991) Science 253, 900-902.

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

13. Wang, W., Carey, M. and Gralla, J.D. (1992) Science 255, 450-453.

14. Conaway, R. C. and Conaway, J.W. (1988). J. Biol. Chem. 263, 2962-2968.

15. Corden, J.L. (1990) Trends in Biol. Sci. 15, 383-387.

16. Laybourn, P.J. and Dahmus, M.E. (1989) J. Biol. Chem. 263, 18880-18885.

17. Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., and Aloni, Y. (1992) Cell 871-881.

18. Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979) Cell 18, 469-484.

19. Shapiro, D.J., Sharp, P.A., Wahli, W.W. and Keller, M.J. (1988) DNA 7, 47-55.

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

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

22. Wang, B.Q., Kostrub, D.F., Finkelstein, A. and Burton, Z.F. (1993) Protein Expression and Purification (in press).

23. Burton, Z.F., Ortolan, L.G. and Greenblatt, J. (1986) EMBO J. 5, 2923-2930.

24. Hawley, D.K. and Roeder, R.G. (1985) J. Biol. Chem. 260, 8163-8172.

25. Luse, D.S., Kochel, T., Kuempel, E.D., Coppola, J. A. and Cai, H. (1987) J. Biol. Chem. 262, 289-297.

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

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

28. Katajima, S., Tanaka, Y., Kawaguchi, T., Nagaoka, T., Weissman, S. and Yasukochi, Y. (1990) Nucleic Acids Res. 18, 4843-4849.

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

30. Sawadogo, M. and Roeder, R. G. (1984) J. Biol. Chem. 259, 5321-5326.

31. Buratowski, S., Sopta, M., Greenblatt, J. and Sharp, P.A. (1991) Proc. Natl. Acad. Sci. USA 88, 7509-7513.

32. Lesley, S.A. and Burgess, R.R. (1989) Biochemistry 28, 7728-7734.

33. Price, D.H., Sluder, A.E. and Greenleaf, A.L. (1989) Mol. Cell. Biol. 9, 1465-1475.

34. Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991) Mol. Cell. Biol. 11, 1195-1206.

35. Izban, M.G. and Luse, D. S. (1992) J. Biol. Chem. 267, 13647-13655.

36. Sopta, M., Carthew, R.W. and Greenblatt, J. (1985) J. Biol. Chem. 260, 10353-10361.

37. Izban, M.G. and Luse, D.S. (1992) Genes and Develop. 6, 1342-1356.

38. Wang, D. and Hawley, D.K. (1993) Proc. Natl. Acad. Sci. USA 90, 843-847.

39. Reines, D., Ghanouni, P., Li, Q.-q. and Mote, J. Jr. (1992) J. Biol. Chem. 267, 15516-15522.

40. Kassavetis, G.A. and Geiduschek, E.P. (1993) Science 259, 944-945.

41. Wiest, D.K., Wang, D. and Hawley, D.K. (1992) J. Biol. Chem. 267, 7733-7744.

42. Marshall, N.F. and Price, D.H. (1992) Molec. and Cell. Biol. 12, 2078-2090.

43. Kephart, D. D., Price, M. P., Burton, Z.F., Finkelstein, A., Greenblatt, J. and Price, D. H. (1992) Nucleic Acids Res. (in press).

44. Spencer, C.A. and Groudine, M. (1990) Oncogene 5, 777-785.

44. Rougvie, A. and Lis, J.T. (1990) Mol. and Cellular Biol. 10, 6041-6045.

46. Krumm, A., Meulia, T., Brunvand, M. and Groudine (1992) Genes and Develop. 6, 2201-2213.

47. Marciniak, R.A., Calnan, B.J., Frankel, A.D. and Sharp, P.A. (1990) Cell 63, 791-802.

48. Marciniak, R.A. and Sharp, P.A. (1991) EMBO J. 10, 4189-4196.

49. Kato, H., Sumimoto, H., Pognonec, P., Chen, C.-H., Rosen, C.A. and Roeder, R.G. (1992) Genes and Develop. 6, 655-666.

50. Cullen, B.R. (1990) Cell 63, 655-657.

CHAPTER III

RAP74 IS REQUIRED FOR STABILITY OF NEWLY-INITIATED TRANSCRIPTION COMPLEXES

ABST

one of

transcr

enable

polyme

paused

radioac

transcr

to cha

transcri

the sta

positio

comple

transcri

enough

require

positiv

transcri

ABSTRACT

The identification of structural transitions accompanying stabilization has been one of the major concerns in the biochemical analysis of transcription. *In vitro* transcription using immobilized templates and high percentage polyacrylamide gels enable us to observe nascent RNAs 11 to 20 nucleotides in length due to stalling of RNA polymerase II on the adenovirus major late promoter shortly after initiation. These paused transcription complexes are mainly due to the limiting concentration of radioactive nucleoside triphosphate; however, some intrinsic property of the basal transcription machinery may also contribute to this stalling. This assay system enables us to characterize the mechanism of transcription, namely the specificity of nascent transcripts, the location of the catalytic site in the transcription complex, and formation of the stable ternary complex. Polypeptide sequence between positions 409-517 and positions 172-205 of RAP74 may play a critical role in stability of the transcription complex. The most interesting observation is of a kinetic lag for the formation of stable transcription complexes. The lag may be the effect of the short RNAs which are not long enough to fill in the first transcript binding site of RNA polymerase II, or it may be required for the conformational change of this enzyme to allow different interactions with positive and negative regulatory factors. This could be a general phenomenon of transcription and an important regulatory point in gene expression.

INTRO

genes

stage i

Refs.

and pr

shown

Prom

facto

befo

conf

relea

facto

whi

stra

tran

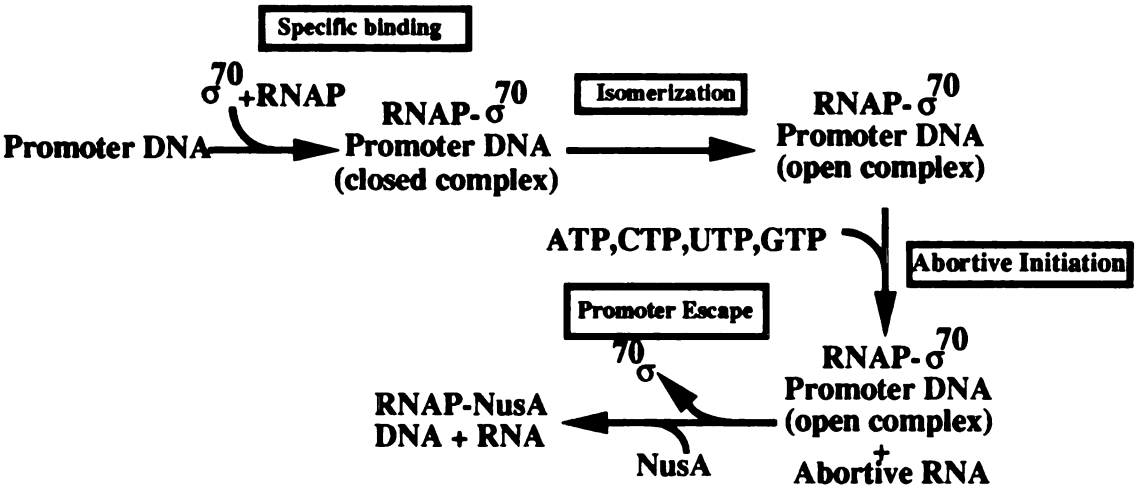
elon

enzy

typi

INTRODUCTION

Both in prokaryotes and in eukaryotes, RNA polymerases (RNAP) transcribe genes accurately. Transcription includes initiation, elongation and terminationp; each stage in the process is a level at which gene expression can be regulated (reviewed in Refs. 1-4). For the initiation of transcription, it includes specific binding, isomerization and promoter escape. Schematically, the process of *E. coli* RNA polymerase initiation is shown as.



E.coli RNA polymerase initiates transcription in the presence of a sigma initiation factor. The primary sigma factor in *E. coli* is σ70. σ70 first binds to RNA polymerase before it associates with the promoter, and binding to RNA polymerase alters the conformation of σ70 to expose a DNA-binding domain (5). Also, the binding of σ70 releases RNA polymerase from non-specific sites on DNA. In the presence of its σ-factor, RNA polymerase binds tightly to promoter DNA, first to form a closed complex, which later becomes an open complex through the separation of the template DNA strands. At many promoters, RNA polymerase then synthesizes numerous short transcripts, called abortive RNAs, before the RNA polymerase can become a productive elongation enzyme. Abortively initiated RNA is released from polymerase without enzyme dissociation from the promoter (6). Commitment to productive elongation typically occurs after the synthesis of a 9-12-nucleotide transcript and is associated with

the release

change in

RNA poly

occur. It

HeLa cell

initiation

In

binding c

that occu

coli; RN

the modi

polymer

pause an

T

promote

RAP30/

TFIIH)

substra

DNA t

shown

promo

associa

compl

is nec

RAP7

the release of sigma factor (7-11). Binding of RNA to the RNA binding sites (12) and a change in conformation of RNA polymerase is required to cause σ70 release, allowing RNA polymerase to escape (12). In eukaryotes, an analogous transition is believed to occur. It has been shown that RNA polymerase II can abortively initiate transcription in HeLa cell nuclear extracts (13). The abortive initiation process represents repeated re-initiation by RNA polymerase without escape from the promoter.

In *E.coli*, the elongation conformation of RNA polymerase is stabilized by the binding of elongation factors such as NusA protein (14). There are also regulatory events that occur during elongation. One example in prokaryotes is that of phage λ infected *E. coli;* RNA polymerase pauses at position +16 on the pR' promoter of phage λ allowing the modification of polymerase by the phage-encoded Q protein (15). The Q-modified polymerase can then leave the +16 pause site in a form that can read through subsequent pause and termination sites.

The essential components of messenger RNA transcription in eukaryotes are: promoter DNA, general transcription factors (TFIID, TFIIA, TFIIB, TFIIF also called RAP30/74 [RAP is an acronym for "RNA polymerase II-associating protein"], TFIIE and TFIIH), RNA polymerase II, ATP as energy source, and nucleoside triphosphates as substrates (reviewed in Ref.1 and 2). The assembly of the pre-initiation complex on the DNA template has been suggested to be a multistep sequential process, schematically shown on the next page.

In general, TFIIA stimulates the TBP binding event on TATA-containing promoters, then TFIIB binds to form a DAB complex (16, 17). RNA polymerase II associated with RAP30/74 (TFIIF) specifically binds the promoter sequence in the DAB complex. Although both RAP30 and RAP74 associate with polymerase *in vivo*, RAP30 is necessary and sufficient for guiding polymerase II to the DAB complex (18, 19). RAP74 functions later during the transcription process (20, Chapter II).

Promoter

Initiat
(Sarkos

Abo

Elongat

(Stable

A

transcrip

complex

the form

terminal

reported

be invol

elongat

detachr

a helica

factor(s

or com

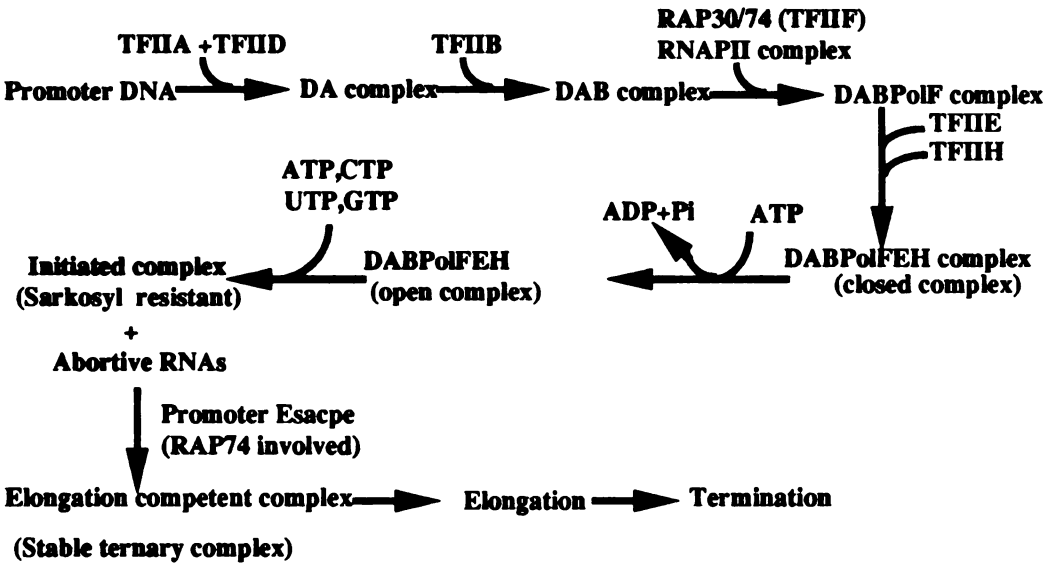
transcr

exist a

open c

aborti

stable



After RNA polymerase II joins the complex, TFIIE and TFIIH enter to form the transcription closed complex (DABpolFEH complex) which is then converted to the open complex which is characterized by the hydrolysis of ATP, the separation of DNA strands, the formation of phosphodiester bonds, and the phosphorylation of the CTD (carboxyl terminal domain) on the largest subunit of polymerase II (21). Recently, TFIIH has been reported to contain a CTD kinase activity and two helicase activities (22-26), which could be involved in this event. Promoter escape occurs after initiation and before productive elongation. It may require a conformational change of RNA polymerase II, the detachment of RNA polymerase II from some general initiation factors, and activation of a helicase to further unwind the DNA template. After this transition, it is not clear which factor(s) move along with RNA polymerase II and which remain bound to the promoter or completely dissociate from the DNA. Since commitment of RNA polymerase to the transcription of a specific gene is a multiple step process, regulatory checkpoints could exist at each stage, notably promoter recognition, assembly of a preinitiation complex, open complex formation, synthesis of the first phosphodiester bond, or transition from abortive to productive elongation. One of the important events is the formation of a stable ternary complex of template DNA, RNA polymerase and nascent RNA. Recent

reports in the literature indicate that such a stable ternary complex may be a checkpoint for transcriptional control in vivo (reviewed in 27, 28). Some RNA polymerase II molecules have been identified as being actively engaged in transcription but also being defective for elongation. Before heat shock on the *hsp70* promoter of *Drosophila melanogaster*, an RNA polymerase II molecule is found to be initiated but stalled within the first 29 nucleotides of the mRNA chain. After heat shock, polymerase molecules are found distributed throughout the *hsp70* gene, and the hsp70 protein is produced (29). The human *c-myc* gene also appears to be regulated at the level of transcription elongation. An attenuation site was identified for this gene at the exon I-intron I boundary, but it may be that this site is secondary to a strong pause site located very close to the promoter (30). A stalled RNA polymerase II molecule has been identified close to position +30 of the human *c-myc* promoter by $KMnO_4$ modification of thymidines and nuclear run-on assays. Paused polymerase molecules at the exon I-intron I boundary may result from elongation-defective polymerase molecules that inefficiently traverse the pause site close to the promoter.

The process by which an RNA polymerase leaves a promoter to become an efficiently elongating enzyme is poorly understood. In order to better understand promoter escape, and to characterize the transition from initiation to processive elongation, methods have been adapted (13, 31, 32) and modified to analyze accurate RNA synthesis corresponding to the transcription initiation site on adenovirus major late promoter (AdMLP) in our laboratory. Biotinylated templates are immobilized on streptavidin-agarose (31, 32), which allows the purification of ternary complexes from other unbound proteins and nucleotides. In this report, experiments were designed to study the RNAP II transcription process at the early stage of nascent RNA synthesis immediately after initiation.

MATERIALS AND METHODS

Construction of the Immobilized Templates

The method for preparation of the immobilized template was adapted from proccedures published by other investigators (31, 32). Template DNA containing the Ad-MLP was synthesized by standard PCR techniques (Perkin-Elmer/Cetus). In PCR reactions, pBluescriptKS$^+$/MLP was the template which includes the sequence of adenovirus major late (AdML) promoter. Primers were complementary to the region around -260 and +220, relative to the start point of the transcription. The sequence of the upstream 5'-biotinylated primer was 5'-biotin-CCCTCGAGCGGTGTTCCGCGGTCCT CCTCG-3', and the sequence of the downstream primer was 5'-CGGTGGCGGCCGCT CTAGAACTAGTGGATC-3'. (Primers were synthesized in the Macromolecular Structure and Synthesis Facility, Michigan State University). The PCR product was digested by SmaI endonuclease to generate the 3'-blunt end and passed through a Sephadex G-50 column to remove the unincorporated biotinylated primer. The purified PCR DNA was incubated with Streptavidin Agarose (GIBCO BRL) in STE buffer (10 mM Tris [pH 7.6], 1 mM EDTA, 100 mM NaCl) with a rotary mixer at room temperature overnight. Immobilized templates were washed with STE buffer several times and then resuspended into H_2O.

Transcription Assays

An extract derived from HeLa cell nuclei was prepared as described in Shapiro et al. (33). HeLa nuclear extract was combined with immobilized DNA and preincubated for 60 min in microfuge tubes. The pre-incubation reaction volume was 15.2 µl in addition to the volume of the beads. Reactions were performed at room temperature in 12 mM Hepes pH 7.9, 12 % glycerol, 60 mM KCl, 8 mM $MgCl_2$, 3.12 mM EGTA, 0.12 mM EDTA and 1.2 mM DTT. 10 µU Hexokinase and 50 µM D-glucose were added to each reaction to deplete ATP and confine transcription initiation to times when a pulse of nucleoside triphosphates (NTPs) was added. For initiation, 100 µM ATP, GTP, UTP and

<1 μM α-[^{32}P] CTP (10 μCi/reaction) or 100 μM ATP, UTP and <1 μM α-[^{32}P] CTP (10 μCi/reaction) were added and incubated for one minute. Afetr initiation, the tubes were spun and washed twice by adding 50 μl 12 mM Hepes pH 7.9, 12 % glycerol, 60 mM KCl, 3.12 mM EGTA, 0.12 mM EDTA, 1.2 mM DTT, and 1mg/ml bovine serum albumin. The reaction was either stopped or chased with NTPs. For analysis of short transcripts, ternary complexes were purified and resuspended in loading buffer (90% formamide, 1% SDS, 10 mM Tris pH7.9, 1 mM EDTA, and 0.1% bromophenol blue and 0.01% xylene cyanol) and visualized by autoradiography of 23% (20% polyacrylamide:3% methylene-bis-acrylamide) or composite 8%/23% (top portion of the gel is 8% and bottom 2/3 is 23%) polyacrylamide gels. For runoff RNA analysis, reactions were phenol-chloroform extracted and ethanol precipitated. Then, transcripts were resolved in 6% (30% polyacrylamide:0.8% methylene-bis-acrylamide) or composite 8%/23% polyacrylamide gels and visualized by autoradiography. Transcripts were quantitated using a Molecular Dynamics Phosphorimager. Details of each experiment are described in Results and in figure legends.

RESULTS

The stalling of RNA polymerase II can be observed with different radioactive labeling, and the precise pause is dependent on which NTP is limiting.

An extract derived from HeLa cell nuclei was incubated with the AdML promoter immobilized on agarose beads (Figure 1). After formation of preinitiation complexes, NTPs were added to initiate transcription. One minute after initiation, ternary complexes were recovered by centrifugation and washing of the beads. Short transcripts were observed in a one minute pulse before the active elongation on the AdML promoter by RNA polymerase II, whether labeling is with radioactive CTP (Figure 2, lanes 2 & 4), GTP (lane 5) or UTP (lane 6). These short transcripts are sensitive to α-amanitin (lanes 1), and paused complexes were elonagted upon addition of chase NTPs (lane 3).

Figure 1. Immobilized template: The schematic drawing of biotinylated PCR template linking to agarose bead through a biotin-streptavidin linkage.

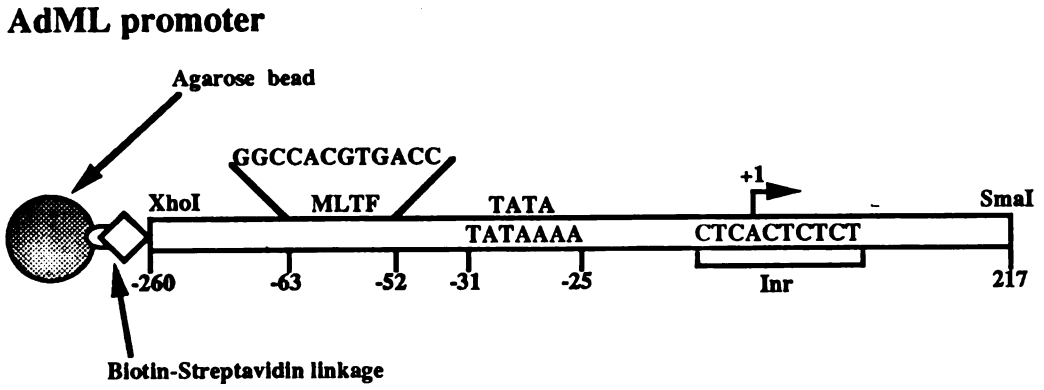

AdML promoter
Agarose bead
GGCCACGTGACC
XhoI
MLTF
TATA
+1
SmaI
TATAAAA
CTCACTCTCT
-260
-63
-52
-31
-25
Inr
217
Biotin-Streptavidin linkage

Figure 1

Figure 2

from Ac

extract o

hr. Pulse

G*) at a

CTP). T

U*. Af

The rea

sequenc

(α-A) v

Marker

ladder

determ

Figure 2. Synthesis of short transcripts by different labeling of the short transcripts from AdML promoter. The reaction scheme is shown at the top of the figure. An extract of HeLa cell nuclei (NE) is mixed with immobilized template and incubate for 1 hr. Pulse NTPs include three unlabeled NTPs at 100 μM and an α[^{32}P]-NTP (C*, U*, or G*) at a concentration < 1 μM. Reactions in lanes 1-4 are labeled with C* (α-[^{32}P] CTP). The reaction in lane 5 is labeled with G*, and the reaction in lane 6 is labeled with U*. After a one minute pulse, complexes are isolated by centrifugation and washing. The reaction shown in lane 3 is chased for 10 min with 1 mM each unlabeled NTP. The sequence of AdML transcripts is shown at the bottom of figure. One μg of α-amanitin (α-A) was added at t = -60 min or during the centrifugation and washing as indicated. Marker lanes are 5' labeled DNA oligonucleotide A) 18 mer; B) 16 mer and C) oligo dT ladder (Betheseda Research Laboratories). The precise size of short RNAs are determined in the experiment shown in figure 4.

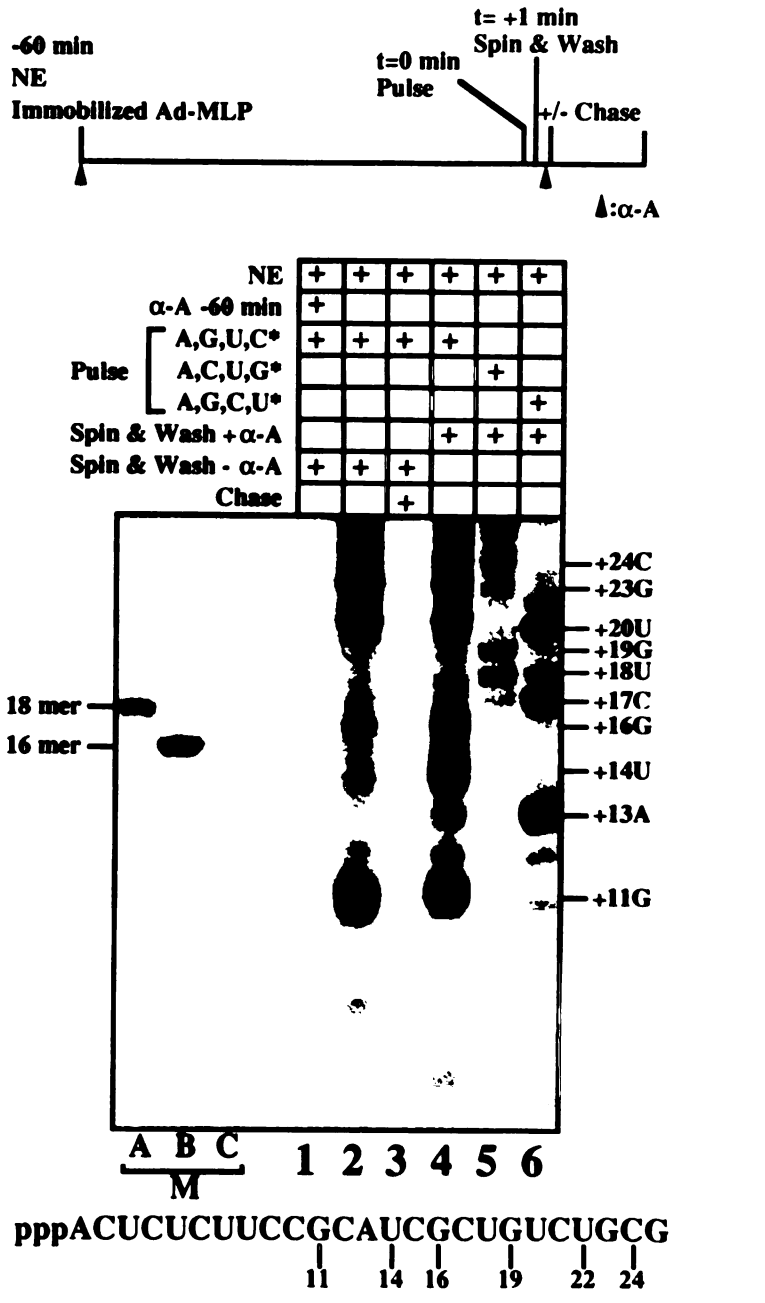

-60 min
NE
Immobilized Ad-MLP
t=0 min
Pulse
t= +1 min
Spin & Wash
+/- Chase
:α-A
NE
α-A -60 min
Pulse
A,G,U,C*
A,C,U,G*
A,G,C,U*
Spin & Wash +α-A
Spin & Wash - α-A
Chase
18 mer
16 mer
+24C
+23G
+20U
+19G
+18U
+17C
+16G
+14U
+13A
+11G
A B C
M
1 2 3 4 5 6
pppACUCUCUUCCGCAUCGCUGUCUGCG
11 14 16 19 22 24

Figure 2

When α-[^{32}P] CTP is the labeled nucleoside triphosphate, major paused sites are observed at +11G, +14U and +16G, just before the addition of +12, +15 and +17C. Of all these short transcripts, +11G is the most significant one since it contains the most radioactivity and the fewest labeled C's. Polymerase appaers to be stalled at very similar positions whether complexes were recovered in the absence or presence of α-amanitin, indicating that little elongation or exonucleolytic degradation of transcripts occurs during the purification process (compare lane 4 to lane 2).

When α-[^{32}P] GTP is the radioactive nucleoside triphosphate, the major paused sites are +18U and +19G (lane 5). Transcription rationally pauses at +18U prior to addtion of +19G undr the limitation of GTP, whereas pauses at +19G is unexpected. When α-[^{32}P] UTP is the radioactive nucleoside triphosphate, the major paused sites are +13A, +17C and +20U (lane 6). The unexpected pause at +20U cannot be simply explained by limitation of UTP.

Pausing primarily occured at positions that precede addition of a limiting nucleoside triphosphate, but under less limiting concentration of NTP (Figure 3) or under severe limitation of GTP (lane 1 of Figure 5) we still observed short transcripts. This implies that the paussing could be an intrinsic property of the transcription complex. Transcripts shorter than +11G are barely detectable after washing the complexes, indicating that such transcripts are not stably associated in the ternary complexes and are abortively initiated. Similar conclusions about the stability of short ternary complexes have been made by Luse and colleagues (34, 35).

Stalling of RNA polymerase II is also observed at higher concentration of radioactive NTP. The primary reason for the stalling of RNA polymerase II is that there is a limiting concentration of CTP (<1 μM) for the radioactive activity in the reaction. In Figure 3, we perform an experiment with the α-[^{32}P] CTP and 10 μM CTP. Reducing the limitation of CTP concentration, we still observe the stalling of RNA polymerase indicated by the accumulation of short RNAs during a one minute pulse. This result

Figure 3.

The pulse

[^{32}P] CTI

Figure 3. Paused transcripts are also observed in less limitation of radioactive NTP. The pulse nucleoside triphosphates are 100 μM A, G, U and 10 μM C with 10 μCi α-[^{32}P] CTP.

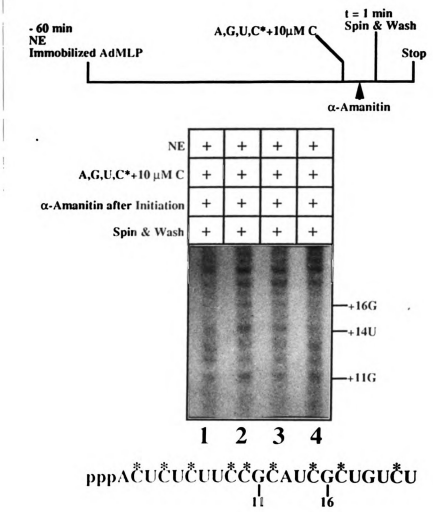

- 60 min
NE
Immobilized AdMLP
A,G,U,C*+10μM C
t = 1 min
Spin & Wash
Stop
α-Amanitin
NE
A,G,U,C*+10 μM C
α-Amanitin after Initiation
Spin & Wash
+16G
+14U
+11G
1 2 3 4
pppA*C*U*C*U*C*U*U*C*C*G*C*A*U*C*G*C*U*G*U*C*U
11
16

Figure 3

sugge

RNA

used a

mRN

3'-dido

depen

demon

synthe

of RN

Since

differe

+1 ATI

transcri

incorpo

move f

position

inhibiti

observe

recover

re-exan

mobilit

transcri

suggests that the limitation of the labeling CTP is not the only reason for the stalling of RNA polymerase II. This result is also true when α-[^{32}P] UTP and 10 μM UTP were used as the labeling condition (data not shown).

+11G synthesis is dependent upon ATP hydrolysis. One criterion of eukaryotic mRNA initiation is the requirement for ATP hydrolysis. In Figure 4, we used ddATP (2'-3'-dideoxy ATP) and AMPPNP (adenosine 5'-β,γ-imino triphosphate) to study the dependence on ATP hydrolysis of +11G synthesis. The results of this experiment demonstrate that a hydrolyzable β–γ bond of ATP or an ATP analog is essential for the synthesis of +11G from the AdML promoter (lanes 1, 2 & 3). As expected, the mobility of RNA with a 5'-end AMPPNP is different from that of a 5'-end ATP (lanes 1 & 3). Since AMPPNP and ATP both incorporate into the body of RNA chain as AMP, this difference is most consistent with initiation of the RNAs with +1 AMPPNP (lane 1) and +1 ATP (lane 3), as expected for the AdML promoter.

Surprisingly, with the presence of α-amanitin we observed the elongation of the transcripts (lane 4). This result suggest that the catalytic sites of RNA polymerase II can incorporate several nucleotides on the 3'end of RNA in the presence of α-amanitin and move from +11, +14, +16 to positions at +17 to +19 as well as from +20, +23, +29 to positions +32 to +34, however the front edge of the emzyes are locked by α-amanitin as inhibition of transcription.

This 23% gel was run for a longer time to obtain better resolution, and we observed double bands for each specific signal. These +11G-RNA and +14U-RNA were recovered from the gel slices and digested with calf intestinal alkaline phosphatase then re-examined on a 23% polyacrylamide gel where they become a single band with a higher mobility (data not shown). This suggests that there may be a phosphatase activity in the transcription complex.

Figure 4. Short transcripts synthesis are dependent on ATP hydrolysis. ATP, ddATP (ddA) and AMPPNP (AN) are used to study the requirement for hydrolyzable β–γ bond. AMPPNP can be incorporated in to the first position of the RNA chain in the presence of ddA (lane 1), but not in the absence (lane 2). Some transcription elongation occurs in the presence of α-amanitin when 1 mM each chase NTP was added for 1 min (lane 4; indicated by open arrows). AMPPNP in the first position causes the mobility of short RNAs to be faster and inhibits dephosphorylation of the 5' triphosphate (compare lanes 1 and 3).

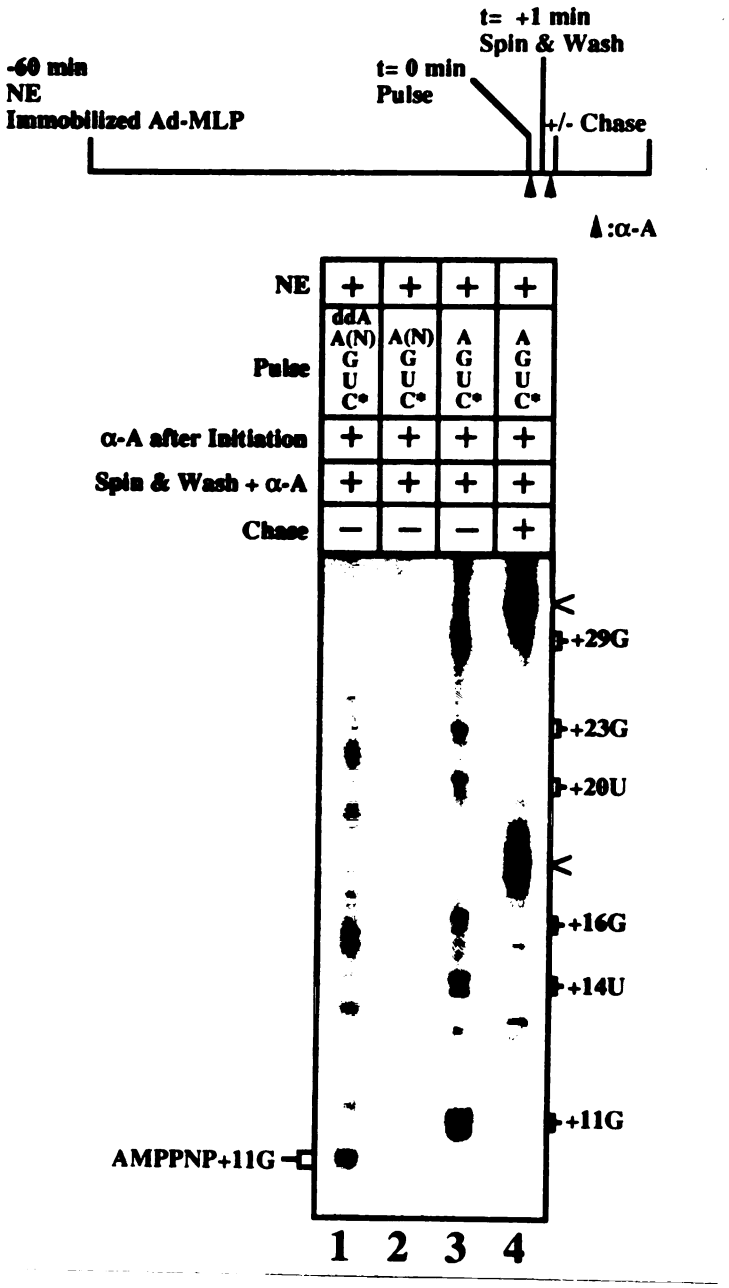
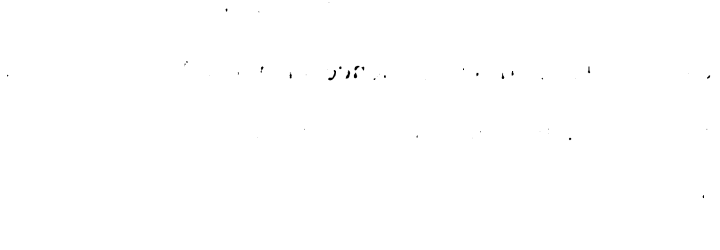

t= +1 min
Spin & Wash
-60 min
NE
Immobilized Ad-MLP
t= 0 min
Pulse
+/- Chase
▲:α-A
NE
Pulse
α-A after Initiation
Spin & Wash + α-A
Chase
ddA A(N) G U C*
A(N) G U C*
A G U C*
A G U C*
+29G
+23G
+20U
+16G
+14U
+11G
AMPPNP+11G
1 2 3 4

Figure 4



Catalytic sites of RNA polymerase II location and precise sizes of short RNAs are determined by conditional chase. Since the migration of RNA is different from that of the DNA size marker, the precise sizes of short transcripts are difficult to determine. Also these short RNAs are characterized as the products of RNA polymerase II using the properties of α-amanitin inhibition (Figure 2, lane 1) and the requirement of ATP hydrolysis (Figure 4, lane 1). A direct evidence such as RNA sequence will confirm the nature of these stalled products. Therefore, we conducted a conditional chase to verify these short transcripts. In the conditional chase experiment (Figure 5) we first initiate transcription to allow the synthesis of +11G transcripts, then we spin and wash the ternary complexes which are assembled on the immobilized templates. Under the size assignment by which the prominent short RNA is +11G, when we supply the appropriate NTPs with the condition for elongation of transcription we should be able to observe the expected products according to DNA sequence if the +11G-assignment is correct. In lane 1, initiation is under severe limitation of GTP and CTP with no additional chase, we observed +11G as the major transcript as well as +14U and +16G identically to those in figures 2, 3, and 4. This +11G-ternary complex is chased to +12C-ternary complex when CTP was present during chase (lane 2). When CTP and ATP are added, the +11G is elongated to +13A (lane 3). When CTP, ATP and UTP are added, the transcript is elongated to +15C (lane 4). When the elongation NTPs are CTP, ATP, UTP and 3'-O-methyl GTP (mG), we observed a +16mG with faster mobility (lane5). This +11G-ternary complex can only be elongated when the CTP is supplied in the reaction (lanes 6-10, compare to lanes 1-5 & 11-13). Where there are some newly synthesized +11mG with higher mobility (lanes 11, 12, 13 & 5) and some transcripts shorter than +11G (lanes 2-5 and lanes 11-13) suggests that these transcripts are generated by S-II mediated shortening of RNA followed by elongation to the indicated positions. This experiment directly confirms that these short transcripts are synthesized by RNA polymerase II from adenovirus major late promoter and these spun-and-washed ternary complexes are

Figur
initiat
comp
of 1
transc
witho
transc
the lo

Figure 5. Conditional chase from +11G of the AdML transcript. Transcription is initiated by a 1 min pulse with ATP, CTP, and $\alpha[^{32}P]$-CTP. After washing of complexes, elongation is allowed to proceed with a 1 min chase of various combinations of 1 mM each CTP, ATP, UTP, and 3'-O-methyl GTP (mG), as indicated. The transcription extract contained enough contaminating GTP to elongate to +11G and +16G without addition of GTP to the reaction (lane 1). These data confirm the size of short transcripts and that short RNAs are initiated from the AdML promoter, also demonstrate the location of the catalytic site of RNA polymerase II in the ternary complexes.

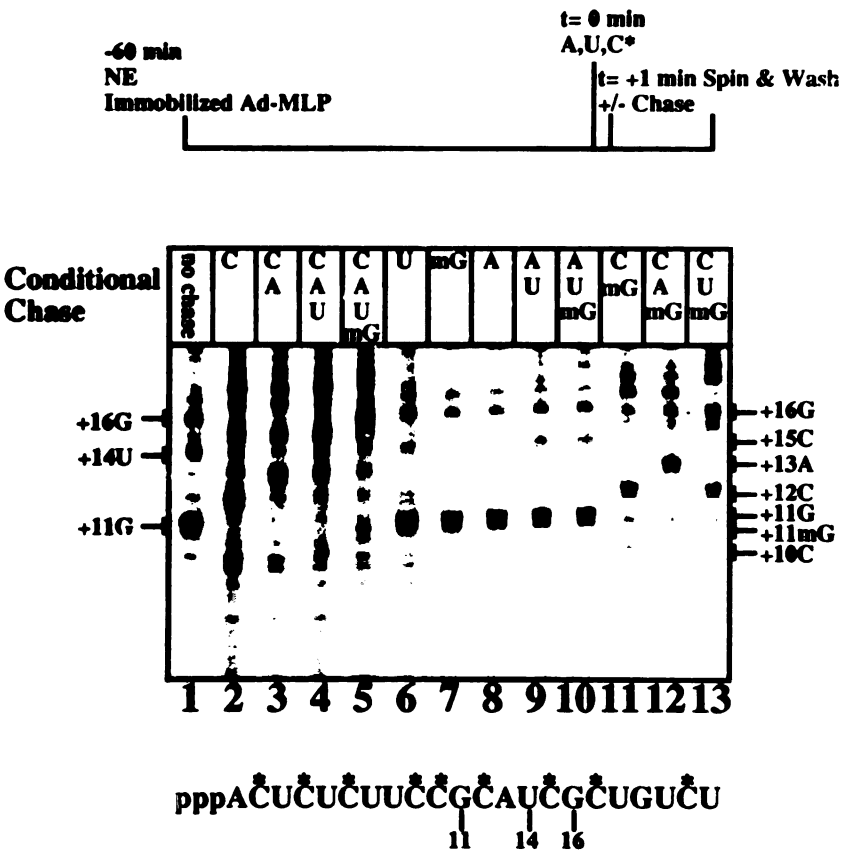

t= 0 min
A,U,C*
-60 min
NE
Immobilized Ad-MLP
t= +1 min Spin & Wash
+/- Chase
Conditional Chase
no chase
C
CA
CAU
CAUmG
U
mG
A
AU
AUmG
CmG
CAmG
CUmG
+16G
+14U
+11G
+16G
+15C
+13A
+12C
+11G
+11mG
+10C
1 2 3 4 5 6 7 8 9 10 11 12 13
pppACUCUCUUCCGCAUCGCUGUCU
11 14 16

Figure 5

transcripti

downstrea

Ru

experimen

that these

autoradiog

short and r

pulse NTP

condition b

amanitin is

are triplica

Lanes 6-8

noted. The

Pre

pausing pro

in transcrip

varying the

washed 0 t

differences

that all of

complexes.

A 1

complex fr

formation i

washing of

extract is i

preinitiation

transcriptionally active with their catalytic sites locating at +11, +14, and +16 positions downstream from the transcription start site.

Runoff RNA can be synthesized from the paused ternary complexes. An experiment with a composite gel (top 1/3rd 8%, bottom 2/3rds 23%) provides evidence that these short transcripts are the precursors of runoff products. Figure 6 is an autoradiograph of 8%/23% composite polyacrylamide gel that allows examination of short and runoff RNAs on the same gel. Lanes 1-4 are those reactions with addition of pulse NTPs only, while lanes 5-8 were prepared identically to lanes 1-4 for the pulse condition but are elongated with 1 mM each NTP during a 10 min chase protocol. If α-amanitin is added neither short (lane 1) nor runoff (lane 5) RNAs are observed. Lane 2-4 are triplicates of the pulse reaction only. +11G, +14 U and +16G transcripts are noted. Lanes 6-8 are triplicates of elongation reaction with chase NTPs, +217 transcripts are noted. The +251 transcripts are due to incomplete SmaI digestion of PCR template.

Preinitiation complexes can be isolated by immobilized template. Since pausing proximal to the AdML promoter might be a process regulated by factors present in transcription complexes, the stability of the preinitiation complex was tested by varying the number of washing treatments prior to initiation (Figure 7). Complexes were washed 0 to 4 times before addition of pulse NTPs without any clearly discernible differences in the level or distribution of short RNAs (lanes 2 to 5). The results suggest that all of the factors required for pausing are stably associated with transcription complexes. Pausing at these positions could be an intrinsic function of polymerase itself.

A 15 to 20 second kinetic lag is observed in formation of stable ternary complex from the AdML promoter. The kinetics of stable ternary complexes formation is measured by addition of α-amanitn at various times after initiation and washing of ternary complexes from adenovirus major late promoter. HeLa nuclear extract is incubated with immobilized AdML template for one hour to allow the preinitiation complexes to form. At t = 0 min, A, G, U and C* are added to initiate the

Figu

a 8%

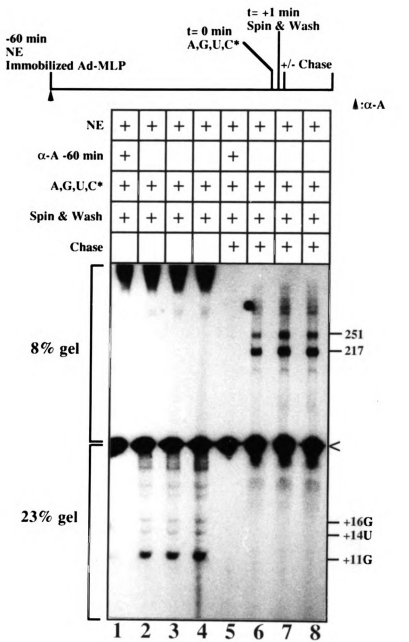
transc

positi

SmaI

comp

Figure 6. Short transcripts are the precursors of runoff RNA. An autoradiograph of a 8%/23% composite gel allows the comparison between short RNAs and runoff transcripts. Short transcripts (lanes 2-4) can be chased quantitatively to the +217 runoff position (lanes 6-8). Some PCR-synthesized templates are not efficiently digested by SmaI generating a +251 runoff position. Isotope that accumulated at the interface of the composite 8% and 23% gel (open arrow) does not represent a specific AdML transcript.


t= +1 min
Spin & Wash
t= 0 min
A,G,U,C*
-60 min
NE
Immobilized Ad-MLP
+/- Chase
▲:α-A
NE
α-A -60 min
A,G,U,C*
Spin & Wash
Chase
8% gel
23% gel
251
217
+16G
+14U
+11G
1 2 3 4 5 6 7 8

Figur 6

Fig

pol

bef

Figure 7. Washes of the preinitiation complexes do not alter the stalling of RNA polymerase II. Different number of washes are applied to the preinitiation complexes before initiation in lanes 3, 4, and 5 compare to no wash treatment in lane 2.

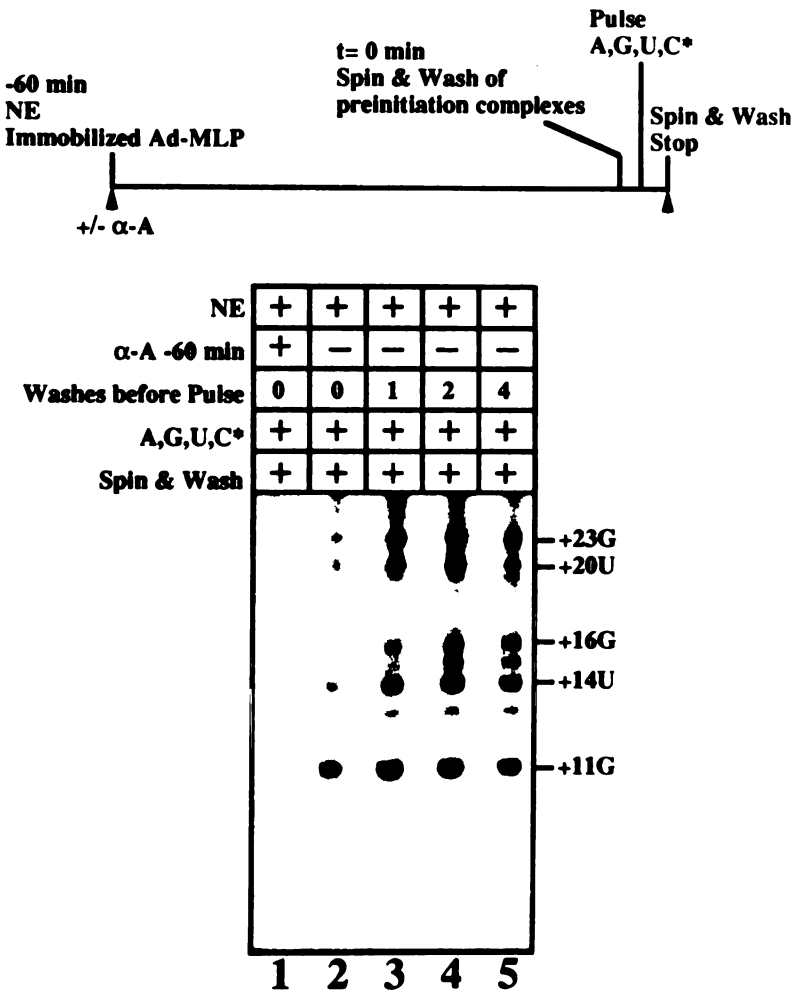

Pulse
A,G,U,C*
t= 0 min
Spin & Wash of
preinitiation complexes
-60 min
NE
Immobilized Ad-MLP
Spin & Wash
Stop
+/- α-A
NE + + + + +
α-A -60 min + − − − −
Washes before Pulse 0 0 1 2 4
A,G,U,C* + + + + +
Spin & Wash + + + + +
+23G
+20U
+16G
+14U
+11G
1 2 3 4 5

Figure 7

synthe

the re

transc

were

plotte

+11G

the se

hydro

preini

pulse

ternar

unaffe

which

from

sugge

bond

betwe

show

migh

not r

RAP

posit

(37),

affec

RAP

synthesis of short RNAs. At t = 0, 5, 10, 20, 40 and 60 seconds, α-amanitin was added to the reactions. After spin and wash, stable ternary complexes were extracted and short transcripts were analyzed on a 23% polyacrylamide gel. The prominent +11G-transcripts were quantitated using a Molecular Dynamics Phosphorimager and quantitation were plotted as Figure 8. Surprisingly, a 15 to 20 second lag precedes the accumulation of +11G (filled squares, ■). ATP hydrolysis has been suggested as an energy source during the separation of template strands prior to phosphodiester bond formation in RNA. ATP hydrolysis would also allow for phosphorylation of the CTD on polymerase II in the preinitiation complex. Therefore, ATP was added one minute before the addition of pulse NTPs to determine whether this treatment will shorten the observed lag in stable ternary complexes formation which is precursor of runoff RNA. However, the lag is unaffected (open circles, ○). Even addition of ATP and α-[^{32}P] CTP (filled circles, ●), which provides for both ATP hydrolysis and formation of the first phosphodiester bond from the adenovirus major late promoter, does not suppress the kinetic lag. These results suggest that the lag is not due to either open complex formation, or to first phosphodiester bond formation, or to phosphorylation of CTD on polymerase II. The slow step occurs between formation of the first and tenth phosphodiester bonds.

RAP74 contributes to the stability of ternary complexes. We have previously shown that RAP74 is required for promoter escape (20). One way in which RAP74 might be involved in this process could be to stabilize short transcripts so that they are not released from polymerase. Short transcripts initiated in the presence of full length RAP74 are very stable to washing and can quantitatively be elongated to a runoff position. Since RAP74 mutants have different efficiencies in a runoff transcription assay (37), these mutants were tested to determine whether the yield of runoff products was affected by initiation efficiency or by the stability of ternary complexes.

An extract derived from the nuclei of HeLa cells was depleted of RAP30 and RAP74 by immunoprecipitation with anti-RAP30. Such an extract is completely

Figure 8. A 15 to 20 seconds lag in stable ternary complexes synthesis. RNA synthesis was inhibited by addition of 1 mg/ml α-amanitin (α-A) at the times indicated, stable ternary complexes were isolated , and the +11G AdML transcript was quantitated. Quantitation of total short transcript synthesis was qualitatively very similar to that shown for +11G, and all short AdML RNAs accumulated with approximately the same kinetics (data not shown). 0, 5 and 10-second time points were done in triplicate to be sure that low signals were not due to experimental error. Addition of 100 μM ATP alone (open circles, ○),or 100 mM ATP + <1 μM α^{32}[P]CTP (filled circles, ●), for 1 min prior to addition of pulse ATP, GTP, UTP, and α[^{32}P]CTP, did not suppress the observed lag.

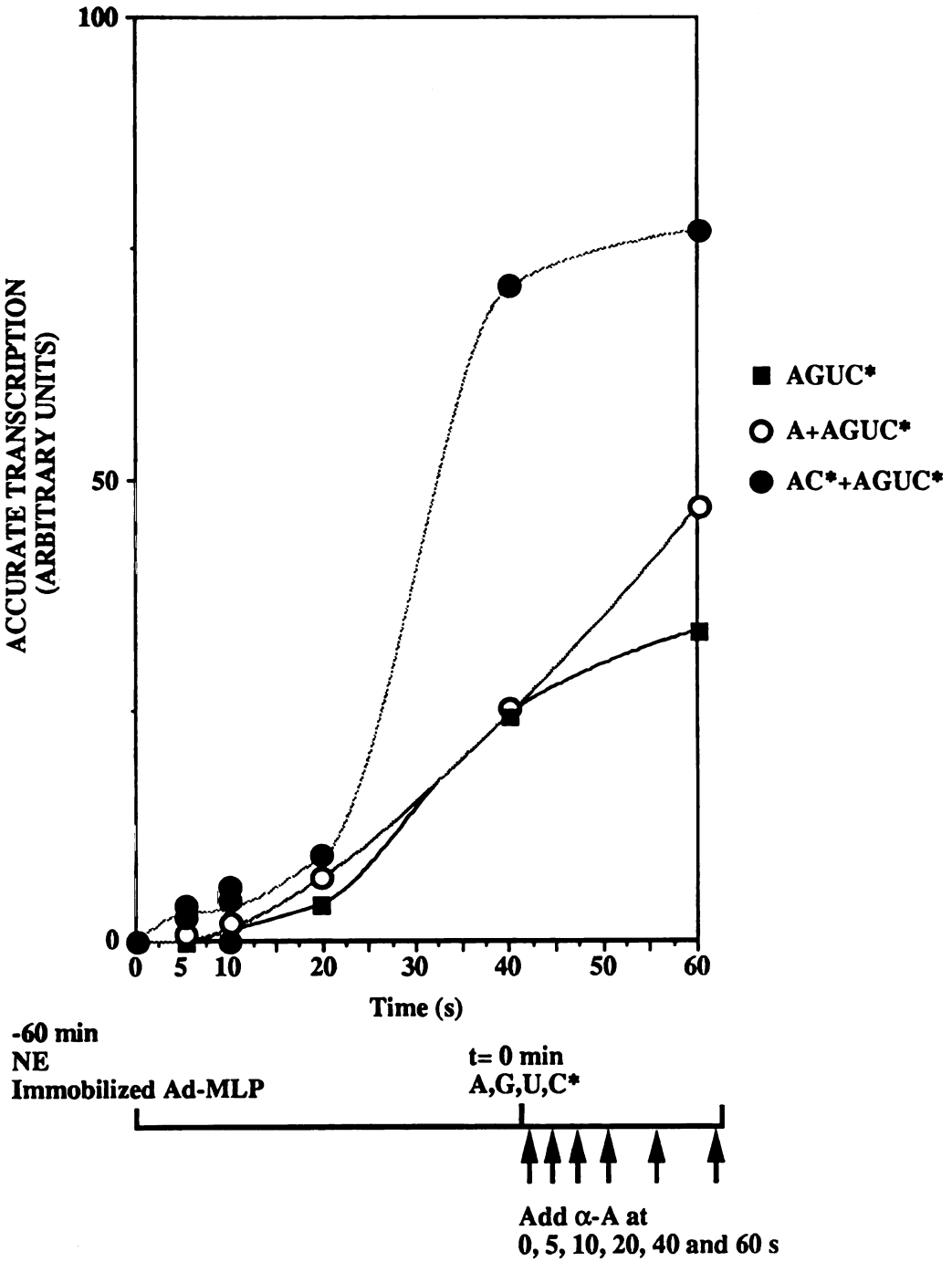

100
50
0
ACCURATE TRANSCRIPTION
(ARBITRARY UNITS)
0 5 10 20 30 40 50 60
Time (s)
AGUC*
A+AGUC*
AC*+AGUC*
-60 min
NE
Immobilized Ad-MLP
t= 0 min
A,G,U,C*
Add α-A at
0, 5, 10, 20, 40 and 60 s

Figure 8

depe

lengt

end

stabl

triph

and

on a

and

chas

RAI

equi

obta

diff

posi

reac

trip

prev

517

pro

diff

attr

was

twi

syn

dependent on addition of RAP30 and RAP74 to restore accurate transcription. Full length RAP74 (1-517), and a set of mutants deleted progressively from the C-terminal end were tested for the ability to support accurate initiation and for the ability to form stable ternary complexes. In previous work, using $\alpha[^{32}P]$-GTP as the limiting nucleoside triphosphate, C-terminal mutants showed an ever-decreasing efficiency of transcription, and mutants deleted beyond 1-205 showed no ability to produce a runoff transcript (37).

The template for transcription is the Adenovirus major late promoter immobilized on agarose beads. Transcripts were initiated with a 1 min pulse with 100 μM ATP, UTP, and GTP, and <1 μM $\alpha[^{32}P]$-CTP. In Figure 9B lanes 1-3, transcripts were immediately chased with a mixture of 1 mM ATP, CTP, UTP and GTP. In single round transcription, RAP74 1-517, 1-205, and 1-172 each supported initiation and runoff transcription with equivalent efficiency. The apparent discrepancy between this result and that previously obtained in which the 1-172 mutant appeared to be inactive may be attributable to the different nucleoside triphosphate concentrations used in the transcription protocols. The position of pausing is dependent on the nucleoside triphosphate that is limiting in the reaction, and the site of pausing may determine transcript stability.

In lanes 4-6, complexes were washed twice with buffer before chase nucleoside triphosphates were added. In this case, the results are much more comparable to those previously obtained with a GTP-limited reaction. The complex initiated with RAP74 1-517 is efficiently elongated to the runoff position. The 1-205 mutant makes a runoff product with much lower efficiency, and the 1-172 mutant makes very little product. The different efficiencies of these mutants for production of the runoff RNA may be attributable to the stability of the initiated complexes.

Short RNA products initiated with these mutants are visualized in Figure 9C after washing. These transcripts were initiated in a 1 min pulse as described above, washed twice, and then visualized on a 23% polyacrylamide gel. Since these transcripts are synthesized under conditions of CTP limitation, the product observed is primarily

Fig

A)

205

Ade

hur

wit

300

ter

µM

1 n

twi

C)

sta

mu

wa

RA

sta

Figure 9. RAP74 is important for the stability of initiated transcription complexes. A) RAP74 deletion mutants. B) Sequences between 205 and 517 and between 172 and 205 can contribute to the efficiency of runoff transcription. Transcription was from the Adenovirus major late promoter immobilized on agarose beads. An extract derived from human HeLa cell nuclei was depleted of RAP30 and RAP74 by immunoprecipitation with anti-RAP30 antibodies. This extract (7 μl) was combined with 150 ng RAP30 and 300 ng RAP74 1-517 or RAP74 deletion mutant and incubated with immobilized template for 60 min. Initiation was by addition of 100 μM ATP, UTP, and GTP, and <1 μM $\alpha[^{32}P]$-CTP for 1 min. Reactions shown in lanes 1-3 were immediately chased with 1 mM ATP, CTP, UTP, and GTP (no wash). Reactions shown in lanes 4-6 were washed twice with 50 μl of wash buffer before addition of chase nucleoside triphosphates (wash). C) Inspection of short RNAs reveals RAP74 sequences that contribute to ternary complex stability. Short RNAs were initiated in the presence of RAP74 1-517 or RAP74 deletion mutants in a 1 min pulse as described for B above. Immobilized ternary complexes were washed twice with buffer and short RNAs were visualized on a 23% acrylamide gel. RAP74 sequences between 409 and 517 and between 172 and 205 contribute to the stability of short ternary complexes.

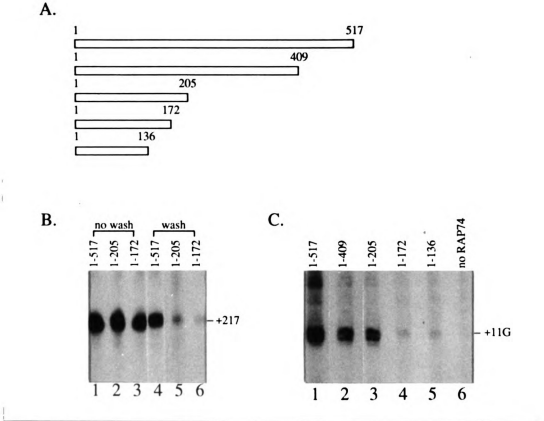

A.
1
517
1
409
1
205
1
172
1
136
B.
no wash
wash
1-517
1-205
1-172
1-517
1-205
1-172
+217
1 2 3 4 5 6
C.
1-517
1-409
1-205
1-172
1-136
no RAP74
+11G
1 2 3 4 5 6

Figure 9

elongated to +11G. The transcript initiated with RAP74 1-517 is most stable. Deletion to 1-409 or 1-205 reduces stability somewhat. A further decrease in ternary complex stability is seen with deletion to 1-172, as predicted from the data shown in Figure 9B lanes 5 and 6. The conclusion of these experiments is that RAP74 mutants as short as 1-172, and possibly shorter, support accurate initiation as well as full length RAP74. Ternary complexes formed with these RAP74 mutants, however, have varying stability. Apparently, sequences between 409 and 517 and between 172 and 205 contribute to the stability of short ternary complexes. This observation is very consistent with the role of RAP74 in promoter escape.

DIS

con

hyd

for

mi

ter

len

ini

the

pos

hy

RN

NT

mi

AT

fir

sy

be

sta

mu

eff

Co

cer

mu

dif

DISCUSSION

Using an immobilized template assay, we have observed the first stable ternary complexes accurately initiated from the AdML promoter. Initiation requires ATP hydrolysis and separation of DNA template strands (39, 40). Once phosphodiester bond formation commences, however, RNA polymerase II must synthesize a chain of a minimum length to form a stable complex (13). Consistent with this idea, no stable ternary complexes shorter than +11G were detected indicating that this is the minimal length for stability (Figure 2). RNA polymerase II pauses transcription shortly after initiation. Varying the nucleoside triphosphate that was limiting in concentration dictated the tendency to pause at particular sites, and in most cases, polymerase stalled at positions preceding addition of the limiting NTP. Synthesis of short RNAs required hydrolysis of ATP and was completely eliminated by addition of α-amanitin. Paused RNAs were quantitatively chased to the expected runoff position with addition of chase NTPs. Short, paused RNAs accumulated after a 15-20 s delay, indicating that this is the minimal time required to synthesize the +11G transcript. Pre-incubation with ATP, or ATP and CTP did not eliminate this delay, indicating that open complex formation and first phosphodiester bond formation do not account for the delay in stable complex synthesis. The delay appears to represent the time required to form phosphodiester bonds between addition of +3U and +11G.

By analysis of mutants, the RAP74 subunit of TFIIF was shown to be required for stability of these short transcripts. RAP74 is a protein of 517 amino acids. Deletion mutants 1-409, 1-205, and 1-172 were all shown to initiate transcription with equivalent efficiency to the full-length protein, but did not support complexes of equivalent stability. Complexes formed in the presence of the full length protein showed no sensitivity to centrifugation and washing. A subset of complexes formed in the presence of the 1-409 mutant were stable but a fraction were disrupted by centrifugation and washing. No differences were noted in stability between the 1-409, 1-356, 1-296 and 1-205 mutants.

Sequence between 409-517 appears to be required for stability of a subset of complexes (about 60%) but not for all. The 1-172 mutant showed lower stability than these longer proteins, although this mutant showed equivalent initiation efficiency. In most cases, two washes completely dissociated the ternary complex formed in the presence of the 1-172 mutant, but in some experiments this mutant has been observed to have some ability to stabilize these complexes (Figure 9B, lane 3, and data not shown), so part of the stability region remains intact in the 1-172 mutant. Essentially all of the newly-initiated complexes appear to require sequence between amino acids 136-205 of RAP74 for stability. Thus, there appear to be at least two distinct classes of elongation complex that may contain slightly different constellations of associated transcription factors. One class requires C-terminal sequences between 409-517. The other class requires sequence between 1-205 but not C-terminal sequences. It is likely that the class that requires C-terminal sequences also requires the N-terminal region, but this point cannot clearly be made by our experiment.

The observation that RAP74 stabilizes short ternary complexes is very consistent with results previously published by our laboratory in which we showed that RAP74 could be completely dispensible for accurate initiation in an extract transcription system (20). Evidence for RAP74-independent initiation was obtained using two methods: 1) RNA was pulse-labeled in the absence of RAP74; and 2) sarkosyl-resistant complexes were formed in the absence of RAP74. In both cases, however, RAP74 was required to elongate the RNA. This result has been somewhat controversial in the literature, however, because RAP74 can clearly contribute to initiation functions. RAP74 contributes to the stability of pre-initiation complexes in polyacrylamide gel mobility shift assays (19). RAP74 also must incorporate into a complex consisting of TBP, TFIIB, RAP30, and RNA polymerase II, for RAP30 to cross-link to DNA (41). Normal assembly of RAP30, therefore, appears to require RAP74. In another report, both RAP30 and RAP74 were required to produce runoff products and short RNAs (42). In our own

unpublished data, we had similarly observed that in some extract systems, both RAP30 and RAP74 were required to initiate and recover stable transcripts.

To reconcile these apparently inconsistent observations, we speculated that initiation by RNA polymerase II did not require the RAP74 subunit of TFIIF, as we had previously observed (20). We further considered that other unidentified factors, present in some systems but absent from others, could stabilize ternary complexes and prevent release of short RNAs from polymerase. Such a stabilizing factor would be similar to *Drosophila* p-TEF (positive transcription elongation factor) described by Price and co-workers (32). In the presence of such a factor or factors, initiation would occur in the absence of RAP74 and the ternary complex would remain stably associated with polymerase. Addition of RAP74 would still be necessary for early elongation because RAP74 is required for promoter escape (20). In a system missing such a stabilizing factor, RAP74 will appear to be required for initiation, but the requirement in this case is to stabilize the ternary complex (Figure 9).

A model is presented to describe the functions of RAP74 in accurate initiation and promoter escape from the AdML promoter (Figure 10). This model has features that are consistent with other models recently presented for RNA polymerase II elongation and "inchworming" (38), and this model can reconcile our and other's experimental results for TFIIF function. RNA polymerase II is indicated as an oval with two structures that interact with RNA, the catalytic center and an RNA binding site (RBS). The catalytic center is the site at which phosphodiester bonds are formed. The RBS is a site at which RNA is tightly bound, but through which RNA translocates during elongation.

Abortive initiation occurs when polymerase synthesizes RNAs that are too short to occupy RBS (<+11G). Such RNAs are released, probably without dissociation or movement of polymerase, and re-initiation can subsequently occur (13, 35). Synthesis of short RNAs (+11G or longer) fills RBS, and this process requires 15-20 s (Figure 8). Occupancy of the RBS is the step designated promoter escape I. The complex that is

Figure 10. A model describing the role of RAP74 in accurate transcription and promoter escape by RNA polymerase II. The catalytic center of RNA polymerase II is the site of phosphodiester bond formation. RBS is an RNA binding site on polymerase through which the chain must translocate. See the text for details.

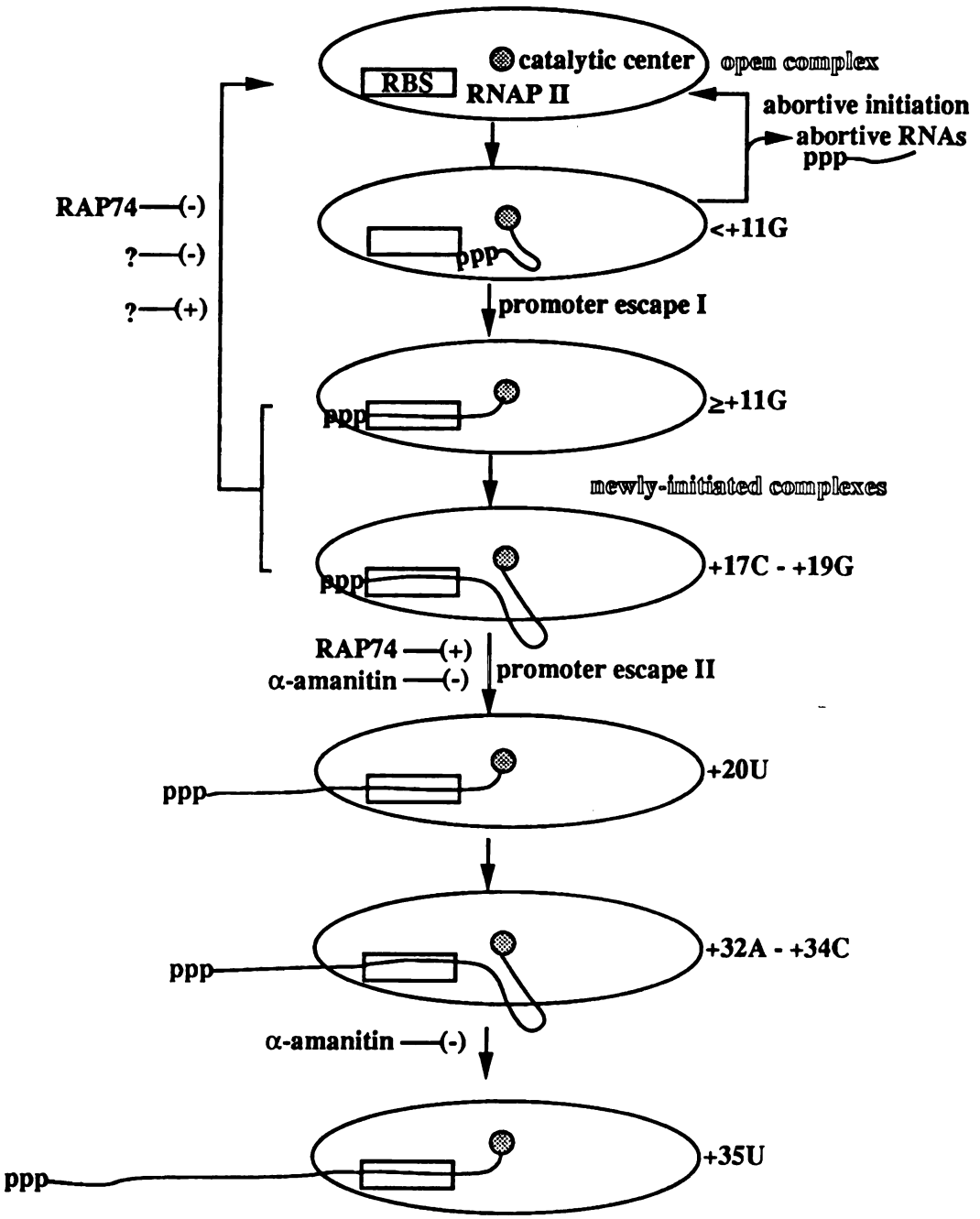

catalytic center
RBS
RNAP II
open complex
abortive initiation
abortive RNAs
ppp
RAP74—(-)
?—(-)
?—(+)
<+11G
promoter escape I
≥+11G
newly-initiated complexes
+17C - +19G
RAP74—(+)
α-amanitin—(-)
promoter escape II
+20U
+32A - +34C
α-amanitin—(-)
+35U

Figure 10

formed is similar to those previously described as initiated or sarkosyl-resistant complexes (20). Goodrich and Tjian (1994) have shown that TFIIE and TFIIH are required for promoter escape I, but our data demonstrate that the RAP74 subunit of TFIIF can be dispensible for this step (43). Other factors that have not been clearly identified enhance stability of the complex.

Particular sequences within RAP74 are required to preserve these complexes. The sequence between amino acids 409-517, at the very C-terminus of RAP74, is involved in making contacts with RNA polymerase II (B.Q. Wang and Z.F. Burton, submitted), TFIIB (S.M. Fang and Z.F. Burton, in preparation), and DNA (B.Q. Wang and Z.F. Burton, submitted). RAP74 sequence between 172 and 205, which appears to be critical for stability, is involved in binding to the RAP30 TFIIF subunit.

"Inchworming" by RNA polymerases has been described as a shift in the distance between two RNA interaction sites (38). This can be drawn as a change in polymerase conformation or, as indicated here, as the RNA distance between the catalytic center and the RBS. According to either view, polymerase reaches boundaries to elongation at which a transition within the enzyme must occur for synthesis to continue. This transition involves shortening the RNA length between the catalytic center and the RBS and translocation of RNA through the RBS.

Our data indicate that RNA polymerase II may undergo two inchworm transitions as it escapes the AdML promoter. When α-aminitin is added with NTPs to initiated complexes, elongation proceeds to two distinct boundaries near +19G and +34C. α-amanitin, therefore, allows phosphodiester bond formation, but blocks the inchworm transition and chain translocation, as indicated in the figure. According to this view, α-amanitin blocks a conformational change in polymerase that is required to decrease the RNA distance between the catalytic center and RBS, or blocks RNA translocation through the RBS, or both.

This same transition that is sensitive to α-amanitin may be promoted by TFIIF, since RAP74 is required for early elongation from the AdML promoter (20). This step is designated promoter escape II. Promotion of similar transitions during many steps in subsequent synthesis could explain how TFIIF stimulates elongation rates (44).

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health, The Michigan State University Agricultural Experiment Station, and Michigan State University.

REFERENCES

1. Zawel, L. and Reinberg, D. (1993) Initiation of transcription by RNA polymerase II: a multi-step process. Progress in Nucleic Acid Res. 44, 67-108.
2. Conaway, R.C. and Conaway, J.W. (1993) General initiation factors for RNA polymerase II. Annu. Rev. Biochem. 62, 161-190.
3. Gross, C. A., Lonetto, M., and Losick, R. (1992) Transcriptional Regulation (McKnight, S. L., and Yamamoto, K. R., eds) pp. 129-178, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
4. von Hippel, P. H., Yager, T. D., and Gill, S. C. (1992) Transcriptional Regulation (McKnight, S. L., and Yamamoto, K. R., eds) pp. 179-202, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
5. Drombowski, A.J., Walter, W.A., Record, M.T. Jr., Siegele, D.A. and Gross, C.A. (1992) Polypeptides containing highly conserved regions of transcription initiation factor σ70 exhibit specificity of binding to promoter DNA. Cell 70, 501-512.
6. Carpousis, A. J., and Gralla, J. D. (1980) Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vivo at the lac UV5 promoter. Biochemistry 19, 3245-3253.
7. Carpousis, A. J., and Gralla, J. D. (1985) Interaction of RNA polymerase with lac UV5 promoter DNA during mRNA initiation and elongation. Footprinting methylation and rifampicin-sensitivity changes accompanying transcription initiation. J. Mol. Biol. 183, 165-177.
8. Hansen, U. M., and Mclure, W. R. (1980) Role of the sigma subunit of E. coli RNA polymerase in initiation. J. Biol. Chem. 255, 9564-9570.
9. Straney, D. C., and Crothers, D. M. (1985) Intermediates in transcription initiation from *E. coli* lac UV5 promoter. Cell 43, 449-459.
10. Levin, J. R., Krummel, B., and Chamberlin, M. J. (1987) Isolation and properties of transcribing ternary complexes of *Escherichia coli* RNA polymerase positioned at a single template base. J. Mol. Biol. 196, 85-100.
11. Krummel, B., and Chamberlin, M. J., (1989) RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes. Biochemistry 28, 7829-7842.
12. Gill, S.C., Weitzel, S.E. and von Hippel, P.H. (1991) *Escherichia coli* σ70 and NusA proteins I. Binding interactions with core RNA polymerase in solution and within the transcription complex. J. Mol. Biol. 220, 307-324.
13. Luse, D.S. and Jacob, G.A. (1987) Abortive initiation by RNA polymerase II in vitro at the Adenovirus 2 major late promoter. J. Biol. Chem. 262, 14990-14997.
14. Horwitz, R. J., Li, J. and Greenblatt, J. (1987) An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage lambda. Cell 51, 631-641.
15. Roberts,J. W. (1988) Phage lambda and the regulation of transcription termination. Cell 52, 5-6.
16. Buratowski, S., Hahn, S., Guarante, L., Sharp, P. A. (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56, 549-561.
17. Maldonado, E., Ha, I., Cortes, P. Weis, L., Reinberg, D. (1990) Factors involve in specific transcription by mammalian RNA polymerase II: Roles of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. Mol. Cell. Biol. 10, 6335-6347.
18. Killeen, M. and Greenblatt, J. (1992) The general transcription factor RAP30 binds to RNA polymerase II and prevent it from binding nonspecifically to DNA. Molec. Cell. Biol. 12, 30-37.

19. Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F. and Reinberg, D. (1991) The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc. Natl. Acad. Sci. USA 88, 9999-10003.
20. Chang, C.-h., Kostrub, C. F. and Burton, Z. F. (1993) RAP30/74 (Transcription Factor IIF) is required for promoter escape by RNA polymerase II. J. Biol. Chem. 268, 20482-20489.
21. Corden, J.L. (1990) Tails of RNA polymerase II. Trends Biol. Sci. 15, 383-387.
22. Flores, O., Lu, H. and Reinberg, D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II: identification and characterization of factor IIH. J. Biol. Chem. 267, 2786-2793.
23. Lu, H., Zawel, L., Fisher, L., Egly, J.-M. and Reinberg, D. (1992) Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature 358, 641-645.
24. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H. J., Chambon, P. and Egly, J.-M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science 260, 58-63.
25. Gulyas, K.D. and Donahue, T.F. (1992) *SSL2*, a suppressor of a stem-loop mutation in the *HIS4* leader encodes the yeast homolog of human *ERCC-3*. Cell 69, 1031-1042.
26. Maxon, M. E., Goodrich, J. A., and Tjian, R. (1994) Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes and Develop. 8, 515-524
27. Lis, J. and Wu, C. (1993) Protein traffic on the heat shock promoter: parking, stalling and trucking along. Cell 74, 1-4.
28. Spencer, C.A. and Groudine, M. (1990) Transcription elongation and eukaryotic gene regulation. Oncogene 5, 777-785.
29. Giardina, C., Perez-Riba, M. and Lis, J.T. (1992) Promoter melting and TFIID complexes on *Drosophila* genes in vivo. Genes and Develop. 6, 2190-2200.
30. Krumm, A., Meulia, T., Brunvand, M. and Groudine, M. (1992) The block to transcriptional elongation within the human *c-myc* gene is determined in the promoter-proximal region. Genes and Develop. 6, 2201-2213.
31. Arias, J. and Dynan, W.S. (1989) Promoter dependent transcription by RNA polymerase II using immobilized enzyme complexes. J. Biol. Chem. 264, 3223-3229.
32. Marshall, N. F., and Price, D. H. (1992) Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol. Cell. Biol. 12, 2078-2090.
33. Shapiro, D.J., Sharp, P.A., Wahli, W.W. and Keller, M.J. (1988) A high efficeincy HeLa cell nuclear transcription extract. DNA 7, 47-55.
34. Jacob, G. A., Luse , S. W., and Luse, D. S. (1991) Abortive initiaiton is increased only for the weakest two members of a set of down mutants of the Adenovirus 2 major late promoter. J. Biol. Chem. 266, 22537-22544.
35. Jacob, G. A., Kitzmiller, J. A., and Luse. D. S. (1994) RNA polymerase II promoter strength in vitro mat be reduced by defects at initiaiton or promoter clearance. J. Biol. Chem. 269, 3655-3663.
36. Goliger, J. A., and Roberts, J.W. (1989) Sequence required for antitermination by phage 82 Q protein. J. Mol. Biol. 210, 461-471.
37. Wang, B.Q., and Burton, Z. F. (1995) Functional domains of human RAP74 including a masked polymerase binding domain. Submit to J. Biol. Chem.
38. Chamberlin, M. J. (1995). New models for the mechanism of transcription elongation and its regulation. Harvey Lect. 88, 1-21.
39. Bunick, D., Zandomeni, R., Ackerman, S. and Weinmann, R. (1982). Cell 29, 877-886.
40. Sawadogo, M. and Roeder, R. G. (1984). J. Biol. Chem. 259, 5321-3526.
41. Coulombe, B., Li, J., and Greenblatt, J. (1994) Topological localization of the human transcription factors IIA, IIB, TATA box binding protein, and RNA polymerase II-associated protein 30 on a class II promoter. J. Biol. Chem. 269,19962-19967.

42. Tan, S., Aso, T., Conaway, R. C., and Conaway, J. W. (1994). Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. J. Biol. Chem. 269, 25684-25691.
43. Goodrich, J. A. and Tjian, R. (1994). Transcription factors IIE, IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II Cell 77, 145-156.
44. Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991). Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. Mol. Cell. Biol. 11, 1195-1206.

CHAPTER IV

YEAST Cdc73p MAY BE A COMPONENT OF AN ALTERNATE VERSION OF TFIIF

ABSTRACT

Cdc73p was isolated using a monoclonal antibody directed against the carboxy terminal domain of RNA polymerase II. Comparison of Cdc73p sequence with that of human RAP30 and bacterial sigma and delta transcription factors led to the idea that this diverse collection of proteins might have similar functions. We propose the model that Cdc73p is a subunit of an alternate form of TFIIF, that functions in transcription of an alternate promoter class recognized by RNA polymerase II, in much the same manner that bacteria utilize different sigma factors to transcribe from alternate promoters. This model makes many predictions that are subject to experimental test. One prediction is that Cdc73p will interact directly with RNA polymerase II through a conserved polymerase binding region. Consistent with this prediction, Cdc73p binds to RNA polymerase II. Cdc73p appears as an additional polymerase subunit, when polymerase is separated from free Cdc73p by gel filtration. Cdc73p sequences within the predicted polymerase interaction region are critical for binding. Another prediction of this model is that Cdc73p may interact with Tfg3p/Anc1p, the smallest subunit of yeast TFIIF. In a preliminary experiment, this expectation appears also to be correct. Several other tests of this model are in progress, and these approaches are discussed.

INTRODUCTION

The yeast gene *CDC73* was initially identified in a suppressor screen of *ste2* deletion mutants that are defective in mating. Ste2p is the α factor receptor. Binding of α factor to Ste2p on **a** type cells stimulates a signal transduction cascade that results in cell cycle arrest at G1/S (1). Cell cycle arrest is necessary prior to fusion of **a** and α type cells to create **a**/α diploids.

The *CDC* designation stands for cell division cycle, and temperature sensitive *cdc73* mutants arrest cell division when shifted to the non-permissive temperature, although it appears that arrest occurs at G2/M rather than at G1/S (2). *CDC73* does not appear to function directly in α factor signalling, and its participation in mating may be indirect, perhaps through its function in regulating the cell cycle.

Transcription factor TFIIF in yeast is composed of three subunits named Tfg1p/Ssu71p, Tfg2p, and Tfg3p/Anc1p (3). In humans TFIIF is composed of two distinct subunits, RAP74 which is homologous to Tfg1p/Ssu71p, and RAP30 which is homologous to Tfg2p. TFIIF is required for accurate initiation by RNA polymerase II and stimulates elongation of transcription (4, 5). TFIIF assists polymerase binding to the promoter and so this factor is at the core of the assembly pathway for an active pre-initiation complex (6). Because of its essential role in polymerase entry, TFIIF must join the complex before TFIIE and TFIIH can stably enter. To a large extent, TFIIF directs the final assembly of the transcription complex.

Human RAP30 is related structurally and functionally to bacterial sigma factors. (7, 8) Our analysis indicates that RAP30 may have sequence similarity to conserved regions 1.2, 2.1, 3.1, and 4.1 of bacterial sigma factors and a region of bacterial delta factor (9). Region 2.1 is the most important sequence for polymerase binding by sigma factors, and mutagenic analysis shows that this is the polymerase binding region of RAP30 as well (10, 11). The C-terminal region of RAP30 contains a masked DNA-

binding domain, just as do similar regions 3.1 and 4.1 of sigma factors (12, 13). RAP30 alters template contacts by polymerase in a manner analogous to bacterial sigma factors, and the functions of these proteins are similar in directing polymerase to the promoter.

Yeast RNA polymerase II has 27 repeats of the heptapeptide consensus YSPTSPS at the C-terminus of the largest subunit (14). Using affinity chromatography with antibody directed against this carboxy terminal domain (CTD) as a leash to capture RNA polymerase II, our laboratory identified TFIIF subunits and Cdc73p as polymerase associated proteins (15). Based on a sequence comparison between Cdc73p, human RAP30, yeast Tfg2p, and bacterial sigma factors, we concluded that yeast Cdc73p was likely to be a subunit of an alternate form of TFIIF (9). Cdc73p appears to be similar in structure to bacterial sigma factors within conserved regions 1.2, 2.1, and 3.1. Cdc73p also has an extended similarity to bacterial delta factor (9). In this report, we have subjected this model to some preliminary experimental tests.

MATERIALS AND METHODS

Expression vector and strains. T7 expression plasmid pET21a and expression strain BL21(DE3) were purchased from Novagen. Manipulation of the expression system was according to manufacturer's instructions and has been described previously (16).

The CDC73 gene was amplified by a standard PCR reaction. The 5' primer (5'-CATATGGCGAACTCATTAGACAGACTGAGAG-3') was designed to create an Nde I site at the position of initiator methionine. The 3' primer (5'-GCGGCCGCACGGTATCC TCTTGAAATAAGTTCC-3') was designed to create an Not I site to fuse the CDC73 open reading frame to a sequence terminated with six histidines. The Nde I to NotI fragment was cloned between the Nde I and Not I sites of the pET21a vector.

The internal deletion mutant, cdc73pΔ129-143, was generated as follows. The first 128 amino acids of Cdc73p using 5'-primer as above and the following primer 5'-CTCGAGGTCTCGCCTTTCTGACCGGGTGCTT-3' which introduces an XhoI site at the end of the first 128 amino acids of Cdc73p. This NdeI-XhoI fragment was cloned into pET21a/CDC73 which had been previously digested by NdeI and SalI, preserving the reading frame and precisely deleting amino acids 129-143.

The C-terminal deletion mutant, cdc73p1-143 was created by digesting the pET21a/CDC73 with restriction enzymes SalI and XhoI, filling in with Klenow enzyme and religating.

Production and purification of Cdc73p, cdc73p Δ129-149 and cdc73p 1-143. Bacterial cultures were grown up to optimal density, induced with IPTG and harvested as described previously (17). Proteins were purified by Ni^{2+} affinity chromatography purchased from QIAGENE, using a previously described protocol (17).

Production of anti-Cdc73p antiserum. Antisera against Cdc73p was produced in rabbits. Titer Max (CtyRx) was included as the adjuvant in the initial injection but not

in the following booster injections. The protocol is essentially the same as that previously used for production of anti-RAP30 and anti-RAP74 antiserum (17).

Gel filtration chromatography binding assay. A three-fold molar excess of Cdc73p, or Δ129-143, or 1-143 were mixed with purified yeast RNA polymerase II in 0.5 M KCl storage buffer and centrifuged through an Amicon filter to concentrate the sample to a volume smaller than 200 μl. Samples were dialyzed against 0.1 M KCl storage buffer (20 mM Hepes pH 7.9, 20 % glycerol, 0.5 M KCl, 0.2 mM EGTA, 0.2 mM EDTA and 2 mM DTT) overnight at 4°C to allow binding. A Superdex 200 HR 10/30 gel filtration column purchased from Pharmacia was equilibrated with 0.1 M KCl storage buffer (20 mM Hepes pH 7.9, 20 % glycerol, 0.1 M KCl, 0.2 mM EGTA, 0.2 mM EDTA and 2 mM DTT) at a flow rate of 0.5 ml/min until a stable base line was reached. Samples were cleared by brief centrifugation before injection. Peaks from elution profiles were collected and spin concentrated before loading on SDS PAGE. Proteins were visualized by Coomassie blue staining.

Anc1p binding assay. Anc1p (120 μg) was immobilized on Affi-gel 10 (Pierce) as reported by Lei and Burton (18). Non-specific binding was blocked by incubating the immobilized Anc1p and control resin with reaction buffer containing 20 mM Hepes, pH 7.9, 20 % glycerol, 0.5 M KCl, , 0.2 mM EGTA, 0.2 mM EDTA, 2 mM DTT and 0.2% BSA at 4°C for one hour. Cdc73p (20 μg) was added and the sample incubated for another 30 min. The beads were collected and washed 5 times with the same reaction buffer but without BSA. Bound Cdc73p was eluted with 40 μl 0.5 M KCl reaction buffer. The eluates were analyzed by SDS PAGE, and proteins were stained by silver staining.

RESULTS

Cdc73p is associated with yeast RNA polymerase II. A yeast whole cell extract was fractionated by affinity chromatography on a column containing covalently immobilized monoclonal antibodies directed against the CTD of RNA polymerase II (19). This antibody captures polymerase along with a complex set of associated proteins (RAPs), which can be dissociated from the core enzyme by treatment with moderate salt (ie. 0.5 M KCl). A negative control column contains an immobilized monoclonal antibody of the same isotype (IgG2a), directed against a protein not found in yeast (15). Initially, RAPs were identified as those proteins that bind to the anti-CTD column but not to the negative control column. The amino acid sequence of an internal peptide derived from a RAP of approximately 54 kDa was identical to a sequence found within the coding frame of a previously identified yeast gene *CDC73*. The selectivity of Cdc73p binding to the anti-CTD column is confirmed in the experiment shown in Figure 1. The 0.5 M KCl eluates of the anti-CTD and anti-β' columns are compared in a Western blot developed with anti-Cdc73p antibodies. Cdc73p is detected in the eluate of the anti-CTD column (lane 2) but not in the eluate of the anti-β' column (lane 3). The recombinant Cdc73p-H6 marker appears slightly larger than the normal yeast protein because of the C-terminal polyhistidine extension.

Cdc73p may share sequence similarity to human RAP30 and bacterial σ factors within their RNA polymerase binding regions. Inspection of the amino acid sequence of Cdc73p indicated that this protein might be evolutionarily related to human RAP30 and bacterial sigma factors within the regions of these proteins that binds to their respective RNA polymerases (9). In Figure 2, a proposed alignment of yeast Cdc73p with the RNA polymerase II binding region of human, *Xenopus*, and *Drosophila* RAP30, and the RNA polymerase binding region of bacterial sigma factors is shown. Hydrophobic cluster plots of the sequences used in the alignment are also shown.

Figure 1. Cdc73p is associated with yeast RNA polymerase II. A yeast whole cell extract that is active in accurate transcription was fractionated over columns containing covalently immobilized monoclonal antibodies directed against the CTD of RNA polymerase II or against the *E. coli* RNA polymerase subunit β'. Yeast RNA polymerase II and associated proteins were immobilized by binding to the anti-CTD antibody. These proteins were not expected to bind to the column containing anti-β' antibody. The 0.5 M KCl eluates of the anti-CTD column (α-CTD) and the negative control column (α-β') (80 μl each) were analyzed in a Western blot developed with anti-Cdc73p antibodies. 100 ng of recombinant Cdc73p-H6 (lane 1) is shown as a reference standard.

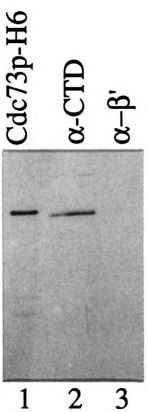


Figure 1

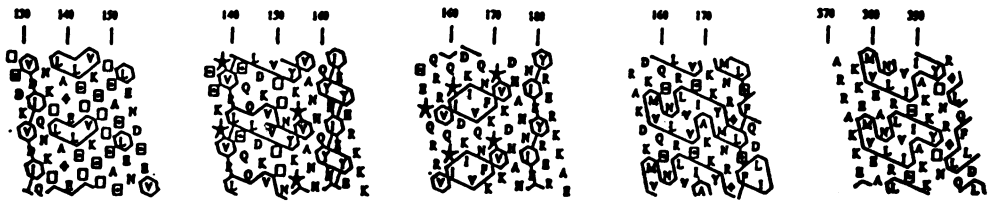
Figure 2. Cdc73p may share sequence similarity to human RAP30 and bacterial σ factors within their RNA polymerase binding regions. At the bottom of the figure are hydrophobic cluster analysis plots of polymerase binding regions of human and *Drosophila* RAP30, and two bacterial sigma factors, sigma A (*Anabena*) and sigma 70 (*E. coli*). These plots are the basis for the alignments shown above. This analysis was performed by Shi Min Fang.

required for pol II binding

```
                                  |-----------------------------------|
hRAP30    142   SQQLDKVVTTNYKPVANHQYN.IEYERKKKEDGKR   176
xRAP30    156   SQQLEKAVTSNYKPVSNHQYN.IEYEKKKKDDGKR   190
dRAP30    158   VQPIDKIV.QNFKPVKDHAHN.IEYRERKKAEGKK   192
                ** ** * ** ** ** * *
Cdc73p    131   VDIQNKTLAGELSTVKSTTSASLENDSEVSDPVVE   165
                |-----------------|
          required for pol II binding

sigma70         RRAKDKMVQSNLRLVVSLAKK.YMNRGLSFQDLIQ
sigmaA          RRAKDKMVQSNLRLVVSLAKK.YMNRGLSFQDLIQ
sigmaG          DSAREKLVNGNLRLVLSVIQR.FNNRGEYVDDLFQ
sigmaF          QQARDLLIEKNMRLVWSVVQR.FLNRGYEPDDLFQ
sigmaH          SDALDYLITKYRNFVRAKARS.YFLIGADREDIVQ
```

(Bold indicates match to >50% of consensus of sigma factors, underline indicates 30-50%, and italic indicates 10-30%, calculated from 31 sigma factors. Stars indicate conserved residues between Cdc73p and RAP30.)



Cdc73p hRAP30 dRAP30 sigmaA sigma70

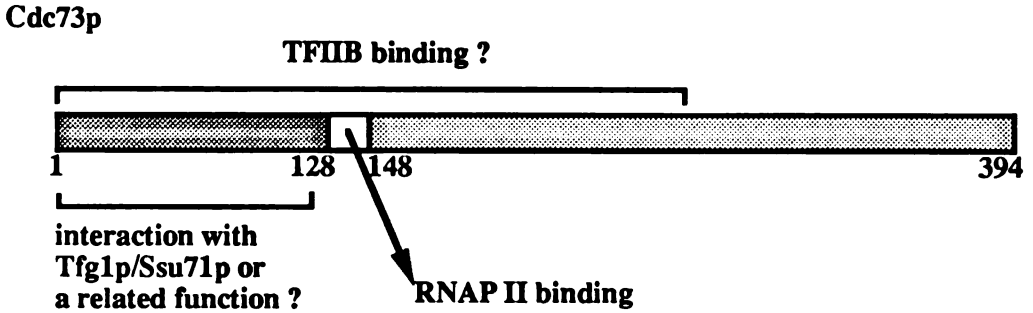
Figure 2

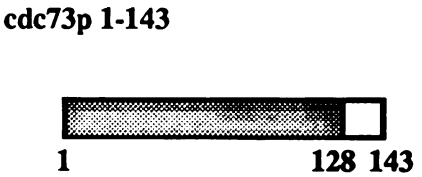
Additional sequence similarities between these proteins (not shown) further suggested that these alignments might be correct.

Production and purification of Cdc73p and Cdc73p mutant proteins. Binding to RNA polymerase II detected by anti-CTD chromatography (Figure 1) might either be direct or indirect. If Cdc73p includes a polymerase binding region, as we propose, Cdc73p should bind directly to RNA polymerase II. Our sequence alignment further predicted that binding would be sensitive to mutation within the proposed polymerase binding domain. Two cdc73p mutant proteins were constructed to determine the importance of this sequence for binding (Figure 3). Mutant cdc73pΔ129-143 has a 15 amino acid deletion within the proposed polymerase binding site, so this mutant was expected to exhibit decreased binding affinity. Mutant cdc73p1-143 has most of its polymerase binding domain intact, but the C-terminal sequences are deleted. The production and purification of Cdc73p and Cdc73p mutant proteins is shown in Figure 4. The cdc73p1-143 mutant was expected to bind RNA polymerase II because this mutant was previously shown to have partial function in vivo (1). Complete truncation of *CDC73* is lethal and the 1-143 truncation is temperature sensitive.

The predicted interaction region of Cdc73p is important for RNA polymerase II binding. Binding to RNA polymerase II was tested using a gel filtration protocol. Yeast RNA polymerase II is a large complex of 10 distinct subunits with a molecular weight of approximately 500 kDa. Cdc73p at 44 kDa appears to be monomeric by gel filtration and is easily separated from RNA polymerase II (data not shown). In Figure 5, Cdc73p and cdc73p mutants were tested for polymerase binding. RNA polymerase II was combined with a three fold molar excess of Cdc73p or cdc73p mutant in buffer containing 0.5 M KCl. Since this buffer was known to release Cdc73p from an anti-CTD column, binding was not expected under this condition, but the solubility of the recombinant proteins was maintained. Binding reactions were then concentrated by centrifugation through a filter and dialyzed against 0.1 M KCl. Under this condition,

Figure 3. Cdc73p mutant proteins. The proposed RNA polymerase II-binding region is indicated as a white box.


Cdc73p
TFIIB binding ?
1
128
148
394
interaction with
Tfg1p/Ssu71p or
a related function ?
RNAP II binding


cdc73p 1-143
1
128
143


cdc73pΔ129-143
1
129
143
394

Figure 3

Figure 4. Production and purification of Cdc73p mutant proteins. Total *E. coli* proteins were visualized 0 and 3 hr after addition of IPTG to cultures carrying production vectors for Cdc73p, cdc73pΔ129-143, and cdc73p1-143. Sample sizes correspond to 0.07 A_{600nm} units of bacterial cells. P indicates recombinant proteins after purification by Ni^{2+} affinity chromatography.

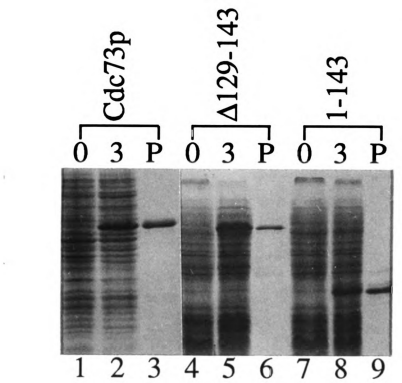


Figure 4

Figure 5. The predicted interaction region of Cdc73p is important for RNA polymerase II binding. Yeast RNA polymerase II (70 μg) was combined with Cdc73p (20 μg; lane 2), cdc73pΔ129-143 (20 μg; lane 4), or cdc73p1-143 (10 μg; lane 6) in buffer containing 0.5 M KCl. Lane 3 contains a sample with RNA polymerase II alone (2.5 μg). Samples were concentrated by centrifugation through an Amicon filter and dialyzed into buffer containing 0.1 M KCl. Insoluble proteins were removed from the solution by centrifugation, and samples were then injected into a Pharmacia Superdex 200 HR10/30 gel filtration column at a flow rate of 0.5 ml per min. RNA polymerase II eluted at 17.5 min after sample injection. Cdc73p, cdc73pΔ129-143, and cdc73p1-143 eluted at 27, 27, and 29.5 min, when fractionated in the absence of RNA polymerase II (data not shown). Apparently, these proteins are monomeric in solution. 280 nm absorbance peaks centered around 17.5 min, corresponding to RNA polymerase II and bound proteins, were collected for each sample, concentrated, and analyzed by SDS-PAGE. RNA polymerase II subunits RPB1-10 are indicated by their apparent molecular weights in kDa along the left side of the figure. Cdc73p (lane 1), cdc73pΔ129-143 (lane 5), and cdc73p1-143 (lane 7) (2 μg each) are included as markers.

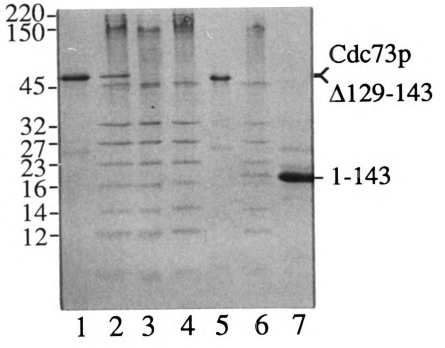


Figure 5

binding to polymerase was expected, but some of the recombinant proteins were expected to precipitate from solution. After removing insoluble protein by centrifugation, samples were injected into a Superdex 200 column (Pharmacia), and the 280 nm absorbance peak corresponding to RNA polymerase II was collected. This fraction was concentrated and analyzed by SDS-PAGE.

Cdc73p bound directly to RNA polymerase II and appeared as an additional band just above the 45 kDa RPB3 subunit (lane 2). Recombinant Cdc73p-H6 is 46 kDa, but migrates slightly slower than expected based on its molecular weight. This protein appears as an additional subunit of polymerase in this experiment. Similarly, the cdc73p1-143 mutant bound to polymerase and appeared as an additional subunit (lane 6). The cdc73pΔ129-143 mutant, however, was severely inhibited for polymerase binding (lane 4). Residual interaction indicates that additional sequences to the 15 deleted amino acids contribute to binding, but demonstrates the critival importance of this short sequence for binding.

Cdc73p may interact with the small subunit of yeast TFIIF. It is likely that most messenger RNA genes are transcribed by the primary form of TFIIF in yeast, which includes three distinct subunits Tfg1p/Ssu71p, Tfg2p, and Tfg3p/Anc1p. Based on the model that Cdc73p is a component of an alternate form of TFIIF, the question arises of whether this protein cooperates with similar subunit partners to its proposed analogue, Tfg2p. Cdc73p might, therefore, interact with Tfg1p/Ssu71p and Tfg3p/Anc1p, or with alternate versions of these proteins. In Figure 6, binding between Cdc73p and Tfg3p/Anc1p was tested in vitro using affinity beads on which Tfg3p/Anc1p was immobilized. Cdc73p bound to these beads in buffer containing 0.1 M KCl and was at least partially eluted with buffer containing 0.5 M KCl. Cdc73p showed no affinity for beads with no immobilized protein ligand. This data is preliminary but may indicate a functional interaction between Cdc73p and Tfg3p/Anc1p. Such an interaction would be consistent with the proposed function for Cdc73p.

Figure 6. Cdc73p interacts with Anc1p/Tfg3p. 30 µl Agarose gel beads with yeast Anc1p/Tfg3p (Anc1p) covalently immobilized (~4 mg bound protein per ml resin) or with no protein ligand (control) were mixed with Cdc73p (20 µg). After washing the beads 5 times with buffer containing 0.1 M KCl, bound protein was eluted with buffer containing 0.5 M KCl, and analyzed by SDS-PAGE developed with silver nitrate.

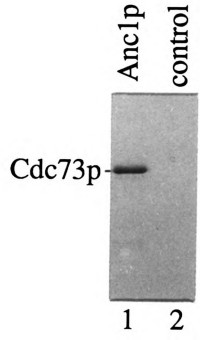


Figure 6

DISCUSSION

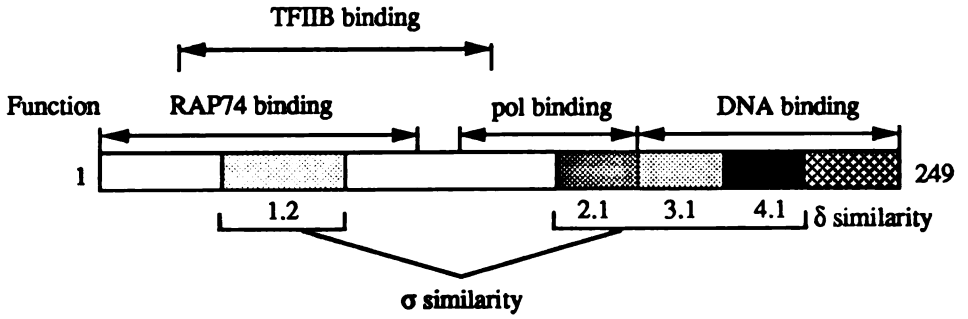
Cdc73p shares structural similarity to a TFIIF subunit and to bacterial transcription factors sigma and delta (Figure 7). Like TFIIF, sigma, and delta, Cdc73p can be isolated based on its affinity for RNA polymerase. Cdc73p binds directly to RNA polymerase II, and we have been able to use structural similarity to predict sequences that are important for polymerase binding. Cdc73p also appears to interact with Tfg3p/Anc1p, a probable transcriptional mediator that interacts with yeast TFIIF and is isolated as the smallest TFIIF subunit. Based on this preliminary analysis, we propose that Cdc73p is a subunit of an alternate form of TFIIF in yeast. Most likely, related functions will be discovered in other eukaryotic organisms.

The N-terminal region of RAP30 binds to the RAP74 subunit of TFIIF (11). The N-terminal region of Cdc73p shows some similarity to RAP30 and sigma factors in this region. Cdc73p may interact with Tfg1p/Ssu71p, the large subunit of yeast TFIIF, or with an alternate version of this factor that has yet to be recognized. The region 2.1 similarity appears to be a polymerase binding region of Cdc73p. C-terminal regions of Cdc73p are likely to bind to DNA, as do related structures in RAP30 and bacterial sigma factors. This domain may be masked as are these domains in RAP30 and *E. coli* sigma 70 (12,13). Deletion of N-terminal sequences of Cdc73p may be important to observe DNA binding by this factor. RAP30 also binds to TFIIB (20), so Cdc73p may also bind TFIIB.

Bacteria use alternate sigma factors to recognize alternate promoter sequences to co-regulate collections of genes (21). For examples, an alternate sigma factor in *E. coli*, sigma 32, regulates heat shock genes, and multiple specific sigma factors are important in separating developmental stages in *B. subtilis* sporulation. But if Cdc73p is important for recognition of alternate promoters in eukaryotic cells, which promoters does it recognize? RNA polymerase II transcribes several different classes of genes including messenger

Figure 7. Proposed similarities between yeast Cdc73p, human RAP30, and bacterial sigma and delta factors. This figure was prepared based on sequence alignments performed by Shi Min Fang.

Domains of human RAP30



Domains of yeast Cdc73p

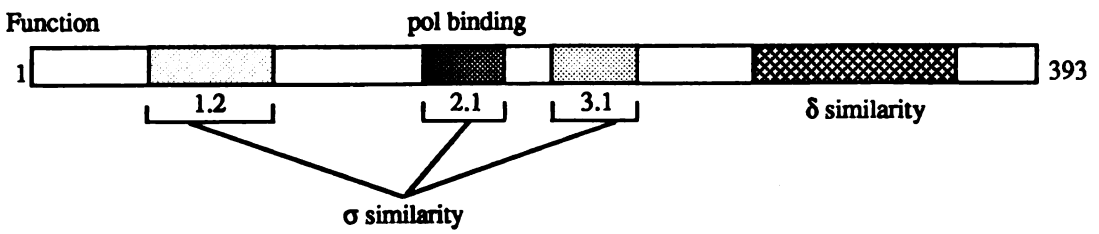


Figure 7

RNA genes that are 3'-polyadenylated, messenger RNA genes that are not 3'-polyadenylated (ie. histone genes in vertebrates), and small nuclear RNAs that have an alternate 5' cap structure and are not 3'-polyadenylated. For some genes alternate 3' end formation is determined by the promoter sequence (22), so polymerase may assemble with a different constellation of factors at different promoters, and these factors may cause different recognition of subsequent processing and termination signals. Therefore, Cdc73p might be involved in transcription of a specific subset of messenger RNA genes, ie. histone genes, or small nuclear RNA genes. If Cdc73p were required for transcription of histone genes, for instance, this might provide clues as to the mechanism of cell cycle arrest in *cdc73* mutants, since histone synthesis is required for the G2/M transition. Alternatively, Cdc73p might be a transcriptional repressor that functions by antagonizing the function of Tfg2p through competition for polymerase binding.

Since temperature sensitive *cdc73* mutants are available (2), RNA levels produced from different genes can be monitored under permissive and non-permissive conditions. Such an approach may permit the identification of genes controlled by Cdc73p. Two hybrid interaction screens may be useful to identify proteins that bind to Cdc73p, for instance to determine whether Tfg1p/Ssu71p or another protein binds to the N-terminal region, as predicted by our model. Affinity tagging Cdc73p in vivo is another approach to identifying additional transcription factors that interact. Using a combination of in vivo and in vitro approaches, therefore, the function of Cdc73p in transcription should be revealed by these studies.

Since TFIIF plays a central function in transcription, an alternate TFIIF might be expected to support a substantially different mechanism. For instance, TFIIE and TFIIH might be dispensible for a Cdc73p dependent initiation mechanism. A mechanism that did not involve TFIIH would likely not require ATP β–γ bond hydrolysis for initiation, since this is a presumed function of the TFIIH helicase activity (23). If TFIIH were not part of the mechanism, the CTD would not be phosphorylated by the CTD kinase that is

another component of TFIIH (24). CTD phosphorylation is thought to regulate promoter entry (25), promoter escape (26, 27), and elongation (28); therefore, a mechanism that did not involve phosphorylation might have less processive transcription and an alternate 3' end formation. For instance, small nuclear RNAs are much shorter transcripts than most mRNA transcripts, so their synthesis would not require as processive a polymerase. A dephosphorylated CTD might associate with an alternate set of processing and termination factors.

Interestingly, some small nuclear RNAs are synthesized by RNA polymerase II and some are synthesized by RNA polymerase III (29). Promoters that are recognized by one or the other polymerase are very similar and both polymerases utilize the same TBP-TAF complex to transcribe these genes (30). Other factor and co-factor requirements for transcription of these genes are less clearly defined. Whether TFIIB, TFIIF, TFIIE, TFIIH, or ATP hydrolysis are required has not been reported. A few straightforward experiments using existing in vitro systems might indicate the possibility of an alternate TFIIF for transcription of these genes.

ACKNOWLEDGEMENT

We thank laboratory members S.-M. Fang for computer analysis, S. Reymez for yeast RNA polymerase II, and L. Lei for immobilized Anc1p. This research was supported by the National Institutes of Health, Michigan State University, The Michigan State University Agricultural Experiment Station.

REFERENCES

1. Reed, S. I., Ferguson, J., Jahng, K. -Y. (1988). Cold Spring Harb.Symp. Quant. Biol. 53, 621-627.

2. Shi, X., and Jaehning, J. A. personal communication.

3. Henry, N. L., Sayre, M. H. and Kornberg, R. D. (1992). J. Biol. Chem. 267, 23388-23392.

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

5. Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991). Mol. Cell. Biol. 11, 1195-1206.

6. Conaway, J. W., and Conaway, R. C. (1990). Science 248, 1550-1553.

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

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

9. Fang, S.- M., and Burton, Z. F. (1995a). Manuscript in preparation.

10. Gribskov, M., and Burgess, R. R. (1986). Nucleic Acids Res. 14, 6745-6763.

11. Fang, S.- M., and Burton, Z. F. (1995b). Manuscrip in preparation.

12. Drombowski, A.J., Walter, W.A., Record, M.T. Jr., Siegele, D.A. and Gross, C.A. (1992) Cell 70, 501-512.

13. Tan, S., Garrett, P. K.,Conaway, R. C., and Conaway, J. W. (1994). Proc. Natl. Acad. Sci. USA. 91, 9808-9812.

14. Corden, J.L. (1990) Trends in Biol. Sci. 15, 383-387.

15. Wade, P. A., Werel, W., Fentzke, R. C., Thompson, N. E., Leykam, J. F., Burgess, R. R., Jaehning, J. A., and Burton, Z. F. (1995). Submit to Proc. Natl. Acad. Sci. USA.

16. Studier, F, W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990). Mehtods Enzymol. 185, 60-89.

17. Wang, B. Q., Kostrub, C. F., Finkelstein, A. and Burton, Z. F. (1993). Protein Expr. Purif. 4, 207-214.

18. Lei, L., and Burton, Z. F. (1995). Manuscript in preparation.

19. Thompson, N. E., Aronson, D. B., and Burgess, R. R. (1990). J. Biol. Chem. 265, 7069-7077.

20. Ha, I., Roberts, S., Maldonado, E., Sun, X., Kim, L. U., Green, M. and Reinberg, D. (1993). Genes Dev. 7, 1021-1032.

21. Gross, C.A., Lonetto, M. and Losick, R. (1992) Transcriptional Regulation, S. L. McKnight and K.R. Yamamoto, eds., Cold Spring Harbor Laboratory Press, pp. 129-178.

22. Hernadez, N. (1992) Transcriptional Regulation, S. L. McKnight and K.R. Yamamoto, eds., Cold Spring Harbor Laboratory Press, pp. 281-313.

23. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulin, W., Hoeijimaker, J. H. J., Chambon, P. and Egly, J.-M. (1993). Science 260, 58-63.

24. Lu, H., Zawel, L., Fisher, L., Egly, J- M., and Reinberg, D. (1992). Nature 358, 641-645.

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

26. Maxon, M. E., Goodrich, J. A. and Tjian, R. (1994). Genes Dev. 8, 515-524.

27. Goodrich, J. A. and Tjian, R. (1994). Cell 77, 145-156.

28. O'Brien, T., Hardin, S. E., Greenleaf, A. L., and Lis, T. J. (1994). Nature 370, 75-77.

29. Lobo, S. M., and Hernandez, N. T. (1994). In: Transcription: Mechanisms and Regulation. edited by R. C. Conaway and J. W. Conaway. (Raven Press, Ltd., New York), pp 127-159.

30. Hernandez, N. (1993). Genes Dev. 7, 1291-1308.