IN VITRO TRANSCRIPTION STUDIES OF ADENOVIRUS

by

Andrew Zachary Fire

A.B. University of California, Berkeley
(1978)

Submitted to the Department of
Biology
in Partial fulfillment of the
requirements of the
degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1983

Signature of Author Signature Redacted

Department of Biology

Certified by Signature Redacted

Thesis Supervisor

Accepted by Signature Redacted

Department Committee on Graduate Students

Archives
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY
JUN 30 1983

LIBRARIES

ABSTRACT

IN VITRO TRANSCRIPTION STUDIES OF ADENOVIRUS

by Andrew Zachary Fire

Submitted to the Department of Biology in June 1983 in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Because of their compact genome and rapid life cycle, adenoviruses provide an excellent model system for studying gene expression in mammalian cells. This thesis presents a series of biochemical experiments directed toward understanding transcription of the adenovirus genome and its regulation during the viral lytic cycle.

A first step was to develop an in vitro system capable of accurately initiating transcription on adenovirus promoters. Such a system was developed in collaboration with J.Manley and M.Gefter. A concentrated and dialyzed whole cell extract was prepared from uninfected HeLa cells. This extract accurately initiates transcription at seven of the nine adenovirus promoter sites. Fine structure mapping of RNA 5' termini was performed in collaboration with C.Baker and E.Ziff. Most adenovirus promoters exhibit heterogeneity in vivo: 5' termini map to specific sites within a 2-7 base region in the DNA sequence. The in vitro system precisely reproduced this microheterogeneity. Further experiments varying free nucleoside triphosphate concentrations demonstrated that these 5' termini resulted from heterogeneous initiation rather than cleavage.

In order to test for the presence of soluble factors involved in regulation, extracts prepared at different stages of infection were assayed for activity on an assortment of viral promoters. Several shifts in promoter activity in vivo during adenovirus infection have been characterized. During a standard adenovirus infection, a product of the "pre-early" E1a region is required for expression of the other early regions. The fact that all early promoters were active in extracts of uninfected cells indicated that the "pre-early" E1a product was not an essential component in promoter recognition. The other major regulatory shift in vivo occurs between early and late promoters, and is concomitant with and dependent on DNA replication. A similar shift could be reproduced in vitro 1) by increasing exogenous DNA concentration added to an uninfected extract, or 2) by using (DNA free) extracts prepared late in infection.

The transcription reaction was studied in more detail using a reconstituted transcription system developed in collaboration with Mark Samuels. A whole cell extract of HeLa cells was resolved through two successive chromatographic steps using an extension of the proceedure of Matsui et. al. (JBC 255, 11992). RNA polymerase II and three of the resulting fractions were necessery and sufficient for transcription of the adenovirus major late and early region IV promoters, The transcriptional activity in each of the HeLa fractions exhibited the following properties: 1) each was sensitive to mild heat treatment, 2) each sedimented as a single peak of activity on sucrose gradients, and 3) each titrated linearly with concentration through a range of at least ten fold.

A multiple step kinetic assay provided information concerning the roles the different components in the reaction. Preincubation of template with a mixture of polymerase and fractions allowed formation of "activated" complexes, defined by the ability to rapidly begin transcription when presented nucleotides. Maximal activation required that all of the components be present in the preincubation. Similar assays were used to define two functional intermediates leading to the activated complex. A complex resistent to interference by excess DNA forms during preincubation of template with just two of the HeLa fractions. Polymerase can associate directly with this intermediate. The third HeLa fraction can act after polymerase has associated with the template

Analysis of the initiation event was greatly facilitated by the ability to prime RNA synthesis using a dinucleotide. A dinucleotide challenge protocol allowed the rate of productive chain initiation to be directly measured. In addition , experiments testing all dinucleotides for specific priming of the major late promoter defined a domain of approximately 9 base pairs around the major late cap site in which initiation with dinucleotides can occur.

Thesis Supervisor: Dr. Phillip A. Sharp, Professor of Biology

TABLE OF CONTENTS

Abstract	2
Dedication	4
Acknowledgements	5
Organization of the Thesis	6
List of Figures and Tables	7
I. Introduction and Summary	10
II. Preparation and Properties of the Whole Cell Extract Transcription System	39
III. Activity of Adenovirus Promoters in Uninfected and Infected Cell Extracts	65
IV. Separation and Characterization of Factors Mediating Accurate Transcription by RNA Polymerase II	83
V. Kinetic Dissection of the Selective Transcription Reaction	93
VI. Use of Dinucleotides to Probe Initiation by RNA Polymerase II	121
Appendix	154
Biographical Note	160

I dedicate the thesis to my family

Acknowledgements

It is a pleasure to thank the many people who have contributed to this thesis and to my graduate education. Indeed I am grateful

to Mark Samuels, whose rigorous and energetic collaboration made this work possible

to Carl Baker and Ed Ziff for their early work and subsequent collaboration

to Malcolm Gefter, who introduced me to enzymes

to Jim Manley, who taught me respect for RNA

to the Sharp Lab postdocs, who have each contributed uniquely to the lab. The critical scientific appraisals of Robert Kingston and cloned DNA segments (and wit) provided by Kathy Berkner have been immesurably useful in shaping this work.

to Barbara Aufiero, Lyn Spencer, Lynne Corboy, Sue Huang, and Mary Esteve, whose technical assistance has made the lab a pleasant and efficient place to work.

to Margarita Siafaca, who translated many an illegible scribble into fine English

to Robert Ely, Andy Hopkins, Rich Calendar, Ziva Reuveny, and Boris Magasanik, who gave me a start in experimental science

to my classmates in the Biology Department who are responsible for much of my education over the last five years

to Frank Laski

to Ihor Lemishka, Connie Cepko, and Dan Donoghue who have shown me what a graduate career could be

to the students (and staff) of course 7.03; I hope they learned as much from the course as I

to my friends outside of Molecular Biology, who have helped keep this all in perspective

Finally, I wish to express my gratitude and respect to my advisor, Professor Sharp, for his time, patience, guidance, support, direction, and enthusiasm.

Organization of the Thesis

The thesis is organized into six chapters. Chapters II-VI each describe one discrete portion of the work. These chapters are self contained papers; as such each includes a separate introduction and discussion of the results. The purpose of chapter 1 is to tie the different parts together as a cohesive unit, and to set the thesis as a whole in the context of previous and current work on transcription. I have attempted in preparing chapter 1 to minimize duplication of introductions and discussions from the other chapters.

List of Figures and Tables

Figure	Page	Title
Figure I-1	15	Transcription Map Of Adenovirus 2
Figure II-1	47	Agarose Gel Analysis of In Vitro Transcription Products
Figure II-2	52	Relative Transcriptional Activities of Two Adenovirus Promoters as a Function of Bulk DNA Concentration
Figure III-1	69	Transcription of E1a and E1b Regions
Figure III-2	70	Transcription of EII, EIII, and EIV Regions
Figure III-3	73	Fingerprints of 250-n runoff transcripts from PEIV
Figure III-4	76	Localization of 5' Termini from the EIV Region
Figure III-5	77	Runoff Mapping of EII
Figure III-6	77	Early Region II Fingerprint
Figure III-7	78	DNA Titration Curves for E1a and the Late Promoter in Uninfected and Infected Extracts.
Figure III-8	79	Mixture of Templates From Different Regions Transcribed in Mock-, Early-, and Late- Infected Extracts
Figure IV-1	85	Scheme for the Resolution of Transcription Factors Contained in Solubilized HeLa Cell Extracts
Figure IV-2	86	Identification of Fractions Required for Transcription
Figure IV-3	87	Short Runoff Analysis
Figure IV-4	88	Dose Response Curves for Transcription Components
Figure IV-5	89	Sucrose Gradient Analyses of Second Column Fractions
Figure IV-6	90	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoretic Analysis of Transcriptionally Active Fractions
Figure IV-7	90	Dependence of Accurate Transcription on Reaction Parameters

Figure V-1	99	Protein Fractions and DNA Template Used In Chapter V
Figure V-2	101	Time Courses of Preincubation, Pulse and Chase
Figure V-3	105	Activation is Specific to the Preincubated Template
Figure V-4	107	Sequential Activation of Two Templates
Figure V-5	109	Factor Requirements During Preincubation
Figure V-6	111	Template Recovery After Incomplete Preincubation
Figure V-7A	113	Inhibition by Poly (dI-dC:dI-dC) Defines a Stable Preinitiation Complex
Figure V-7B	115	Polymerase Associates with Template in the Absence of [CB]
Figure V-8	117	Proposed Scheme for the Transcription Reaction
Figure VI-1	130	The Reconstituted System and Three Stage Protocol
Figure VI-2	132	Analysis of M13-truncated EIV RNA
Figure VI-3	135	T1 Oligonucleotide Analysis of Major Late Promoter RNA
Figure VI-4	140	Dinucleotide Primed Initiations
Figure VI-5	142	RNase U2 Digestions of Dinucleotide-Primed Products
Figure VI-6	144	Dinucleotide Challenge Experiment
Figure VI-7	148	Dinucleotides Used to Prime RNA Synthesis
Figure A-1	156	The 5' End of the Ad2 Late Transcription Unit
Figure A-2	156	Extract and DNA Concentration Optima for Accurate Transcription
Figure A-3	157	α-amanitin sensitivity of transcription from the Ad2 Late Promoter
Figure A-4	157	Fingerprint Analysis of RNA Synthesized in Vitro
Figure A-5	158	Transcription from Other Adenovirus Promoters

Tables

Table	Page	Title
Table II-1	57	Summary of Initial Work With In Vitro Polymerase II Systems
Table III-1	72	DNA Sequences Preceding mRNA CAP Sites
Table III-2	75	Analysis of RNase T1 Oligonucleotides from Ad5 EIV RNA Transcribed In Vitro
Table IV-1	88	Dependence on Individual Fractions
Table IV-2	88	Heat Lability of Individual Fractions
Table IV-3	89	Purification of Second Column Fractions
Table IV-4	90	Dependence on Reaction Components
Table VI-1	138	Secondery Analysis Of 5' Terminal Oligonucleotides

Chapter I

INTRODUCTION AND SUMMARY

General Introduction

DNA serves as template for two fundamental cellular processes: replication and transcription. Replication duplicates the genetic material and allows a precise copy to be passed on to each daughter at cell division. In general, each segment of DNA acts as a substrate for replication exactly once during each cell division cycle. Transcription copies the genetic material into RNA; this process provides input from the genome into the cell's metabolic system. Chemically, the two functions for DNA are similar in that both involve the synthesis of nucleotide polymers complementary to the DNA in sequence. Transcription differs biologically from replication in that different RNA sequences are required by the cell in different amounts. Thus transcription has evolved as a selective process: some sequences are transcribed many times during each cell cycle, some only a few times, and some are probably never transcribed.

Different cell types or cells in differing physiological states have different structural and metabolic requirements. In some cases regulation is mediated post transcriptionally, either through modifications in the function of existing proteins or through differences in mRNA stability, translation or processing. In several cases, however, regulation has been shown to result from differences in transcription.[a] Thus the machinery mediating transcriptional selectivity must be responsive to the developmental and physiological state of the cell.

This thesis, "In Vitro Transcription Studies of Adenovirus", presents a series of experiments directed toward the goal of understanding transcriptional selectivity in mammalian cells. Several different approaches toward this goal are currently being pusued in this laboratory and elsewhere. One useful approach in bacterial systems has been to isolate and study the effect of mutants in the transcriptional machinery. Unfortunately, difficulties inherent in mammalian genetics will probably limit the availability of such mutants in the near future. Another approach involves detailed structural analyses of the DNA template and RNA products in vivo. The object of such studies is to obtain a correlation between template structure (i.e. sequence or nucleoprotein conformation) and the structure or

amount of the complementary RNA products. Such correlations can be strengthened by the analysis of mutated templates. The range of templates available is greatly enhanced by the ability to specifically mutate and re-introduce cloned DNA segments back into cells.

These in vivo approaches have certain limitations. First, manipulation of the system is restricted by constraints inherent in experiments with living cells. Second, the structure and amount of RNA complementary to a given region are functions of processing and stability as well as of transcription; it is not always possible to distinguish among effects of these processes in vivo.

In vitro transcription provides a complementary approach. In a soluble system it is possible to directly manipulate the template, the transcriptional machinery, and the conditions of the reaction; in addition, transcription rates can be directly measured. By itself, the enzymology is impossible to interpret without some indication of whether the relevant in vivo process is carried out similarly. Indeed the strongest attack comes from a combination of in vivo and in vitro approaches. Genetic data and structural analysis of RNA in vivo define selective transcription and provide criteria for reconstructing faithful assays for regulation in vitro. Any models deduced from the enzymology should then be confirmed by testing critical predictions in vivo.

Prokaryotic Models

Transcription in prokaryotic systems has been extensively analyzed (1). E. Coli RNA polymerase was originally purified using a simple assay for template directed synthesis of RNA. This assay demanded no selectivity in the enzyme and led to the purification of a "core" polymerase which did not exhibit the selectivity expected from previous genetic data. Selective transcription could be retained by careful purification of the polymerase. A subunit responsible for selectivity (called sigma) had been lost during the original purification (2).

A general model emerges from studies with the sigma containing "holoenzyme." After mixing template and polymerase, initial [polymerase:template] complexes called "closed" complexes are rapidly formed (3,4). These structures subsequently isomerize to form "open" complexes, defined by the ability to rapidly initiate transcripton when

presented nucleotides (3). Stable open complexes can also be characterized by their resistance to heparin (a DNA analog) (5), by electron microscopy (6) or by protection of specific sequences in the DNA (7). On certain promoters, transcription is observed under conditions which do not permit detection of stable open complexes. It is surmised that the open complexes exist as unstable intermediates in these reactions.

Studies of gene expression in bacteria have shown that a wide variety of regulatory mechanisms are used. In particular, transcriptional regulation has been demonstrated at the levels of polymerase binding (8,9), open complex formation (10), and elongation (anti-termination) (11). Many of the relevant regulatory proteins have been characteized: some interact specifically either with template (8) or with polymerase (12); some interact with both (10,13). These studies have provided a wide variety of paradigms, some of which will undoubtedly be relevant to eukaryotic systems.

Eukaryotic RNA Polymerases

Like E.Coli RNA polymerase, the eukaryotic RNA polymerases were purified using an assay for template directed RNA synthesis that did not demand selectivity (in fact denatured templates were generally used). Three forms of polymerase were isolated, designated I, II, and III (14,15). Conclusions about the physiological roles of these enzymes were made possible by the availability of α-amanitin, a mycotoxin which had been shown to specifically inhibit messenger RNA synthesis (16). The demonstration that purified RNA polymerase II was acutely sensitive to the toxin suggested that this polymerase was responsible for messenger RNA synthesis (17,18). This was confirmed genetically by demonstrating that mutant cell lines resistant to amanitin (for both growth and mRNA synthesis) contained an amanitin resistant RNA polymerase II activity (19).

Polymerase III is sensitive only to very high concentrations of amanitin (20). These concentrations were too high to diffuse into whole cells. However, experiments using permeable nuclei showed that the sensitivity of 5S and tRNA transcription to amanitin paralleled that of the purified RNA polymerase III (20). RNA polymerase I was completely resistant to amanitin, as was transcription of large rRNA

precursors (21). Both RNA polymerase I and the newly synthesized large rRNA precursors had previously been localized in the nucleolar compartment (22,23). These data suggested that RNA polymerase I was responsible for the synthesis of large rRNA precursors, and that RNA polymerase III was responsible for 5S and tRNA synthesis.

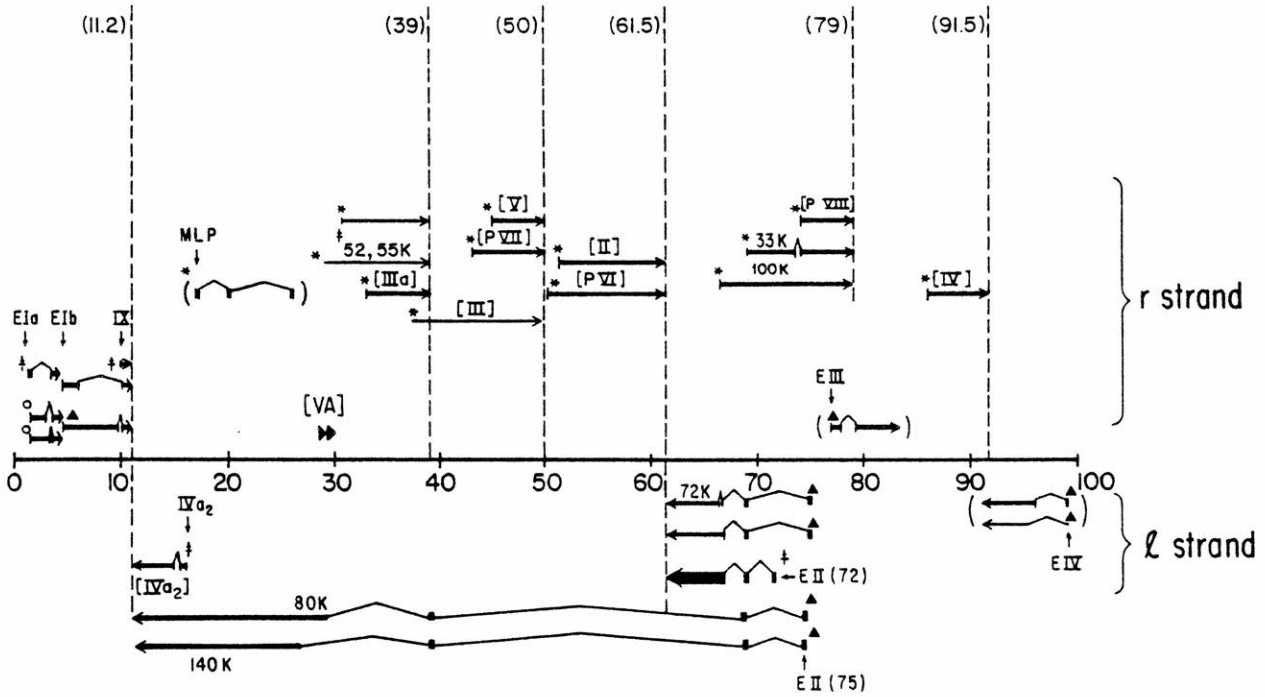
With no other assay than the transcription of denatured templates, it was difficult to conclude anything about the mechanism of transcriptional selectivity. What was needed was a template whose transcriptional properties were well defined in vivo. This should allow analysis of a small number of genes in detail. Since the advent of molecular cloning it has become possible to use cellular genes for this purpose. Ten years ago this was not possible.

Adenovirus as a Model System:

Because of their compact genome and rapid life cycle, adenoviruses have provided an excellent model system for studying gene expression in mammalian cells. The structure and biology of these viruses has been reviewed (24). The features of the viral life cycle important for our studies can all be described in relation to the physical map of the genome shown in figure 1. This map has been compiled over the last ten years in several labs (for reviews see references 25, 26).

Figure I-1 Transcription Map of Adenovirus 2

The structures of the major RNAs expressed from different regions of adenovirus 2 are presented above or below the heavy line marked in map units (1 genome = 100 map units = 35,000bp). Sequences spliced together are joined by a caret symbol. The mRNAs are divided into four groups: pre-early, early, intermediate, and late. These are denoted by the four symbols o, ▲, ‡, and * respectively. By convention, regions of the genome transcribed into early mRNAs are referred to as early regions 1A (0-4.4), 1B (4.4-11.2), II (75-11.2), III (76-84), and IV (92-99). All RNAs from the major late transcription unit (denoted by a *) have a tripartite leader set from 16.5, 19.5, and 26.5 spliced to their 5' termini. These RNAs can be assigned to one of five families based on the site of polyadenylation at 39 (L1), 50 (L2), 61.5 (L3), 79 (L4), or 91.5 (L5).



The virus encodes about thirty messenger RNAs, most of which are believed to encode a single polypeptide. These RNAs are divided into nine transcription units. In addition, two small non-messenger RNAs (denoted VA I and VA II) are encoded from the middle of the genome.

Two aspects of the life cycle are key to the use of adenovirus as a model system. First, the virus uses host systems to express its genes; initial studies with ∝-amanitin demonstrated that the mRNA genes are transcribed by RNA polymerase II and the VA genes by RNA polymerase III (27-29). Second, there is an orderly progression of gene expression during the lytic cycle. Four stages have been defined: pre-early, early, intermediate and late. During each of these stages, a different class of mRNA is most prominent. RNA pulse labeled very early in infection (0.5-2.0hr) hybridizes primarily to early region Ia at the left end of the map (30). One of the products of this region is involved in positive regulation of other early genes (31,32). Between 2 and 8 hours post infection, RNAs from five different "early" transcription units (including EIa) are prominent. Several functions have been assigned to these early regions. A product from early region Ib has been implicated in translational control (33). Three proteins from the EII transcription unit are involved in viral DNA replication: a 140 kd DNA polymerase, a 72K single stranded DNA binding protein, and an 80K protein which acts as a covalent primer for DNA synthesis (34-36). The functions of EIII and EIV products have not been determined.

As DNA replication begins, two intermediate mRNAs are produced (37-40): the resulting proteins form a part of the virion capsid. At least eight other virally encoded proteins are necessery for synthesis of virion particles. These are coded for by five families (each 3' coterminal) of mRNA from the major late transcription unit (25). A tripartite leader sequence spliced to form the 5' terminus is characterisic of all the major late RNAs (41, 42). Transcription from the major late unit is extremely prominent at late times, accounting for as much as 30% of total RNA synthesis in the cell (43). A novel form of mRNA for the 72kd DNA binding protein is also observed late in

infection (38, 40).

Promoters, Cap Sites, and 5' Termini

The sensitivity of the different adenovirus mRNAs to ultraviolet light indicated, on a rough level, that RNAs from each of the transcription units were initiated at a unique site (44, 45). In a series of experiments which are of immeasurable value to this work, Baker, Ziff, and Evans used DNA and RNA sequencing to precisely map the 5' ends of all the mRNA families observed in vivo. In the case of the major late and EIa transcription units, all 5' termini occur at a single residue (46, 47); the other transcription units each exhibit microheterogeneity over a 2-7 base region (47, 48).

It was conceivable that the 5' termini could have represented cleavage sites rather than sites of initiation. This was considered unlikely for several reasons. First 5' termini agreed well with the localization of initiation sites obtained from UV sensitivity and nascent transcript mapping (25). Second, several groups had rigorously demonstrated that nucleotides labeled in the beta position were incorporated into capped 5' termini in permeabilized cells or nuclei (49, 50). This would not be expected if all 5' termini resulted from post-transcriptional cleavage.

Termination and Processing

The 3' structures of a wide variety of mRNAs have been determined. In general, genomic sequences stop at a single base with a tail of about 200 A residues attached at that point (51). Further analysis revealed that a hexanucleotide sequence is conserved some 20 nucleotides upstream of the poly A, with a C invariably lying adjacent to the poly A (52). These sequences are necessery, but not sufficient for generation of the 3' terminus (53). All of the sequenced adenovirus messages carry a similar consensus sequence and polyA structure at their 3' termini, suggesting strongly that the virus makes use of cellular mechanisms to form these termini (54).

It was difficult to determine in vivo whether generation of 3' ends was due to termination of transcription or posttranscriptional cleavage followed by poly A addition. Transcription was observed to extend beyond poly A sites in vivo, suggesting cleavage (45). The demonstration of a precursor-product relationship requires a pulse

chase experiment, however. Technical complexities make *in vivo* pulse chase experiments difficult to interpret.

Toward an in vitro Transcription System

Elucidation of the adenovirus RNA map provided an ideal assay for selective transcription *in vitro*: a system that faithfully reproduces the *in vivo* reaction should yield 5' termini identical to those obtained *in vivo*. Negative results were obtained with the purified RNA polymerases. Polymerase II would not transcribe efficiently from double stranded regions in the template, but rather started at nicks, gaps, and ends with a high efficiency (14). A very low level of transcription apparently initiating from double stranded regions was nonspecific. A similarly negative result was obtained with RNA polymerase III (14).

One obvious approach was to turn to crude systems. A number of studies had previously examined transcription in isolated nuclei (55), although none had demonstrated specific initiation. Nuclei prepared from adenovirus infected cells late in infection provided an excellent system to study RNA metabolism, since adenovirus RNA accounts for a major fraction of RNA produced. Although most synthesis in such a system results from elongation of previously initiated chains, Manley et al were able to demonstrate that nuclei were capable of initiating correctly at the major late cap site (56). This was a step toward developing a soluble transcription system for polymerase II genes. The isolated nuclei also produced RNAs with polyadenylated 3' termini identical to those seen *in vivo* (57). Because the nuclei were freely permeable to nucleotides as well as to inhibitors such as α-amanitin and actinomycin D, it was possible to perform pulse chase experiments (58). Indeed, longer precursors could be chased into shorter polyadenylated RNA species in the presence of amanitin or actinomycin.[b] This showed that 3' end formation could proceed by cleavage of a longer precursor. These studies strongly suggest that in devising assays for *in vitro* transcription by RNA polymerase II, one should not expect production of the *in vivo* 3' terminus.

A transcription system for RNA polymerase III

The first eukaryotic *in vitro* transcription system was developed using adenovirus as a template. It was known that polymerase III was

present at high concentrations in a cytoplasmic extract (called an S100) prepared at a moderate salt concentration (59). The use of moderate salt during extraction allowed many nuclear proteins to leak out. G.J. Wu showed that such an extract would efficiently and specifically synthesize VA RNAs from the viral DNA (60). These RNAs were both initiated and terminated at the proper nucleotide.

Transcription by RNA polymerase III *in vitro* has since been extensively studied. The first surprise came from deletion analyses, which revealed a promoter internal to the gene (61, 62). Similar results were subsequently obtained by microinjection of mutant DNAs into oocytes (63). Selective transcription by RNA polymerase III requires multiple protein components in addition to the purified polymerase (64). One of these components was shown to be specific to the 5S genes (64-66). This factor binds specifically to the internal promoter region of the 5S gene; its binding correlates well with sequences required for transcription (67). Recognition of the promoter is thus independent of the polymerase. Although similar promoter specific components have not been identified for other polymerase III genes, a striking specificty in DNA competition experiments suggests their existence (68, 69).

Two Polymerase II Transcription Systems

The success in developing the RNA polymerase III transcription system immediatly suggested a similar approach for an RNA polymerase II system. The S100 extract contains very little of the cell's RNA polymerase II. Two strategies were used to overcome this. One strategy was to add purified polymerase II back to the S100 extract in the hope that the S100 extract contained factors which would stimulate accurate transcription. Weil et al showed that this S100 + purified RNA polymerase II system could accurately initiate transcription at the adenovirus 2 major late promoter (70). This was the first demonstration of selective transcription *in vitro* by RNA polymerase II.

The second strategy for preparing an *in vitro* transcription system was to use a whole cell extract known to contain endogenous RNA polymerase II. Such an extract was first used by Sugden and Keller as an early step in RNA polymerase purification (71). This "whole cell extract" was capable without any added polymerase of accurate

transcriptional initiation at the adenovirus major late cap site (72).

Preparation and Properties of The Whole Cell Extract System

In order to use such systems it was necessery to have some understanding of their biochemical properties. Chapter 2 describes preparation of the whole cell extract and some of its properties relevant to transcription.

In Vitro Activity of Different Adenovirus Promoters

Given the recognition of the major late promoter, an immediate question was whether other promoters would be accurately transcribed as well. This question was of particular import if in vitro transcription was to be useful in studying adenovirus gene expression. Two techniques were used to map the RNA products observed in vitro.

The first technique was to examine "runoff" transcripts. A recombinant plasmid[c] carrying the putative promoter element was cleaved with a variety of restriction enzymes and transcribed in the whole cell extract. By correlating the lengths of the resulting transcripts with the restriction map, it was possible to localize 5' termini to within about 30 bases. The runoff transcription analysis demonstrated that the WCE system specifically initiated transcription at all five of the early promoters and at one intermediate promoter (pIX) as well as at the major late promoter. Two other promoters (for the IVa2 and late 72K messages) produced no runoff RNA in vitro. These promoters are transcribed in vivo only during the intermediate and late stages of infection.

More detailed 5' end mapping was performed in collaboration with C. Baker and E. Ziff. Their characterization of 5' termini in vivo had made use of RNA fingerprinting techniques. A similar analysis of in vitro transcripts from EII and EIV templates showed that the whole cell extract precisely reproduced the 5' termini that had been observed in vivo at these two promoters. This was particularly instructive in the case of EIV. The 5' termini for EIV mRNAs correspond to seven adjacent residues in the DNA (TTTTTTA); exactly the same pattern of capped termini was observed under the standard conditions in vitro. This provides a striking demonstration that the transcription reaction in the whole cell extract reflects at least some component of the in vivo process.

In addition, faithful reproduction *in vitro* of the microheterogeneity allowed us to directly address the question of whether the multiple 5' termini resulted from initiation or cleavage. This was done by varying concentrations of ATP and UTP in the reaction. At high ratios of ATP to UTP, the A start predominated; at low ratios of ATP to UTP, the U starts predominated. Such a shift would not be expected if the heterogeneous 5' termini arose through cleavage. This indicated that the 5' termini indeed resulted from heterogenous initiation.

Lee and Roeder mapped *in vitro* transcripts from the adenovirus promoters using the S100 + purified RNA polymerase II system, obtaining relative activities similar to those observed in the whole cell extract system (73). To date accurate transcription of over thirty viral and cellular promoters has been demonstrated in the *in vitro* systems. A list of such studies is presented at the end of chapter 2.

Applications of in vitro transcription

The availability of *in vitro* transcription systems active on a wide variety of templates should allow several aspects of RNA metabolism to be defined and studied biochemically.

I. Which template sequences promote accurate initiation of transcription?

II. What is the nature of proteins and other cofactors required to permit selective transcription by polymerase?

III. How do the required factors, polymerase and template interact with each other?

IV. What is the effect of higher order template structure (e.g. supercoiling or chromatin conformation) on transcription?

V. What is the nature of regulatory elements that interact with the transcriptional apparatus?

A number of studies have begun to address these questions using the lysate transcription systems described above. It is of great importance to compare these results both quantitatively and qualitatively to the results of parallel studies *in vivo*. The results of such comparisons define both uses and limitations of the *in vitro* assays. A good correlation indicates that the *in vitro* system provides a faithful assay for the *in vivo* phenomenon. A lack of correlation can

be interpreted in a number of ways:

First, one could propose that the in vitro reaction occurs by a different mechanism than that in vivo. The precise correspondence between initiation sites indicates that the polymerase - factor - template complex responsible for initiation has the same conformation in vivo and in vitro. It is conceivable, however that the complex forms by a non-physiological pathway in vitro.

Second, relevant factors could be absent or inactive in the extract. The inefficiency of the system might support such a hypothesis. Estimates of the fraction of template and polymerase molecules participating in the selective transcription during the reaction range from 0.1%-10%.

Finally, some differences observed between the in vitro and in vivo reactions could reflect solely differences in relative rates for component steps in the reaction. Transcription is a kinetically complex, multi-step reaction involving at minimum a binding step, helix opening, initiation, and elongation. If different reaction steps limit the in vivo and in vitro assays, then the two sets of results would not be expected to correlate.

Sequence Dependences In Vitro and In Vivo:

A myriad of templates and assays have been used for promoter mapping studies in vitro and in vivo. For the purposes of this discussion, I will simplify matters by dividing sequence elements into "close upstream" sequences (-40 to the cap) and "far upstream" sequences (beyond -40). The general consensus from in vitro transcription studies is that the "close upsteam" sequences are both necessery and sufficient to allow selective initiation at the proper site[d,e](74, 75); these sequences include a homology 25 to 30 nucleotides upstream of the cap site called the "TATA" box (76). Specific point mutations in the TATA sequence have been shown to drastically reduce in vitro transcription (77).

In vivo results from several different systems differ in the indicated role for "close upstream" sequences. In some studies, deletion of sequences around the TATA caused heterogeneity in initiation without any apparent affect on the level of transcription (78); in most studies, deletion of TATA sequences significantly reduced

but did not eliminate transcription (79). This discrepency could reflect either differences between promoters or differences between *in vivo* assay systems.

Virtually all *in vivo* deletion studies have detected "far upstream" sequences required for transcription (80, 81). Some of these elements, called "enhancers", can function at different positions upstream or downstream of the gene (82). Although the *in vitro* system does not absolutely require any of the "far upstream" elements, some quantitative dependence on these sequences has been observed (83, 84). In many of these cases, the *in vitro* effect was tightly dependent on the assay conditions and the extract used. Further analysis of these examples may provide considerable information concerning the role of the upstream sequences.

Regulation in vitro?

It would be unreasonable to expect the first *in vitro* systems to precisely reflect all of the regulatory activities in the cell. Indeed the initial demonstration of late promoter function in an uninfected extract was something of a surprise, since transcripts from this promoter are prominent only late in infection (24). A number of cellular promoters that are transcriptionally inactive in HeLa cells are also transcribed in the *in vitro* system (85, 86). Recent *in vivo* data has shown that the late promoter is indeed transcribed during the early stages of infection, albeit at a low level (45, 87, 88), and that some cellular promoters are active in inappropriate hosts when introduced by DNA transfection (89, 90). These data suggest that the *in vivo* regulation may be more complex than originally imagined.

The first demonstration of transcriptional regulation *in vitro* made use of a genetically identified regulatory protein (SV40T) which had been purified using several independent biochemical assays (91-93). *In vivo*, this protein negatively regulates its own promoter (SV40 early) (94). If an SV40 template is prebound with purified SV40T, the early promoter is specifically repressed in a subsequent transcription reaction (95, 96). A similar *in vitro* result (97) has been obtained with high concentrations of purified adenovirus 72K protein, which had previously been shown to specifically repress transcription from the EIV promoter *in vivo* (97).

Strategies developed with previously purified viral proteins cannot be readily extended to other systems, since the genetic and biochemical techniques used in identification and purification do not exist in most cases. The ability to use in vitro transcription assays to identify and purify regulatory factors could provide a much more general approach. Such an approach would be useful in analyzing the molecular basis of regulatory shifts during adenovirus infection.

In particular, it was known that an EIa product was required for transcription of the other early promoters during a standard lytic infection (31, 32). The fact that all of the early promoters were active in the whole cell extract indicated that EIa was not a necessery component in promoter recognition. Given the comparable activity of the EIa promoter and other early promoters in uninfected extracts, it was not surprising that attempts to observe a specific stimulation of these other promoters in early infected extracts were not successful. Indeed extracts prepared from early infected cells gave the same transcriptional pattern as extracts from mock infected cells. It has since been shown that at very high multiplicities of infection or in the presence of various metabolic inhibitors, efficient transcription of the early promoters in vivo does not require EIa activity (99, 100). The above in vitro and in vivo experiments suggested to us that construction of biochemical assays for EIa activity will require more detailed characterization of the in vivo dependence.

More encouraging results were obtained in reconstructing the early to late shift in adenovirus transcription. Two late onset promoters, IVa2 and LII were not transcribed in uninfected extracts. This provides an assay for activity of the components that stimulate these promoters in late infected cells. Unfortunately, the standard transcription extract prepared from late infected cells did not contain such activities.

Comparison of uninfected and late infected extracts did show a striking shift between those early and late promoters that were actively transcribed in vitro. The late to early ratio was five to ten fold higher in late infected extracts. In vivo pulse labeling experiments show a 20 fold shift in relative transcriptional rates (87). The in vivo shift is dependent on and concominant with viral DNA

replication (25, 26). During this period the template concentration is rapidly increasing, with a potential shift in conformation.[f] Addition of bulk DNA to the uninfected extract produced a shifted late to early promoter ratio similar to that seen in the late extract. This indicates that the different promoters are recognized differently in the uninfected extract. Whether this reflected template-specific positive or negative effectors, or simply a different set of reaction rates for the different promoters was difficult to test without a better understanding of the transcription reaction. Indeed the use of the whole cell extract to identify factors distinguishing these promoters was hampered by the fact that any component affecting total "available" DNA concentration will shift the early to late ratio. A critical test for the correspondence of the in vitro and in vivo early to late shifts should be provided by a comparison of sequences involved in the two effects. In vivo assays for stimulation of the late promoter after DNA replication are currently being developed (101, 102).

A similar early to late shift is observed during SV40 replication (24); this shift can also be mimicked in vitro by increasing DNA concentrations (95, 96). Dynan and Tijan (103) and Hansen et al (104) have suggested that this represents the titration of a positive regulatory factor specifically required for the SV40 early promoter.

Tsuda and Suzuki have studied regulation of fibroin transcription in the silk gland of Bombyx Mori. Fibroin is synthesized in the posterior part of silk gland (105). Comparison of fibroin promoter activity in extracts of posterior and middle silk gland reveals no tissue specific enhancement (106). However, comparison of extracts from either silk gland with a HeLa extract showed enhancement of fibroin transcription relative to a control promoter. This enhancement apparently depended on upstream sequences. Further data from this system should prove interesting.

Conformation of the Template

Chromatin structures in transcriptionally active and inactive regions differ by a number of criteria (see reference 107 for a review). It has been difficult to determine which, if any of these differences has a causal effect on transcription. In vitro

transcription of chromatin templates may eventually provide the crucial tests. Current systems, however, are probably not sufficient for such experiments. Any template added to the whole cell extract becomes rapidly complexed with protein. The structure formed is not chromatin. The low efficiency of the system in terms of templates transcribed (<10%) adds further uncertainty to the structure of active templates.

It has been proposed that superhelical density may play a role in transcription in vivo. This has been difficult to test using current in vitro systems because of significant amounts of topoisomerase activities.

Towards a Reconstituted Transcription System

The whole cell extract and S100 systems have provided a useful beginning in adressing a number of specific questions in transcription. In order to make any further progress it will be necessary to have a better understanding of the basic transcription reaction. This will come primarily with better defined and more efficient transcription system. To this end, a number of groups have begun to fractionate the transcription components. The first such fractionation was obtained by Matsui et al, who chromatographed the HeLa S100 extracts to yield four mutually dependent fractions, each of which must be mixed with the purified polymerase to observe production of correctly initiated transcripts (108).

Mark Samuels and I have extended the proceedure of Matsui et al. We resolve the whole cell extract through two succesive chromatographic steps. Three of the resulting fractions and purified RNA polymerase II are necessary and sufficient for selective transcription of the adenovirus major late promoter. Activities in the three required fractions exhibit the following properties:

1)Each is sensitive to mild heat treatment

2)Each sediments as a peak of activity on sucrose gradients

and 3)Each titrates linearly through a range of at least tenfold

The latter result is particularly useful, since it defines linear assays for each of the fractions.

Using HeLa cells as a source it was difficult to obtain sufficient material to further purify each of these components. It will therefore be advantageous to extract the transcriptional factors from more

abundant (and cheaper) tissue sources. The linear assays constructed from HeLa factors should be very helpful in identifying and purifying factors from tissue extracts.

Are Factors Required In Vivo?

The fact that a protein factor stimulates a reaction in vitro is far from an indication that the same protein has this role in vivo. It is perhaps instructive that two of the four factors originally identified by Matsui et. al. (108) have since been shown by that group to have no direct role in the transcription reaction[g]: one acts as an RNase inhibitor in the in vitro reaction, and the other acts to supress nonspecific background by binding tightly to nicks and ends in DNA (109, 110). Apparently neither fraction is required with "clean" preparations of template and fractions (109, 110). Similar reconstitutions of multi-component systems in bacteria made use of extensive genetics to confirm, wherever possible, the roles of any factors identified. The dearth of genetics in mammalian cells makes this difficult. Some drug resistant mutants have been identified which may prove useful. In particular, polymerase II purified from α-amanitin resistant cell lines (19) functions and is α-amanitin resistant for specific transcription (109). Although not a surprising result (since we already knew that polymerase II was involved in vivo) this type of experiment may prove useful in demonstrating physiological roles for other components which can be identified by drug resistance.

A second approach to this problem is to use lytic viruses which shut off host transcription (poliovirus, reovirus, VSV, etc). Poliovirus shuts off polymerase II transcription very rapidly upon infection (see ref 111 for a review). By preparing extracts from infected cells at a series of points after infection, Crawford et al show that this shutoff can be mimicked in vitro (112). The time course of the shutoff in vitro correlates well with that observed in vivo. Transcription in the infected extracts can be restored by adding back one of the partially purified HeLa transcription factors, suggesting that inactivation of this factor is responsible for the shutoff in vivo. Although the initial evidence for involvement of a specific factor is at best a correlation from such studies, a detailed understanding of the shutoff mechanism could provide methods for

studying the role of a factor *in vivo*.

Kinetic Analysis

Using the reconstituted transcription system consisting of RNA polymerase II supplemented with partially purified HeLa cell transcription factors, we have begun to kinetically dissect the transcription reaction. These experiments were all based on a three stage preincubation - pulse - chase transcription protocol. Variations in this protocol provide assays for several different component steps in the reaction.

A. Stable Complex Formation

The transcription reaction is normally sensitive to high concentrations of DNA. Preincubation with just two of the HeLa fractions (which we call [AB] and [DB]) allows the preincubated template to overcome this block. These experiments define a stable functional complex that can form on the template in the absence of polymerase. Analogous results were obtained by Davison et al (113). Stable template complexes formed during transcription of 5S genes by RNA polymerase III have been shown to be competent for multiple rounds of transcription (114). It has been speculated that formation of such stable complexes might allow DNA to remain transcriptionally active for long periods *in vivo*.

B. Polymerase Binding

Polymerase will form a functional association with the stable complex described above. Significantly, this polymerase association event does not require the third HeLa fraction (which we call [CB]). This fraction is, however, absolutely required for transcription.

C. Activation

The dependence on the preincubation phase in the three stage protocol defines an "activation" assay. Maximal incorporation during the pulse requires that each of the three HeLa fractions and polymerase be present during the preincubation with template. These data suggest that polymerase and each of the fractions participate in forming a template associated "activated" complex capable of rapidly initiating when presented nucleotides.

D. Initiation

Priming by dinucleotides at the adenovirus major late and EIV

promoters has been used to probe the initiation reaction. A "dinucleotide challenge" protocol allows the rate of productive initiation to be directly measured. In addition, experiments testing all 16 possible dinucleotides define a region of nine consecutive bases centered around the late promoter cap site in which initiation with dinucleotides can occur. E. Coli RNA polymerase has been shown to initiate from an open complex in which about 10 bases of the template (one helix turn) are unwound (115). The dinucleotide priming data suggests that a similar "open" complex might exist as an intermediate in accurate transcription by RNA polymerase II.

E. Elongation

The preincubation and pulse produce a relatively synchronous population of labeled nascent chains. Time courses of chase can provide information about the rate and efficiency of elongation. It is intriguing that elongation rates observed for specific transcription in the whole cell extract are comparable to those calculated by Kadesch and Chamberlin (116) for nonspecific transcription by purified RNA polymerase II.

Prospects

The primary goal at this point should be to understand the basic transcription reaction on one or a few well defined templates. This will undoubtedly require the development of efficient transcription systems with purified factors. The kinetic assays described above should facilitate this analysis.

An understanding of this reaction will almost certainly entail the definition of interactions with the TATA sequences. If the current reports of upstream sequence dependences *in vitro* can be developed as reproducible assays, then enzymological approaches should succeed in clarifying the role of these sequences in the reaction. A number of questions are not likely to be amenable to biochemical analysis without further *in vivo* definition. These include the EIa effects in adenovirus, effects of "enhancer" sequences, and regulatory mechanisms dependent on chromosomal position. If chromatin structure is crucial for the latter activities, then a combination of chromatin reconstitution and *in vitro* transcription approaches may be advantageous.

Conclusion

A biochemical assay is a valuable tool in attempting to forge order from chaos. The major contributions of this thesis and other recent transcription studies are biochemical assays for factors, mechanistic steps, and regulatory phenomena. These assays have begun to yield information concerning the mechanism of transcription.

Footnotes

[a]Perhaps the best documented example of tissue specific gene regulation at the level of transcription is reported by Derman et. al. for the case of four cDNA clones of liver specific RNAs (117)

[b]Two of the three 3' termini examined were formed post-transcriptionally. A third 3' terminus at 38.5 m.u. could only form co-transcriptionally. This 3' terminus may thus form by a novel mechanism

[c]These plasmids were kind gifts of K. Berkner, G. Chu, J. Manley, and F. Laski

[d]The large number of promoter mapping studies that have recently been reported make a complete list of citations unreasonable. I have tried to cite one or two exemplary cases for each point.

[e]SV40 provides a striking exception to this. In that case TATA sequences are apparently required neither in vivo nor in vitro (78, 104, 118). This may relate to the involvement of an extra factor in SV40 transcription (103).

[f]Early in infection adenovirus DNA is organized into a chromatin structure; late in infection the DNA is not organized into chromatin, but rather is complexed with two adenovirus encoded proteins (see reference 24 for a detailed discussion)

[g]I do not wish to give a negative impression of the Matsui paper. In my opinion, it is amoung the most important papers that has been published in the field.

References

1. Losick, R., and Chamberlin, M. eds, RNA Polymerase (1976) Cold Spring Harbor Press, N.Y.
2. Burgess, R., Travers, A., Dunn, J., and Bautz, E. (1969) Factor stimulating transcription by RNA polymerase. Nature 221, 43-46
3. Chamberlin, M. (1974) The selectivity of transcription. Annu. Rev. Biochem. 43, 721-772
4. Hinkle, D., and Chamberlin, M. (1971) The role of sigma subunit in site selection by RNA polymerase. Cold Spring Harbor Symposia On Quantitative Biology 35, 65-72
5. Zillig, W., Zechel, K., Rabussay, D., Schachner, M., Sethi, V., Palm, P., Heil, A., and Seifert, W. (1971) On the role of different subunits of DNA dependent RNA polymerase from E. coli in the transcription process. Cold Spring Harbor Symposia On Quantitative Biology 35, 47-58
6. Williams, R. (1977) Use of polylysine for adsorbtion of nucleic acids and enzymes to electron microscopic specimen films. Proc. Nat. Acad. Sci. 74, 2311-2315
7. Sibenlist, U., Simson, R., and Gilbert, W. (1980) E. coli RNA polymerase interacts homologously with two different promoters. Cell 20, 269-281
8. Nakanishi, S., Adhya, S., Gottesman, M., and Pastan, I. (1973) Studies on the mechanism of action of the GAL repressor. J. Biol. Chem. 248, 5937-5942
9. Maurer, R., Maniatis, T., and Ptashne, M. (1974) Promoters are in the operators in phage lambda. Nature 249, 221-223
10. Hawley, D., and McClure, W. (1983) The effect of a lambda repressor mutation on the activation of transcription initiation from the lambda P_{rm} promoter. Cell 32, 327-333
11. Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. Nature 289, 751-758
12. Losick, R., and Pero, J. (1981) Cascades of sigma factors. Cell 25, 582-585
13. Hochschild, A., Irwin, N., and Ptashne, M. (1983) Repressor structure and the mechanism of positive control. Cell 32, 319-325
14. Roeder, R. (1976) Eukaryotic nuclear RNA polymerases; in "RNA Polymerase" (Losick, R. and Chambrerlaim, M. eds)pp 285-329, Cold Spring Harbor Press, N.Y.
15. Roeder, R., and Rutter, W. (1969) Multiple forms of DNA dependent RNA polymerase in eukaryotic organisms. Nature 224, 234-237
16. Zylber, E., and Penman, S. (1971) Products of RNA polymerases in HeLa cell nuclei. Proc. Nat. Acad. Sci. 68, 2861-2865
17. Lindell, T., Weinberg, F., Morris, P., Roeder, R., and Rutter, W. (1970) Specific inhibition of nuclear RNA polymerase II by amanitin. Science 170, 447-449
18. Kedinger, C., Gniazdowski, M., Mandel, J., Gissinger, F., and Chambon, P. (1970) α-amanitin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. Biochem. Biophys. Reas. Comm. 38, 165-171
19. Chan, V., Whitmore, G., and Siminovich, L. (1972) Mammalian cells with altered forms of RNA polymerase II. Proc. Nat. Acad. Sci. 69, 3119-3123
20. Weinman, R., and Roeder, R. (1974) Role of DNA dependent RNA

polymerase III in the transcription of the tRNA and 5S RNA genes. Proc. Nat. Acad. Sci. 71, 1790-1794
21. Portnoff, L. and Roeder, R. Unpublished Results (cited in 14)
22. Roeder, R. and Rutter, W. (1970) Specific nucleolar and nucleoplasmic RNA polymerases. Proc. Nat. Acad. Sci. 65, 675-682
23. Reeder, R., and Roeder, R. (1972) Ribosomal RNA synthesis in isolated nuclei. J. Mol. Biol. 67, 433-441
24. Tooze, J. (1980) DNA Tumor Viruses. Cold Spring Harbor Press, N.Y.
25. Ziff, E. (1980) Transcription and RNA processing by the DNA tumor viruses. Nature 287, 491-499
26. Flint, S. J. (1982) Expression of adenoviral genetic information in productively infected cells. Bioch. Biophys. Acta. 651, 175-208
27. Price, R. and Penman, S. (1972) A distinct RNA polymerase activity, synthesizing 5.5S, 5S and 4S RNA in nuclei from adenovirus 2-infected HeLa cells. J. Mol. Biol. 70, 435-450
28. Price, R., and Penman, S. (1972) Transcription of the adenovirus genome by an α-amanitine sensitive ribonucleic acid polymerase in HeLa cells. J. Virol. 9, 621-626
29. Weinman, R., Raskas, H., and Roeder, R. (1974) Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc. Nat. Acad. Sci. 71, 3426-3430
30. Nevins, J., Ginsberg, H., Blanchard, J.-M., Wilson, M., and Darnell, J. (1979) Regulation of the primary expression of the early adenovirus transcription units. J. Virol. 32, 727-733
31. Berk, A., Lee, F., Harrison, T., Williams, J., and Sharp, P.A. (1979) Pre-early adenovirus 5 gene function regulates expression of other early viral genes. Cell 17, 935-944
32. Jones, N., and Shenk, T. (1979) An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc. Nat. Acad. Sci. 76, 3665-3669
33. Persson, H., Monstein, H.-J., Akusjarvi, G., and Phillipson, L. (1981) Adenovirus early gene products may control viral mRNA accumulation and translation in vivo. Cell 23, 485-496
34. Stillman, B., Tamanoi, F., and Matthews, M. (1982) Purification of an adenovirus-coded DNA polymerase that is required for initiation of DNA replication. Cell 31, 613-623
35. Lichy, J., Enomoto, T., Field, J., Guggenheimer, R., Ikeda, J.-E., Nagata, K., Horwitz, M., and Hurwitz, J. (1982) Isolation of proteins involved in the replication of adenovirus DNA. Cold Spring Harbor Symposia on Quantitative Biology 47, in press
36. Stillman, B., Lewis, J., Chow, L., Mathews, M., and Smart, J. (1981) Identification of the gene and mRNA for the adenovirus terminal protein precursor. Cell 23, 497-508
37. Pettersson, U., and Mathews, M. (1977) The gene and messenger RNA for adenovirus polypeptide IX. Cell 12, 741-750
38. Galos, R., Williams, J., Binger, M. and Flint, J. (1979) Location of additional early gene sequences in the adenovirus chromosome. Cell 17, 945-956
39. Chow, L., Broker, T., and Lewis, J. (1979) Complex splicing patterns of RNA from the early regions of adenovirus 2 messenger RNA. J. Mol. Biol. 134, 265-303
40. Chow, L., Lewis, J., and Broker, T. (1980) RNA transcription and splicing at early times after adenovirus 2 infection. Cold Spring

Harbor Symposia on Quantitative Biology 44, 401-414
41. Berget, S., Moore, C., and Sharp, P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Nat. Acad. Sci. 74, 3171-3175
42. Chow, L., Gelinas, R., Broker, T., and Roberts, R. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNAs. Cell 12, 1-8
43. Lucas, J., and Ginsberg, H. (1971) Synthesis of virus-specific ribonucleic acid in KB cells infected with type 2 adenovirus. J. Virol 8, 203-213
44. Berk, A., and Sharp, P.A. (1977) Ultraviolet mapping of the adenovirus-2 early promoters. Cell 12, 45-55
45. Nevins, J., and Wilson, M. (1981) Regulation of adenovirus-2 gene expression at the level of transcription termination and RNA processing. Nature 290, 113-118
46. Ziff, E., and Evans, R. (1978) Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus 2 major late transcription unit. Cell 15, 1463-1475
47. Baker, C., and Ziff, E. (1981) Promoters and heterogeneous 5' termini of the messenger RNAs of adenovirus serotype 2. J. Mol. Biol. 149, 189-221
48. Baker, C., Herisse, J., Courtois, G., Galibert, F., and Ziff, E. (1979) Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end. Cell 18, 569-580
49. Contreras, R., and Fiers, W. (1981) Initiation of transcription by RNA polymerase II in permeable, SV40-infected or noninfected CV1 cells; evidence for multiple promoters of SV40 late transcription. Nucleic Acids Research 9, 215-239
50. Gidoni, D., Kahana, C., Canaani, D., and Groner, Y. (1981) Specific initiation of transcription of simian virus 40 early and late genes occurs at various capped nucleotides including cytidine. Proc. Nat. Acad. Sci. 78, 2174-2178
51. Brawerman, G. (1976) Characteristics and significance of the polyadenylate sequence in mammalian messenger RNA. Progress in Nucleic Acid Research and Molecular Biology 17, 117-148
52. Proudfoot, N., and Brownlee, G. (1976) 3' noncoding region sequences in eukaryotic messenger RNA. Nature 263, 211-214
53. Fitzgerald, M., and Shenk, T. (1981) The sequence 5'AAUAAA3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24, 251-260
54. Fraser, N., Baker, C., Moore, M., and Ziff, E. (1982) Poly (A) sites of adenovirus serotype 2 transcription units. J. Mol. Biol. 155, 207-233
55. Mory, Y., and Gefter, M. (1977) Synthesis of messenger RNA-like molecules in isolated myeloma nuclei. Nucleic Acids Research 4, 1739-1757
56. Manley, J., Sharp, P.A., and Gefter, M. (1979) RNA synthesis in isolated nuclei: in vitro initiation of adenovirus 2 major late mRNA precursor. Proc. Nat. Acad. Sci. 76, 160-164
57. Manley, J., Sharp, P.A., and Gefter, M. (1979) RNA synthesis in isolated nuclei: identification and comparison of adenovirus 2 encoded transcripts synthesized in vitro and in vivo. J. Mol. Biol. 135, 171-197

58. Manley, J., Sharp, P.A., and Gefter, M. (1982) RNA synthesis in isolated nuclei: processing of adenovirus serotype 2 late messenger RNA precursors. J. Mol. Biol. 159 581-599
59. Seifart, K., Benecke, B., and Juhasz, P. (1972) Multiple RNA polymerase species from rat liver tissue; possible existence of a cytoplasmic enzyme. Arch Bioch Biophys 151, 519-532
60. Wu G.-J. (1978) Adenovirus DNA-directed transcription of 5.5S RNA in vitro. Proc. Nat. Acad. Sci. 75, 2175-2179
61. Sakonju, S., Bogenhagen, D., and Brown, D. (1980) A control region in the center of the 5S RNA gene directs specific initiation. I: the 5' border of the region. Cell 19, 13-25
62. Bogenhagen, D., Sakonju S., and Brown, D. (1980) A control region in the center of the 5S RNA gene directs specific initiation. II: the 3' border of the region. Cell 19, 27-35
63. Hofstetter, H., Kressman, A., and Birnstiel, M. (1981) A split promoter for a eukaryotic tRNA gene. Cell 24, 573-585
64. Segall, J., Matsui, T., and Roeder, R. (1980) Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J. Biol. Chem. 255, 11986-11991
65. Engelke, D., Ng, S.-Y., Shastry, B., and Roeder, R. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. (1980) Cell 19, 717-728
66. Pelham, H. and Brown, D. (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc. Nat. Acad. Sci. 77, 4170-4174
67. Sakonju, S., Brown, D., Engelke, D., Ng, S.-Y., Shastry, B., and Roeder, R. (1981) The binding of a transcription factor to deletion mutants of a 5S ribosomal RNA gene. Cell 23, 665-669
68. Fowlkes, D., and Shenk, T. (1980) Transcriptional control regions in the adenovirus VA1 gene. Cell 22, 405-413
69. Sharp, S., Dingerman, T., Shaack, J., Defranco, D., and Soll, D. (1983) Transcription of eukaryotic tRNA genes in vitro. I Analysis of control regions using a competition assay. J. Biol. Chem. 258, 2440-2446
70. Weil, P., Luse, D., Segall, J., and Roeder, R. (1979) Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18, 469-484
71. Sugden, B., and Keller, W. (1973) Mammalian deoxyribonucleic acid dependent RNA polynerases. J. Biol. Chem. 248, 3777-3788
72. Manley, J., Fire, A., Cano, A., Sharp, P.A., and Gefter, M. (1980) DNA dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Nat. Acad. Sci. 77, 3855-3859
73. Lee, D., and Roeder, R. (1981) Transcription of adenovirus 2 genes in a cell-free system: apparent heterogeneity of initiation of some promoters. Mol. Cell. Biol. 1, 635-651
74. Hu, S.-L., and Manley, J. (1981) DNA sequences required for initiation of transcription in vitro from the major late promoter of adenovirus 2. Proc. Nat. Acad. Sci. 78, 820-824
75. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corse, P., Kedinger, C., and Chambon, P. (1980) Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406-1414
76. Goldberg, M. (1979) Ph.D. Thesis, Stanford University
77. Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule,

R., and Chambon, P. (1980) Specific in vitro transcription of the conalbumin gene is drastically decreased by single point mutation in TATA box homology sequence. Proc. Nat. Acad. Sci. 77, 7024-7028

78. Ghosh, P., Lebowitz, P., Frisque, J., and Gluzman, Y. (1981) Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Nat. Acad. Sci. 78, 100-104
79. Grosschedl, R., and Birnstiel, M. (1980) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Nat. Acad. Sci. 77, 1432-1436
80. Grosschedl, R., and Birnstiel, M. (1980) Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription in vivo. Proc. Nat. Acad. Sci. 77, 7102-7106
81. Gruss, P., Dhar, R., and Khory, G. (1981) Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Nat. Acad. Sci. 78, 943-947
82. Fromm, M., and Berg, P. (1982) Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. Journal of Applied and Molecular Genetics 1, 457-481
83. Grosschedl, R., and Birnstiel, M. (1982) Delimitation of far upstream sequences required for maximal in vitro transcription of a H2A histone gene. Proc. Nat. Acad. Sci. 79, 297-301
84. Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M., and Chambon, P. (1982) Sequences upstream of the T-A-T-A box are required in vivo and in vitro for efficient transcription from the adenovirus serotype 2 major late promoter. Proc. Nat. Acad. Sci. 79, 7132-7136
85. Proudfoot, N., Shander, M., Manley, J., Gefter, M., and Maniatis, T. (1980) Structure and in vitro transcription of human globin genes. Science 209, 1329-1336
86. Luse, D., and Roeder, R. (1980) Accurate transcription initiation on a purified mouse β globin DNA fragment in a cell free system. Cell 20, 691-699
87. Shaw, A., and Ziff, E. (1980) Transcription from the adenovirus-2 major late promoter yield a single family of 3' coterminal mRNAs during early infection and five families at late times. Cell 22, 905-916
88. Lewis, J., and Mathews, M. (1980) Control of adenovirus early gene expression: a class of immediate early products. Cell 21, 303-313
89. Mantei, N., Boll, W., and Weissman, C. (1979) Rabbit β-globin mRNA production in mouse L cells transformed with cloned rabbit B-globin chromasomal DNA. Nature 281, 40-46
90. Dierks, P., van Ooyen, A., Mantei, N., and Weissman, C. (1981) DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5' terminus. Proc. Nat. Acad. Sci. 78, 1411
91. Tjian, R. (1978) The binding site on SV40 DNA for a T antigen related protein. Cell 13, 165-179
92. Jessel, D., Landau, T., Hudson, J., Lalor, T., Tenen, D., and Livingston, D. (1976) Identification of regions of the SV40 genome which contain preferred SV40 T antigen binding sites. Cell 8,

535-545
93. Reed, S., Ferguson, J., Davis, R., and Stark, G. (1975) T antigen binds to SV40 DNA at the origin of replication. Proc. Nat. Acad. Sci. 72, 1605-1609
94. Tegtmeyer, P., Schwartz, M., Collins, J., and Rundell, K. (1975) Regulation of tumor antigen synthesis by simian virus 40 gene A. J. Virol. 16, 168-178
95. Rio, D., Robbins, A., Myers, R., and Tjian, R. (1980) Regulation of simian virus 40 transcription in vitro by a purified tumor antigen. Proc. Nat. Acad. Sci. 77, 5706-5710
96. Hansen, U., Tenen, D., Livingston, D., and Sharp, P.A. (1981) T antigen repression of SV40 early transcription from two promoters. Cell 27, 603612
97. Handa, H., Kingston, R., and Sharp, P.A. (1983) Inhibition of adenovirus early region IV transcription in vitro by a purified viral DNA binding protein. Nature 302, 545-547
98. Nevins, J., and Winkler, J. (1980) Regulation of early adenovirus transcription: a protein product of early region 2 specifically represses early region 4 transcription. Proc. Nat. Acad. Sci. 77, 1893-1897
99. Gaynor, R., Tsukamoto, A., Montell, C., and Berk, A. (1982) Enhanced expression of adenovirus transforming proteins. J. Virol. 44, 276-285
100. Nevins, J. (1981) Mechanism of activation of early viral transcription by the adenovirus E1a product. Cell 26, 213-220
101. Kaufman, R. and Sharp, P.A. (1982) Construction of a modular dihydrofolate reductase cDNA gene: analysis of signals utilized for efficient expression. Mol. Cell. Biol. 2, 1304-1319
102. Berkner, K., and Sharp, P.A., Manucript in preparation
103. Dynan, W., and Tjian, R. (1983) Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell 32, 669-680
104. Hansen, U., Fromm, M., and Sharp, P.A., manuscript in preparation
105. Suzuki, Y. (1976) Differentiation of the silk gland. in "Results and Problems in Cell Differentiation" 8, (W. Beerman, ed.) Springer Verlag, Berlin and Heidelberg. 1-44
106. Tsuda, M., and Suziki, Y. (1981) Faithful transcription initiation in a homologous cell free system reveals an enhancing effect of 5' flanking sequence far upstream. Cell 27, 175-182
107. Felsenfeld, G. (1978) Chromatin. Nature 271, 115-122
108. Matsui, T., Segall, J., Weil, P., and Roeder, R. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. 255, 11992-11996
109. Samuels, M., Fire, A., and Sharp, P.A. Unpublished Results
110. Slattery, E., Dignam, D., Matsui, T., and Roeder, R. (1983) Purification and analysis of a factor which supresses nick induced transcription by RNA polymerase II and its identity with poly (ADP-ribose) polymerase. J. Biol. Chem. 258, 5955-5959
111. Lucas-Lenard, J. (1979) Inhibition of cellular protein synthesis after virus infection. in "The Molecular Biology of PicoRNAviruses" (R.Berkoff ed) Plenum, N.Y. pp.73-99
112. Crawford, N., Fire, A., Samuels, M., Sharp, P.A., and Baltimore, D. (1981) Inhibition of transcription factor activity by poliovirus. Cell 27, 555-561

113. Davison, B., Egly, J.-M., Mulvivill, E., and Chambon, P. (1983) Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature 301, 680-686

114. Bogenhagen, D., Wormington, W., and Brown, D. (1982) Stable transcription complexes of Xenopus 5S RNA genes: a means to maintain the differentiated state. Cell 28, 413-421

115. Wang, J., Jacobsen, J., and Saucier, J.-M. (1977) Physiochemical studies on interaction between DNA and RNA polymerase. Unwinding of the DNA helix by Esherichia coli RNA polymerase. Nucleic Acids Research 4, 1225-1241

116. Kadesch, T., and Chamberlin, M. (1982) Studies of in vitro transcription by calf thymus RNA polymerase II using a novel duplex DNA template. J. Biol. Chem. 257, 5286-5295

117. Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M., and Darnell, J. (1981) Transcriptional control in the production of liver-specific mRNAs. Cell 23, 731-739.

118. Myers, R., Rio, D., Robbins, A., and Tjian, R. (1981) SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell 25, 373-384

Chapter II

PREPARATION AND PROPERTIES OF THE WHOLE CELL EXTRACT TRANSCRIPTION SYSTEM*

*The original description of Whole Cell Extract mediated accurate transcription is presented as appendix 1: Manley, J.L., Fire, A. , Cano, A. , Sharp, P.A. , and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. 77, 3855-3859

This chapter is derived from an article that I wrote jointly with J.L. Manley, M. Samuels, and P. Sharp. (Methods in Enzymology 101, 568-582)

I. Introduction

Three classes of nuclear DNA-dependent RNA polymerases have been identified in eukaryotic cells (1). RNA polymerase I catalyzes the synthesis of rRNA precursors, RNA polymerase II transcribes primarily the genes which give rise to mRNA, while RNA polymerase III transcription results in the production of tRNAs, 5S RNA, and several other RNAs of unknown function. It has been clear for many years that in order to study the mechanisms of transcription as well as to identify the factors and nucleotide sequences which control and regulate gene expression, cell free systems that accurately and specifically transcribe exogenously added DNA are required. Early attempts at achieving this aim, which utilized purified RNA polymerases, were unsuccessful. In the last several years, however, in vitro systems have been developed in which accurate transcription by all three types of RNA polymerases can be obtained. Two basic approaches have been successful. In one, purified RNA polymerase is supplemented with cell extracts which contain factors required for accurate transcription (2). This method has been used primarily for RNA polymerase II mediated transcription and is described in reference 56. The other approach is to prepare concentrated cell lysates which contain not only the factors required for transcription, but also sufficient amounts of RNA polymerase so that addition of purified enzyme is not required. For RNA polymerase I (3) and III (4) such systems can be simply prepared from cytoplasmic extracts, because sufficient amounts of these polymerases and their required factors leak out of the nucleus at isotonic salt concentrations. RNA polymerase II, on the other hand remains almost entirely within the nucleus. Thus, to obtain extracts containing this

activity, a whole cell extract must be prepared (5). We describe here preparation and properties of such an extract, which contains all of the factors and enzymatic activities necessary for accurate and specific transcription, not only by RNA polymerase II, but also by RNA polymerases I and III.

Most experiments to date have utilized extracts prepared from human cells which grow in suspension culture (HeLa or KB). Such extracts show quite broad species specificities for RNA polymerase II and III. Polymerase III genes from virtually all higher eukaryotes that have been tested are accurately transcribed in HeLa lysates. Whole cell extracts do not seem able to accurately transcribe yeast polymerase II genes, but have been shown to be capable of transcribing Drosophila (6) and chicken (7) polymerase II genes as well as many such genes from higher eukaryotes and their viruses.

Synthesis of mature RNA molecules, of course, requires additional enzymes and factors other than those needed to bring about accurate transcriptional initiation (e.g. processing enzymes). Soluble HeLa lysates appear to contain virtually all of the enzymes required for tRNA processing, including the splicing enzymes (8,9). Although work with RNA polymerase I systems is just beginning, soluble extracts appear to contain at least one processing enzyme (10,11). RNA polymerase II transcripts synthesized in vitro are efficiently capped and methylated at their 5' ends (2,5). Although one report in the literature claims that an extract efficiently spliced RNA (12), we have not observed in a variety of experiments either splicing or creation of polyadenylated 3' termini in the cell-free system described here.

II. Methods

A. Preparation of Extract

Extracts are prepared by modification of a procedure originally described by Sugden and Keller (13) who used the method as a first step in RNA polymerase purification. We have used HeLa cells almost exclusively. These cells are easy to obtain in large quantities, and the resultant extracts are relatively free of nuclease activity at 30°C. Lysates with transcriptional activity have been prepared from a few other cell lines; most other cell lines and tissues, however, have yielded extracts without detectable levels of transcription or with high levels of nuclease. Cells are grown in suspension culture in Eagle's minimal essential medium supplemented with 5% horse serum to a density of $4\text{-}8\text{x}10^5$ cells/ml. The cell density appears not to be crucial, although we have recently obtained slightly more active extracts with cells harvested at the lower end of the range indicated.

The following operations are carried out at 0-4°C.

1) Cells are harvested by centrifugation and washed two times with phosphate buffered saline.

2) The volume of the resultant cell pellet is determined, and the cells resuspended in four packed-cell volumes (PCV) of 0.01M Tris-HCl (pH 7.9), 0.001M EDTA and 0.005M DTT ($6\text{x}10^8$ cells yield approximately 2 ml of packed cells). At this point, the cells should visibly swell.

3) After 20 minutes, the cells are lysed by homogenization in a Dounce homogenizer with eight strokes using a "B" pestle.

4) Four PCV of 0.05M Tris-HCl (pH 7.9), 0.01M $MgCl_2$, 0,002M DTT, 25% sucrose and 50% glycerol are added, and the suspension is gently mixed. With continued gentle stirring, one PCV of saturated $(NH_4)_2\ SO_4$

is added dropwise. After this addition, the highly viscous lysate is gently stirred for an additional 30 minutes. Stirring must be very gentle to prevent shearing of the DNA, which would interfere with its removal in the next step. Nuclear lysis can be detected by increased viscosity after approximately half the $(NH_4)_2\ SO_4$ has been added. Occasionally, lysates appear clumpy and only slightly viscous, rather than extremely viscous and uniform as usually observed. We have obtained active extracts from both types of lysates, although more reproducibly from the latter.

5) The extract is carefully poured into polycarbonate tubes and centrifuged at 45,000 rpm in a SW 50.2 rotor for three hours.

6) The supernatant is decanted so as not to disrupt the pellet (the last one or two ml is left behind) and protein and nucleic acid precipitated by addition of solid $(NH_4)_2\ SO_4$ (0.33 gm/ml of solution). After the $(NH_4)_2\ SO_4$ is dissolved, 1N NaOH [0.1 ml/10 gm solid $(NH_4)_2\ SO_4$] is added and the suspension stirred for an additional 30 minutes.

7) The precipitate is collected by centrifugation at 15,000xg for 20 minutes (the supernatant should be completely drained off), and resuspended with five percent the volume of the high-speed supernatant with 0.025M Hepes (adjusted to pH 7.9 with NaOH), 0.1M KCl, 0.012M $MgCl_2$, 0.5 mM EDTA, 2 mM DTT, and 17% glycerol.

8) The suspension is dialyzed against two changes of 50-100 volumes each of the resuspension buffer for a total of 8-12 hours. The volume of the solution increases 30-50% during dialysis. The conductivity of a 1:1000 dilution of dialyzed extract into distilled H_2O ($23^{\circ}C$) should be 12-14 umho.

9) The dialysate is centrifuged at 10,000xg for ten minutes to

remove insoluble material. The supernatant is divided into small aliquots (0.2-0.5 ml), quick frozen in liquid nitrogen or powdered dry ice and stored at -80°C. Extract can be thawed and quick frozen several times without loss of activity and retains full activity at -80°C for at least a year.

10) Lysates contain between 15-30 mg/ml of protein and up to 2 mg/ml of nucleic acid. We routinely start with 1-50 liters of cells. One liter of cells should yield about 2 ml of WCE, or enough for 100-400 assays. More concentrated extracts are desirable because with these the same optimal protein concentration can be obtained in reaction mixtures with a smaller volume of lysate. In this manner, the salt concentration in the in vitro reaction mixture can be lowered (high salt severely inhibits transcription; see below). Attempts to obtain more concentrated extracts by resuspending the pellet in a smaller volume after precipitation, or by tying the dialysis bag tightly (to reduce expansion during dialysis), have not been reproducibly successful due to increased protein precipitation during dialysis. Likewise, dialysis against buffer containing lower salt concentrations results in less active lysates, again as a result of increased protein precipitation.

B. The Transcription Reaction:

Reactions can be done in volumes of a few ul or more . Analytical reactions are conveniently performed in 20 ul. A typical reaction mix might contain the following: 30-60% whole cell extract in its dialysis buffer, 0.2-1.5 ug of template DNA (see below for effects of DNA and

extract concentrations), 50 uM ATP, 50 uM GTP, 50 uM CTP, 5 uM UTP, 5 mM creatine phosphate, and 10 uCi of α^{32}P-UTP [commercial preparations of aqueous nucleotides can be obtained at a high enough concentration (~10 mCi/ml) and a sufficient specific activity (>200 ci/mMol) to be added directly to the transcription]. After incubation for 30-120 minutes the reactions can be extracted directly or placed at -80°C for up to a week before extraction.

C. Extraction of RNA and Resolution of Products by Gel Electrophoresis

To terminate transcription, 200 ul of stop buffer (7M urea, 100 mM LiCl, 0.5% sodium dodecyl sulfate [SDS], 10 mM EDTA, 250 ug/ml tRNA, 10 mM Tris Cl [pH 7.9]) and 300 ul of PCIA (phenol-chloroform-isoamyl alchohol; 1:1:0.05; water saturated and buffered with 20 mM Tris [pH 7.9]) are added, the tubes blended in a vortex mixer and centrifuged at 12,000xg for 15 minutes. The aqueous phase (discarding interface) is extracted once more with PCIA and once with chloroform and then pooled with 200 ul of 1.0M NH_4 acetate and precipitated with 900 ul of ethanol. The pellet is washed with ethanol and resuspended in 20 ul of 10 mM Na_2HPO_4 (pH 6.8)-1 mM EDTA; to this is added 50 ul of 1.4M deionized glyoxal-70% dimethyl sulfoxide-10 mM Na_2HPO_4 (pH 6.8)-1 mM EDTA-0.04% bromophenol blue. After 1 hr at 50°C, 25 ul of the sample is loaded on 1.4% agarose gels in 10 mM PO,-1 mM EDTA (14). This extraction procedure removes most of the free nucleotides from the RNA preparation.

For some techniques, larger reaction volumes are necessary. The above extraction protocol can be scaled up with modifications as follows: After removal of the first aqueous phase, three reaction

volumes of stop buffer are added to the first organic phase, and the mixture is again homogenized and spun at 12,000xg (for 1 minute). The organic phase is removed and an equal volume of chloroform added. After brief homogenization and centrifugation, the aqueous phase can be easily removed. The two aqueous phases are then pooled and reextracted once with PCIA, and twice with chloroform. Any precipitate at the interface of these extractions should be discarded. After the first ethanol precipitation, the pellet is resuspended in 200 ul of 0.2% SDS and 1 mM EDTA. An equal volume of 2M NH_4 acetate is added and nucleic acid reprecipitated with ethanol. The pellet is washed with ethanol and can be resuspended in the buffer of choice.

For analysis by hybridization and S1 nuclease digestion (15) it is important to first remove the template DNA. The pellet is resuspended in 0.3M NaAcetate [pH 5.2] and reprecipitated and washed with ethanol. The dried pellet is resuspended in 100 ul of 10 mM Tris HCl [pH 7.5], and 100 mM NaCl. RNase free DNase (treated with iodoacetate; 16) to 50 ug/ml, and $MgCl_2$ to 10 mM are added. After 5 minutes at 37°C, 100 ul of 10 mM EDTA, 0.2% SDS, 150 mM NaCl is added, and the mixture extracted with PCIA and chloroform. The final aqueous is reprecipitated with 0.25 ml of ethanol as above and the pellet resuspended in 60 ul of 0.2% sarkosyl-1 mM EDTA [pH 8.0] and stored at -20°C.

III. Sizing and mapping in vitro RNA:

In general, RNA polymerase II does not terminate transcription in vitro. However, distinct length RNA products can be generated by the "run-off" assay. This method uses, as template, DNA molecules which have been cleaved by a restriction enzyme that cuts downstream from a

Figure 1. Analysis of RNA products by denaturation with glyoxal and agarose gel electrophoresis. Recombinant plasmids containing the adenovirus major late promoter were cleaved with a variety of restriction enzymes and used as templates for in vitro transcription. In lanes A-E cleavage should generate runoff transcripts of 0.97, 3.22, 3.90, 4.18, and 1.75 Kb respectively. The migration of the RNA bands correlated well with migration of glyoxal treated DNA restriction fragments of known legnth. The high molecular weight RNA (>6000n.) in lane B results from an incomplete restriction digest in preparing this template.



putative transcription start site. RNA polymerases which transcribe this DNA will stop or fall off when they reach the end of the DNA. If a substantial number of enzymes initiate transcription at the same site, then a population of molecules of a discrete size will be produced; such a population will migrate as a band on gel electrophoresis. DNA segments that have been cleaved by different restriction enzymes are used as templates in separate reaction mixtures, the transcription start site can be deduced by comparison of the sizes of the RNAs produced. This technique has been widely used for promoter mapping with in vitro transcription systems.

An example of the technique is shown in Figure 1. The RNAs were transcribed from recombinant plasmids containing the adenovirus late promoter and various segments of the long (30 kb) late transcription unit. Several points are exemplified by this experiment. The in vitro system is capable of synthesizing very long RNAs, up to 7 to 8 kb, and hence contains little nuclease activity. Also, the glyoxal method of analyzing RNA is sensitive over a wide range of sizes. Plots of log molecular weight vs. migration are linear for transcripts from 0.2 kb to over 5 kb.

Analysis of runoff transcripts is a simple, sensitive and accurate method for determining the structure of in vitro synthesized RNA. However, it does have some limitations. The WCE contains relatively high levels of nucleic acid. Since most of this is 18S and 28S rRNA, it is impossible to load more than 25-50% of the RNA obtained from a 20 ul reaction mix onto a standard size gel slot (6 mm x 3 mm) without producing severe overloading of the gel in the regions occupied by these RNAs. An additional problem is that specific transcripts produced by

very weak promoters can sometimes be obscured by nonspecific initiation or termination, or by exogenous nucleic acids labeled by end labeling activity in the extract (particularly rRNA and its breakdown products). This latter activity is amanitin and actinomycin D insensitive and can thus be distinguished from de novo RNA synthesis. Use of radioactive GTP as tracer produces the least end labeling. Use of CTP or ATP produces high levels of tRNA labeling by enzymes exchanging the 3' terminal CCA. A third limitation of analyzing runoff RNAs by denaturation and agarose gel electrophoresis is that end points of transcripts can only be mapped to within ~20 nucleotides at best. To map the 5' ends of RNAs more precisely, short runoff transcripts can be analyzed on polyacrylamide sequencing gels.

The above analysis can be extended, and some of the problems circumvented, by using several variants on the technique of hybridization and S1 nuclease digestion. By using labeled RNA and a nonradioactive DNA probe, the problem of spurious RNA labeling can be eliminated, since RNA not complementary to the DNA probe is destroyed by the nuclease. Use of nonradioactive RNA and a DNA probe 5' end-labeled 50-200 nucleotides downstream from the promoter allows resolution of ±1 nucleotide. The structure of the 5' end of in vitro synthesized RNA can also be studied by classical RNA fingerprinting techniques (2,5).

IV. Transcriptional Activity

Extracts made from different cell preparations can vary in activity over a five to ten fold range, with about two in three extracts exhibiting activity within two fold of the observed maximum. Extracts should be compared for their activity using a runoff assay from a standard

polymerase II promoter, such as the major late promoter of adenovirus 2. With optimal DNA and extract concentrations, a good extract (20 ul) will yield 10^6 dpm, or 20 ng of a 2200 nucleotide runoff transcript from the Ad2 late promoter, in one hour (^{32}P-UTP at 100 Ci/mMol). This represents the synthesis of one RNA molecule per 10 DNA template molecules present. However, the extract may actually be utilizing a smaller fraction of templates with multiple rounds of initiation per active template.

A) DNA and Extract Concentrations:

Titrations both of DNA and of extract yield nonlinear responses. At a constant extract concentration, measuring runoff transcription as a function of DNA concentration yields (i) a threshold DNA concentration below which no transcription occurs, and (ii) an inhibitory effect of high DNA concentration (5). The requirement for a minimal DNA concentration is nonspecific, i.e. by using a concentration of a promoter specific DNA which is below the threshold, carrier DNA such as pBR322 or E. coli DNA can be added to stimulate specific transcription. The duplex alternating copolymers poly(dIC:dIC) and poly(dAT:dAT) will also act as carrier DNA, thereby demonstrating a total lack of sequence specificity in the bulk DNA requirement (17). A further advantage of these copolymers as carrier DNA is that the transcribed RNA products of the carrier poly(dIC:dIC) and poly(dAT:dAT) contain only two nucleotides. Thus, poly(dIC:dIC) carrier in a reaction containing ^{32}P-UTP yields no radioactive background. The key aspect of bulk DNA dependence is that at a fixed total DNA concentration, the molar yield of transcripts per promoter is constant and independent of the source of

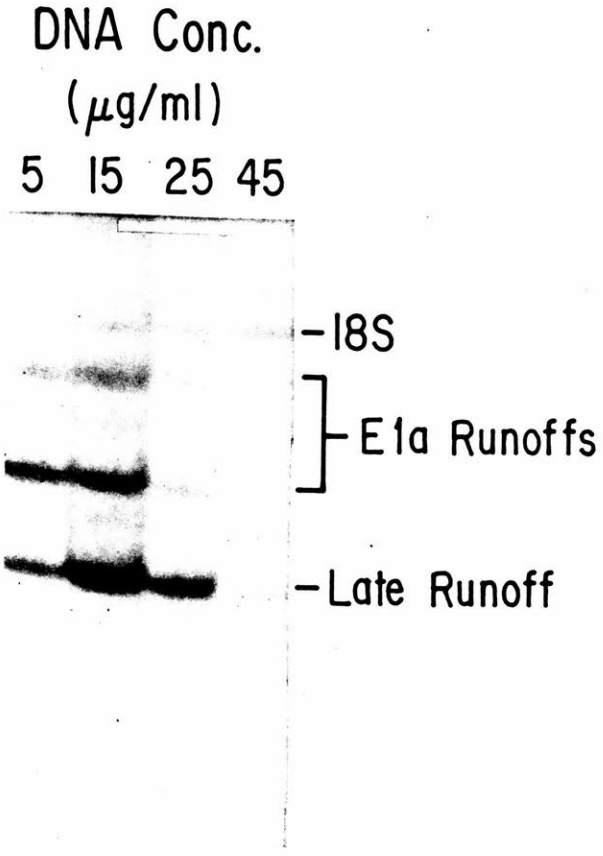
carrier. In general, specific competition between promoter containing fragments is not observed.

A critical dependence of transcription upon extract concentration is also observed (5). In fact, DNA concentration dependence and extract protein concentration dependence are not independent (18). Specific transcription can be obtained in a range of 4 to 18 mg/ml of extract protein. At low extract concentration the DNA optima tend to be much lower (in the range of 10 ug/ml). There is still a bulk DNA dependence, but it is less steep and the threshold concentrations are lower. At a high extract concentration the DNA titration becomes sharper, and the threshold becomes higher. Under such conditions it is often necessary to use 60 ug/ml of DNA in order to see any transcription. Thus, for each new extract it is necessary to do careful DNA and extract titrations, to determine optimal conditions.

For a given promoter, very short runoff transcripts (<300 n) have a higher optimum DNA concentration than longer runoff transcripts (18). This effect can be taken into account by measuring the synthesis of different length runoff products from the same promoter. No length dependence has been observed with runoff products between 400 and 4000 nucleotides.

To further complicate matters, the ratio of activity from two promoters can vary as much as 20 fold over a range of DNA and extract concentration (17,18,19,20). An example of this is shown in Figure 2 where the relative activities of an early and a late Ad2 promoter are compared at increasing total DNA concentrations. The ratio of these activities varies 10 fold at different DNA concentrations in uninfected extracts. Comparison of promoter strengths in different extracts must thus be cautiously

Figure 2. Relative transcriptional activities of two adenovirus promoters as a function of bulk DNA concentration. Each reaction contained 4 ug/ml of a plasmid containing the adenovirus Ela promoter cleaved to give a 1220 n runoff and 1 ug/ml of a plasmid containing the late promoter cleaved to give a 974 n runoff. Bulk DNA concentration was increased by addition of poly (dIdC:dIdC). The transcription products were resolved on glyoxal gels as described.



controlled and interpreted, a crucial point in assaying for regulatory phenomena.

B) Reaction Conditions

One unusual feature of the WCE is the temperature dependence of the reaction. Transcription is routinely done at 30°C, where the in vitro synthesized RNA product is stable for 8 hours. Increasing the temperature to 37°C greatly enhances the rate of RNA degradation; RNA made at 30°C is degraded within 10 minutes after shifting to 37°C. Transcription assayed at 23°C yields the expected Arrhenius effect.

Specific transcription in the WCE is highly sensitive to ionic strength. Concentrations of KCl and NaCl above 60 mM significantly inhibit the reaction; concentrations in the 30-40 mM range are optimal. Reactions can also be performed in 15-30 mM $(NH_4)_2SO_4$. The divalent cations Zn++ and Mn++ inhibit transcription and 0.5 mM EDTA is added to control their effect and the effect of other heavy metal contaminants. Reactions are done at pH 7.9, which is optimal for purified RNA polymerase II (1).

Even after extensive dialysis, most WCEs seem to contain a free pool of 1 uM nucleotides (18). The extract also contains creatine kinase (CPK) and other kinases and phosphatases so that the β and γ phosphates in nucleotidess are labile (21). For example, γ^{32}P-ATP will rapidly exchange label with other triphosphates. Label at the α position of the nucleotide triphosphates does not exchange in the WCE, thus allowing RNA to be uniquely labeled with ^{32}P in the α position of each triphosphate. Addition of 5 mM creatine phosphate to the reaction mix insures charging of the triphosphates and allows reduction in

triphosphate concentrations, thus permitting use of higher specific activities (21). Extracts from some cell lines tested (e.g. L cells) lack CPK activity and the enzyme must be added exogenously to maintain nucleotide concentrations (22). One must also recall that $(NH_4)_2SO_4$ inhibits creatine kinase. In the presence of creatine phosphate, concentrations of UTP, CTP and GTP as low as 5 uM saturate specific transcription; higher concentrations (up to 500 uM) do not inhibit specific transcription (21). Because of endogenous pools, the transcription reaction is not fully dependent on addition of these three nucleotides (18). A higher concentration of ATP is required for optimal activity (50 uM); ATP concentrations above 500 uM inhibit the reaction (21).

The dialyzed extract also contains sufficient S-adenosyl methionine (SAM) to methylate the 5' ends of the *in vitro* transcripts (5). Internal methylation has not been studied, however. Addition of exogenous SAM does not affect the reaction (18,23).

C) Time Course:

The rate of elongation in the WCE is approximately 300 nucleotides/min (23). The rate *in vivo* is 10 fold higher, but one must recall the difference of 7°C in temperature between the two. After DNA template, nucleotides, and extract are mixed, there is about a 5 min lag before specific transcription commences. The lag cannot be eliminated by preincubation of extract alone or with nucleotides, but is eliminated by preincubation of extract together with DNA (in the absence of nucleotides) (24). This suggests that the lag represents the time required for assembly on the DNA of factors required for initiation.

The rate of accumulations of runoff transcripts is approximately linear for over an hour after the initial lag period (5).

V. Some Other Properties of the WCE

The preparation procedure for the WCE was originally designed for solubilization of RNA polymerase II from mammalian cells. A standard 20 ul reaction mix typically contains 2-3 units of RNA polymerase II (5). Under optimal conditions at most one in ten polymerase II molecules gives rise to a specific transcript in a one hour reaction. Supplementation with excess purified polymerase has no significant effect on the WCE (22). Endogenous RNA polymerase II in the WCE is inhibited by α-amanitin at 0.5 ug/ml; addition of a purified mutant enzyme resistant to α-amanitin reconstitutes specific transcription (6,25).

RNA polymerase II preferentially initiates transcription at the termini of DNA fragments and at internal nicks (26). The enzyme is also capable of end labeling DNA fragments with α^{32}P-NTPs to yield full length labeled molecules, which are resistant to RNase digestion (25). These reactions are each sensitive to α-amanitin (0.5 ug/ml), and are suppressed *in vitro* by the addition of a 110,000 dalton ADP-ribosyltransferase which may blockade nicks and ends (27). This 110K protein is present in large quantities in the WCE (up to 0.1% of total WCE protein).

The WCE should contain most of the soluble proteins in the cell. Most of these are of no concern; however, some can interfere with interesting experiments. Most extracts have high levels of topoisomerase type I and II activities as well as DNA ligase. Thus DNA

topologies can change rapidly in the reaction mix, preventing, for instance, studies of supercoiled DNA. The extract also contains RNA polymerases I and III (13). Their contribution to the background pattern can be assessed with α-amanitin. Most template DNAs do not contain promoters for these enzymes and their contribution to background incorporation is small. Some genomic clones contain dispersed / repetitive elements, which often contain polymerase III genes.

VI. Partial Fractionation of the WCE

A number of inhibitory activities can be removed by fractionation on phosphocellulose, yielding a more efficient transcription extract (24,25,28). After dialysis to remove Mg^{++} and dilution to 40 mM KCl, a WCE is chromatographed on phosphocellulose (Whatman P11) yielding a breakthrough fraction, and two higher salt washes (0.35 and 1.0 M KCl). Reconstitution of the breakthrough and the dialyzed 1.0 M wash with purified RNA polymerase II in optimal ratios yields a mixture capable of specifically transcribing DNA at 10 times the efficiency of the original WCE. Tsai et al. have used a similar protocol to remove inhibitors from an extract of hen oviduct (29).

VII Summary of Initial Work with In Vitro Polymerase II Systems

Since the first demonstration of accurate transcription by RNA polymerase II in a soluble S-100 system and development of the whole cell extract proceedure, a large number of investigators have used these systems. A partial listing of such studies is shown in table I.

Table I

In vitro transcription studies using a whole cell extract

Template	Comments	References
Adenovirus	almost all *in vivo* promoters are recognized *in vitro*	2,5,7,12,18,20
	detailed 5' terminal analysis, dependence on upstream sequences	18,20,30,31,32,33
	inactivation of transcription in WCE's of poliovirus infected cells	34
	changes in transcriptional pattern in WCE's of adenovirus infected cells	18
Globins	α- and β-globin genes are recognized *in vitro*, dependence on upstream sequences	35,36,37,38
	mutant α- and β-thalassemia globin genes are transcribed *in vitro*	36,39,40,41
SV40	early and late promoters are recognized *in vitro*	19,21
	detailed 5' terminal analysis of RNA from early promoter	17,42,43
	in vitro inhibition of transcription by T antigen	17,19,44
	cell-free translation of *in vitro* synthesized RNA	45

Conalbumin and ovalbumin	promoters are recognized *in vitro*, dependence on upstream sequences, effects of altering TATA box transcription in a homologous system	7,30,46,47 29
Type C retroviruses	promoters in the long terminal repeat (LTR) of several RNA tumor viruses are recognized *in vitro*, dependence on upstream sequences	48,49,50
Fibroin	promoter is recognized *in vitro* by WCE's of HeLa cells and of silk worm glands, dependence on upstream sequence in both extracts	51,52
Herpes Simplex Virus	early promoters are recognized *in vitro* in uninfected cell WCE's	53
Histone H2A	promoter is recognized *in vitro*, dependence on upstream sequences using linear or circular DNA template	54
Adeno-associated virus	identification of a new promoter, detailed 5' terminal analysis	55

Acknowledgements

The work of R. Jove, H. Handa, U. Hansen, C. Cepko, N. Crawford, and A. Cano in helping to characterize the whole cell extract, and the advice and guidance of M. Gefter are gratefully acknowledged.

1. R.G. Roeder, in "RNA Polymerase" (R. Losick and M. Chamberlin, eds.), p. 285. Cold Spring Harbor Lab., Cold Spring Harbor, New York, 1976.

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

3. I. Grummt, Proc. Natl. Acad. Sci. USA 78, 727 (1981).

4. G.J. Wu, Proc. Natl. Acad. Sci. USA 75, 2175 (1978).

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

6. R. Morimoto, unpublished observations.

7. B. Wasylyk, C. Kedinger, J. Corden, O. Brison and P. Chambon, Nature (London) 285, 366 (1980).

8. D.N. Standring, A. Venegas, and W.J. Rutter, Proc. Natl. Acad. Sci., USA 78, 5963 (1981).

9. F. Laski, A. Fire, U. Rajbhandary, and P.A. Sharp, unpublished observations.

10. I. Grummt, E. Roth and M.R. Paule, Nature (London), in press.

11. K.G. Miller and B. Sollner-Webb, Cell 27, 165 (1981).

12. B. Weingartner and W. Keller, Proc. Natl. Acad. Sci. USA 78, 4092 (1981).

13. B. Sugden and W. Keller, J. Biol. Chem. 248, 3777 (1973).

14. G.K. McMaster and G.C. Carmichael, Proc. Natl. Acad. Sci. USA 74, 4835 (1977)

15. P.A. Sharp, A.J. Berk, and S.M. Berget, this series, Vol. 65, p. 750.

16. S.B. Zimmerman and G. Sandeen, Anal. Biochem. 14, 269 (1966).

17. U. Hansen, D.J. Tenen, D.M. Livingston, and P.A. Sharp, Cell 27, 603 (1981).

18. A. Fire, C.C. Baker, J.L. Manley, E.B. Ziff, and P.A. Sharp, J. Virol. 40, 703 (1981).

19. D. Rio, A. Robbins, R. Myers, and R. Tjian, Proc. Natl. Acad. Sci. USA 77, 5706 (1980).

20. D.C. Lee and R.G. Roeder, Mol. Cell. Biol. 1, 635 (1981).

21. H. Handa, R.J. Kaufman, J.L. Manley, M.L. Gefter, and P.A. Sharp, J. Biol. Chem. 256, 478 (1981).

22. N. Crawford, unpublished results

23. R. Jove, unpublished results

24. A. Fire, unpublished results

25. M. Samuels, unpublished results

26. M.K. Lewis and R.R. Burgess, J. Biol. Chem. 255, 4928 (1980).

27. D. Dingman and R.G. Roeder, unpublished results.

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

29. S. Tsai, M.-J. Tsai, L. Kops, P. Minghetti, and B. O'Malley, J. Biol. Chem.256, 13055.

30. J. Corden, B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon. Science 209, 1406 (1980).

31. S.-L. Hu and J.L. Manley. Proc. Natl. Acad. Sci. USA 78, 820 (1981).

32. O. Hagenbuchle and U. Schibler. Proc. Natl. Acad. Sci. USA 78, 2283 (1981).

33. D.J. Mathis, R. Elkaim, C. Kedinger, P. Sassone-Corsi, and P. Chambon. Proc. Natl. Acad. Sci. USA 78, 7383 (1981).

34. N. Crawford, A. Fire, M. Samuels, P.A. Sharp, and D. Baltimore. Cell 27, 555 (1981).

35. D.S. Luse and R.G. Roeder. Cell 20, 691 (1980).

36. N.J. Proudfoot, M.H.M. Shander, J.L. Manley, M.L. Gefter, and T. Maniatis. Science 209, 1329 (1980).

37. C.A. Talkington, Y. Nishioka, and P. Leder. Proc. Natl. Acad. Sci. USA 77, 7132 (1980).

38. G.C. Grosveld, C.K. Shewmaker, P. Jat, and R.A. Flavell. Cell 25, 215 (1981).

39. R.A. Spritz, P. Jagadeeswaran, P.V. Choudary, P.A. Biro, J.T. Elder, J.K. DeRiel, J.L. Manley, M.L. Gefter, B.G. Forget, and S.M. Weissman. Proc. Natl. Acad. Sci. USA 78, 2455 (1981).

40. S.H. Orkin, S.C. Goff, and R.L. Hechtman. Proc. Natl. Acad. Sci. USA 78, 5041 (1981).

41. S.H. Orkin and S.C. Goff. J. Biol. Chem. 756, 9782 (1981).

42. D.J. Mathis and P. Chambon. Nature (London) 290, 310 (1981).

43. P. Lebowitz and P.K. Ghosh. J. Virol. 41, 449 (1982).

44. R.M. Myers, D.C. Rio, A.K. Robbins, and R. Tjian. Cell 25, 373 (1981).

45. C.L. Cepko, U. Hansen, H. Handa, and P.A. Sharp. Cell Biology 1, 919 (1981).

46. S. Tsai, M.-J. Tsai, and B.W. O'Malley. Proc. Natl. Acad. Sci. USA 78, 879 (1981).

47. B. Wasylyk and P. Chambon. Nuc. Acids Res. 9, 1813 (1981).

48. T. Yamamoto, B. deCrombugghe, and I. Pastan. Cell 22, 787 (1980).

49. M.C. Ostrowski, D. Benard, and G.L. Hager. Proc. Natl. Acad. Sci. USA 78, 4485 (1981).

50. L.A. Fuhrman, C. Van Beveren, and I.M. Verma. Proc. Natl. Acad. Sci. USA 78, 5411 (1981).

51. Y. Tsujimoto, J. Hirose, M. Tsuda, and Y. Suzuki. Proc. Natl. Acad. Sci. USA 78, 4838 (1981).

52. M. Tsuda and Y. Suzuki. Cell 27, 175 (1981).

53. R.J. Frink, K.G. Draper, and E.K. Wagner. Proc. Natl. Acad. Sci. USA 78, 6139 (1981).

54. R. Grosschedl and M.L. Birnstiel. Proc. Natl. Acad. Sci. USA 79, 297 (1982).

55. M.R. Green and R.G. Roeder. Cell 22, 231 (1980).

56. J.D. Dignam, P.L. Martin, B.S. Shastry, and R.G. Roeder. Methods Enzymol 101, 582 (1983)

Chapter III

ACTIVITY OF ADENOVIRUS PROMOTERS IN UNINFECTED AND INFECTED CELL EXTRACTS*

*The RNA fingerprinting experiments in this paper were performed in collaboration with C. C. Baker and E. B. Ziff

JOURNAL OF VIROLOGY, Dec. 1981, p. 703-719
0022-538X/81/120703-17$02.00/0

Vol. 40, No. 3

In Vitro Transcription of Adenovirus

ANDREW FIRE,[1] CARL C. BAKER,[2] JAMES L. MANLEY,[3] EDWARD B. ZIFF,[2] AND PHILLIP A. SHARP[1]*

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[1]; Rockefeller University[2] and Department of Biology, Columbia University,[3] New York, New York 10021

Received 5 June 1981/Accepted 17 July 1981

A series of recombinants of adenovirus DNA fragments and pBR322 was used to test the transcriptional activity of the nine known adenovirus promoters in a cell-free extract. Specific initiation was seen at all five early promoters as well as at the major late promoter and at the intermediate promoter for polypeptide IX. The system failed to recognize the two other adenovirus promoters, which were prominent in vivo only at intermediate and late stages in infection. Microheterogeneity of 5′ termini at several adenovirus promoters, previously shown in vivo, was reproduced in the in vitro reaction and indeed appeared to result from heterogeneous initiation rather than 5′ processing. To test for the presence of soluble factors involved in regulation of mRNA synthesis, the activity of extracts prepared from early and late stages of infection was compared on an assortment of viral promoter sites. Although mock and early extracts showed identical transcription patterns, extracts prepared from late stages gave 5- to 10-fold relative enhancement of the late and polypeptide IX promoters as compared with early promoters.

Initiation of transcription by RNA polymerase II is the first step in synthesis of mRNA in eucaryotic cells. Weil et al. (40) have demonstrated that this process can be duplicated in a soluble in vitro reaction by mixing purified RNA polymerase II and an S100 cytoplasmic extract. Manley et al. (24) found that an extract of whole HeLa cells will also specifically initiate transcription by endogenous RNA polymerase II at promoter sites on both viral and cellular DNAs. In fact, these in vitro systems produce RNA with 5′ termini identical to those of mature mRNA's both in nucleotide sequence and 5′ modification by capping and methylation. The exogenously added template DNA can then be manipulated by deletion and substitution methods to define sequences directing the in vitro reaction. In general, sequences in the vicinity of the TATA consensus sequence or Goldberg-Hogness box seem essential for in vitro initiation (9, 20, 26, 36, 38, 39). In addition, alterations in sequences as far as 30 nucleotides on either side of the TATA sequence can affect the level of in vitro initiation at a promoter site.

The availability of cell-free transcription systems recognizing eucaryotic promoters allows direct access to questions of transcriptional regulation. It has already been shown, for instance, that the binding of a purified analog of simian virus 40 T antigen is capable of specifically repressing in vitro transcription from the simian virus 40 early region (29; U. Hansen and P. A. Sharp, personal communication). In most eucaryotic systems one is not so fortunate as to have purified regulatory factors whose activities have been well characterized in vivo. In vitro systems should provide a tool for identifying such factors as well as in characterizing their modes of action.

One situation where such factors must exist is during productive infection of human cells by adenovirus. The course of synthesis of different adenovirus genes during the lytic cycle follows a pattern: "early" functions begin expression at 1 to 4 h postinfection, and their expression is seen to decrease during the latter part of the early phase (4 to 7 h postinfection) (33). At 6 to 12 h postinfection DNA replication begins, thus increasing template copy number and potentially shifting the configuration of the template. The intermediate mRNA's for polypeptides IX and IVa_2 and the late form of 72K DNA-binding protein mRNA, all of which are undetectable at early times, are made in copious amounts during the first part of the late stage (see references 35 and 42 for review). The initiation of transcription occurs at the late promoter during the early and intermediate stages, but these transcripts are processed to yield only two mRNA's (8, 21, 28, 33). During the late stage of infection at least 13 different RNAs are processed from the late transcription unit and account for 50% of total mRNA synthesis (35, 42). This high level of

transcription from the late promoter may be due to a shift in the specific activity (initiations per template per minute) during the course of infection. However, positive regulation of the late promoter is hard to assess given the tremendous expansion in the pool of functional DNA template at late times.

MATERIALS AND METHODS

Cells and virus. HeLa S-3 cells were grown in suspension in RPMI or minimal essential medium with 5% horse serum. For virus infection, cells were concentrated 10-fold in the absence of serum, and 50 PFU of an adenovirus type 2 (Ad2) lysate stock per cell was added. Adsorption was carried out for 60 min, after which medium containing 2% horse serum was added to original volume. This was taken as the zero point of infection. To insure that infection had indeed occurred, a sample of each infection was carried to 32 h postinfection, fixed in phosphate-buffered saline with 10% formaldehyde, and then stained with fluorescent antibodies to adenovirus late protein; 96 to 98% of the cells were infected according to this assay.

Preparation of extract. Extracts were prepared as described by Manley et al. (24). Briefly, cells were washed in phosphate-buffered saline and then swelled and homogenized in a Dounce homogenizer in 4 volumes of hypotonic buffer. After addition of sucrose and glycerol the nuclei were lysed with the addition of ammonium sulfate to 10% saturation. The chromatin and other debris were removed by ultracentrifugation at 175,000 × *g* for 3 h, and the supernatant was concentrated by ammonium sulfate precipitation. This precipitate was resuspended and dialyzed in 100 mM KCl–12.5 mM $MgCl_2$–2 mM dithiothreitol–17% glycerol–0.5 mM EDTA, and samples were stored at −70°C after freezing in liquid N_2. Protein concentrations varied from 10 to 20 mg/ml, and extracts contained 0.1 to 0.5 mg of RNA per ml. Repeated freezing in dry ice and thawing for 1 min at 30°C produced no change either in levels or specificity of transcription. All transcriptions described in this paper were performed with extract frozen and thawed only once.

Transcription conditions. Except where noted, conditions were as described by Handa et al. (18). Preparative reactions were done at somewhat lower salt concentrations (40 mM KCl, 5 mM $MgCl_2$, 0.8 mM dithiothreitol, 7% glycerol), which enhanced transcription severalfold without changing specificity.

DNA and extract concentrations. Titrations of both DNA and extract yield nonlinear responses. Fixing extract and measuring runoff transcripts as a function of DNA concentration yields (i) a threshold DNA concentration below which no transcription occurs, and (ii) an inhibitory effect of high DNA concentration (24, 40). The requirement for a minimal DNA concentration is nonspecific—i.e., by using a concentration of a promoter-specific DNA which is below the threshold, "carrier" DNA, such as pBR322 or *E. coli* DNA, can be added to stimulate specific transcription. U. Hansen (personal communication) has shown that the duplex alternating copolymers polydeoxyinosinic-deoxycytidylic acid [poly(dIC:dIC)] and polydeoxyadenylic-deoxythymidylic acid [poly(dAT:dAT)] will act as carrier DNA, thereby demonstrating a total lack of sequence specificity in the bulk DNA requirement. A further advantage of these copolymers as carrier DNA is that the transcribed RNA products of the carrier poly(dIC:dIC) and poly(dAT:dAT) contain only two nucleotides. Thus, poly(dIC:dIC) carrier in a reaction containing [α-^{32}P]UTP yields no radioactive background. The key aspect of bulk DNA dependence is that at a fixed total DNA concentration, the molar yield of transcripts per promoter site is constant and independent of the source of carrier DNA. An exception to this generality is that at high carrier DNA concentrations, synthetic copolymers tend to more severely inhibit transcription than natural DNA.

A critical dependence of transcription upon extract concentration has been previously reported (24). In fact, DNA concentration dependence and extract protein concentration dependence are related. Specific transcription can be obtained in a range of 4 to 18 mg of extract protein per ml. At low extract concentration the DNA optima tend to be lower (in the range of 10 μg/ml). There is still a bulk DNA dependence, but it is less steep and the threshold concentrations are lower. At a high extract concentration the DNA titration becomes much sharper, and the threshold becomes much higher. At higher protein concentrations, often it is necessary to use 60 μg of DNA per ml for any transcription. Thus, for each new extract it is necessary to do careful DNA and extract titrations to determine optimal conditions.

For a given promoter, very short runoff transcripts (<300 nucleotides [n]) have a higher DNA optimum than longer runoff transcripts. This effect has been taken into account in the regulation and DNA titration experiments described here by measuring the synthesis of different-length runoff products from a promoter. No length dependence for transcription was observed with runoff products between 400 and 4,000 n.

Analysis of RNA products. Analytical reactions were done in 20 μl. After 90 min at 30°C, 150 μl of stop buffer (7 M urea, 100 mM LiCl, 0.5% sodium dodecyl sulfate, 10 mM EDTA, 350 μg of tRNA per ml, 10 mM Tris [pH 7.9]) and 300 μl of phenol-chloroform-isoamyl alcohol buffered by 20 mM Tris (pH 7.5) were added, and the tubes were blended in a Vortex mixer and centrifuged at 12,000 × *g* for 15 min. The aqueous phase (without interface) was extracted once more with phenol-chloroform-isoamyl alcohol and once with chloroform and then pooled with 200 μl of 1.0 M NH_4 acetate and 900 μl of ethanol. The RNA precipitate was collected after 60 min at −70°C and washed with 1 ml of ethanol. The pellet was air dried in an inverted tube for 5 min and resuspended in 20 μl of 10 mM $Na_2H(PO_4)$ (pH 6.8)–1 mM EDTA; to this was added 50 μl of 1.4 M deonized glyoxal–70% dimethyl sulfoxide–10 mM $Na_2H(PO_4)$ (pH 6.8)–1 mM EDTA–0.04% bromophenol blue. After 1 h at 50°C, 25 μl of the sample was loaded on 1.4% agarose gels in 10 mM $Na_2H(PO_4)$–1 mM EDTA (27). DNA restriction fragments were denatured with glyoxal and run as markers; 0.1% sodium dodecyl sulfate was added to the running buffer, and the gel was prerun for 10 min at 100 V to prevent the appearance of a sodium dodecyl sulfate front.

Preparative reactions for RNA fingerprinting. Large-scale reactions were as above, except that a second precipitation with 0.5 M NH_4 acetate and 3

volumes of ethanol was done. Autoradiography of 7 M urea–8% acrylamide gels was done at room temperature. Bands of interest were cut out, electroeluted, and subjected to fingerprinting as described previously (6, 31). End labeling of RNA by decapping and kinasing and selection on filters has been described previously (2, 3a, 43). Alkali breakage before selection and RNase A treatment on filters has been omitted.

RESULTS

In vitro transcription of Ad5: runoff mapping. Sizing of transcription runoff RNAs provides a sensitive means of identifying promoter sites. For this purpose, cloned restriction endonuclease fragments encompassing each of the nine known promoter regions of adenovirus have been constructed (K. Berkner et al., manuscript in preparation; see Table 2 for list of promoter sites). These recombinant pBR322-viral DNAs are cleaved with a variety of restriction enzymes and then used as template in a whole cell extract (WCE) reaction mix. The RNA products were specifically labeled by incorporation of [α-^{32}P] UTP and detected by autoradiography after electrophoresis in denaturing glyoxal gels (27). Detection of a particular length of α-amanitin-sensitive RNA product positions the promoter site relative to the terminus of the viral DNA fragment. Transcription of fragments generated by cleavage with other endonucleases confirms uniquely the position of the promoter site(s).

Some of the bands resolved after electrophoresis of [α-^{32}P]UTP-labeled RNA are not due to the initiation of transcription at promoter sites on viral DNA. First, some faint products are insensitive to α-amanitin (1 μg/ml) and are not RNA polymerase II products. These products are primarily due to various amounts of end labeling of ribosomal RNA present in the extracts (24). A second source of labeled bands is the inefficient but specific RNA polymerase II transcription of the pBR322 DNA present in the template (R. Kaufman and F. Laski, personal communication). High-molecular-weight RNA products are also generated by ligation of DNA segments containing active promoters to other fragments (unpublished results). Finally, it is possible that some of the faint bands generated during transcription are the consequence of RNA processing. In several cases, detailed analysis has failed to confirm this, and, where tested, transcripts were found to be stable on 2-h chases in both uninfected and infected extracts. For the above reasons, it is essential that a variety of templates be used to map a particular transcriptional region.

The *Hin*dIII G (0.0 to 7.9-map-unit [m.u.]) fragment of Ad5 DNA contains sites for initiation of EIa and EIb early mRNA's. This fragment was cloned into pBR322 through *Eco*RI linkers attached at 0.0 m.u. and through the *Hin*dIII site at 7.9 m.u. and was a kind gift of Kathleen Berkner. When cleaved with *Kpn*I, this plasmid generated 1,550- and 345-n RNAs. These sizes agree with the predicted size runoff RNAs for initiation at the EIa (1,550 n) and EIb (350 n) promoter sites (Fig. 1A). Similarly, cleavage of the plasmid with *Xba*I or *Hpa*I also yielded the correct size runoff RNAs (Fig. 1A). Additionally, RNA polymerase II and factors in the WCE recognize a site for initiation some 300 n upstream from the EIa promoter (PEIa; denoted with a dashed arrow in Fig. 1A). Initiation at this site generated α-amanitin-sensitive runoff transcripts of 1,850, 1,400, and 1,165 n after cleavage with *Kpn*I, *Hpa*I, and *Xba*I, respectively. mRNA's with 5′ termini mapping upstream from the EIa promoter have been reported (5, 12, 30, 41); thus, this may be an active promoter in vivo.

A comparison of in vitro transcription from the EIb and polypeptide IX promoter sites can be obtained by use of the Ad5 *Sma*I F fragment (2.8 to 11.1 m.u.). This fragment was cloned into pBR322 by attachment of *Eco*RI linkers (Fig. 1B). In vitro transcription of this plasmid after cleavage with *Eco*RI generated products of 2,250 and 380 n from the EIb and polypeptide IX promoter regions, respectively (Fig. 1B). The products predicted from DNA sequence are 2,240 and 360 n. Similarly, transcription of the *Kpn*I- or *Hin*dIII-cleaved *Sma*I F-pBR322 plasmid also yielded the expected runoff products (Fig. 1B). Thus, both the EIb and polypeptide IX promoter sites are recognized and are about equally active. No other major in vitro RNA polymerase II promoter site was observed in the *Sma*I F fragment.

The *Hin*dIII B fragment (72.8 to 89.1 m.u.) of Ad5 encompasses initiation sites for both early regions II (EII) and III (EIII). This fragment was cloned into pBR322 by ligation to the *Hin*dIII cleavage sites. Somewhat surprisingly, in vitro transcription of this plasmid after cleavage with a variety of different restriction endonucleases yielded strong runoff transcripts only from the EIII promoter. For example, transcription of plasmid DNA after cleavage with either *Eco*RI, *Xho*I, *Kpn*I, or *Xba*I generated RNAs of lengths 2,380, 2,200, 1,175, and 980 n, respectively, from the EIII promoter site (Fig. 2A). However, transcription of the same plasmid after cleavage with *Sal*I or *Bam*I produced only very faint bands of lengths 1,320 or 1,080 n, consistent with initiation at the EII promoter site. Thus the EII promoter is relatively inactive in the WCE system; however, other experiments show that the levels of EII transcription seen in vitro are indeed specific (see below).

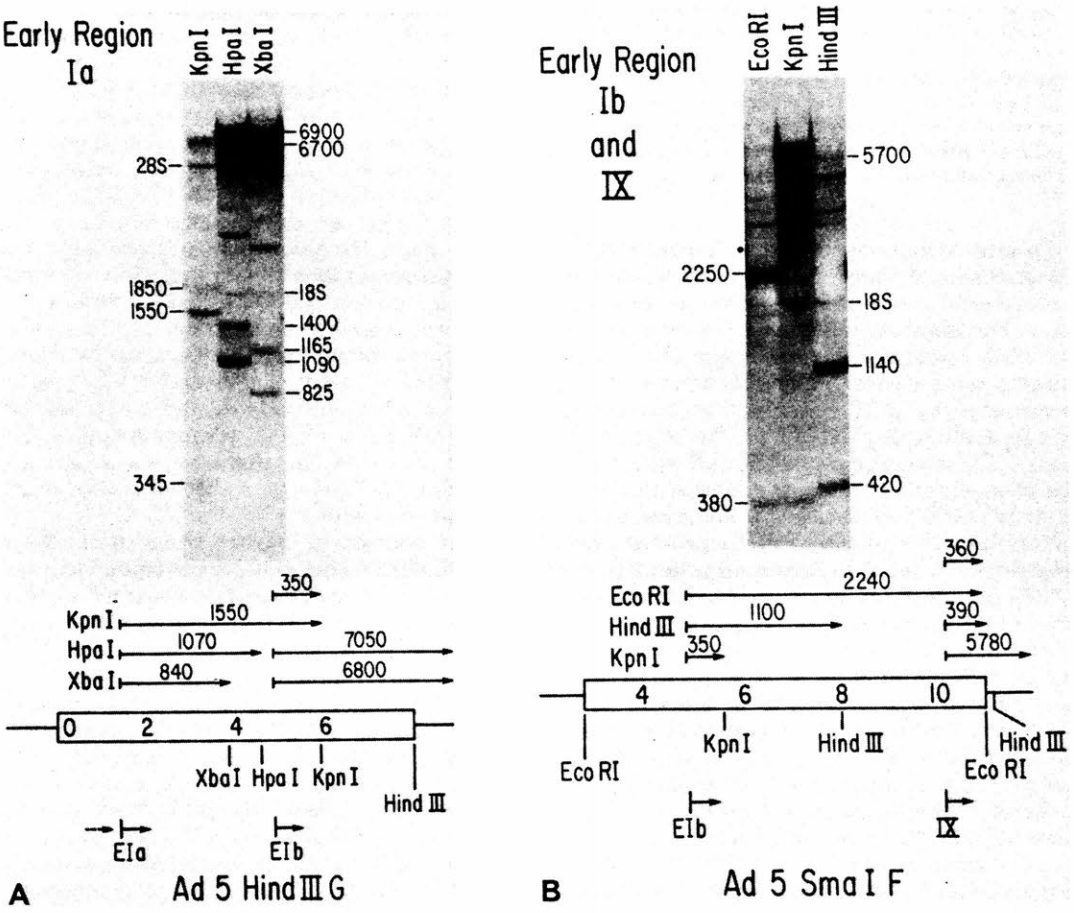


FIG. 1. *Transcription of EIa and EIb regions. Clones of HindIII G and SmaI F fragments of Ad5 were cleaved as shown, and unfractionated cut DNA was transcribed in an uninfected extract. Each 20-μl reaction mix contained 160 μg of WCE protein; 1 μg of specified DNA; 500 μM ATP, CTP, and GTP; 50 μM UTP at 2 Ci/mmol; 60 mM KCl; 10% glycerol; 7.5 mM $MgCl_2$; 5 mM EDTA; 15 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid-NaOH (pH 7.9); 8 mM creatine phosphate; and 0.5 mM dithiothreitol. Transcription was for 75 min at 30°C. The autoradiograms shown are 40-h exposures with intensifying screen. A, Digests of Ad5 HindIII G recombinant show EIa and an upstream initiation site. Nucleotide lengths alongside gel denote sizes calculated from DNA and RNA markers. Diagram shows localization of in vivo promoters and expected lengths of runoff transcripts (22, 23, 37). The bands of 1,550, 1,090, and 825 n are generated by EIa runoffs. Bands of 1,850, 1,400, and 1,165 n result from initiation at 0.7 m.u. Bands of 345, 6,900, and 6,700 n correspond to the EIb promoter. Bands between 18S and 28S are produced by initiation in the pBR322 vector. B, Digests of Ad5 SmaI F fragment show initiation at EIb and IX promoter sites. The runoff transcripts expected from both the EIb and IX promoter sites are indicated by their respective lengths. End-labeled 18S RNA is also indicated.*

Runoff transcription of early region IV of Ad5 was tested by using the *Eco*RI B fragment (84 to 100 m.u.). This fragment was cloned into pBR322 by attachment of *Eco*RI linkers to the right terminus of the genome and then insertion through the *Eco*RI site at 84 m.u. The addition of this plasmid after cleavage with either *Sma*I, *Hin*dIII, or *Kpn*I to an in vitro reaction mix generated RNAs of 235, 660, and 2,020 n, respectively—the sizes expected for products initiated at the EIV promoter site. There was another specific initiation site for RNA polymerase II weakly detected in the *Eco*RI B fragment. This site maps at 96.3 m.u., and its transcripts are in the same direction as the EIV promoter. Initiation at this site generated the 980- and 1,485-n products seen from templates cleaved with *Kpn*I and *Sma*I, respectively (Fig. 2B). There is no known in vivo Ad5 mRNA with 5′ termini mapping at 96.3 m.u. It is difficult to decide whether

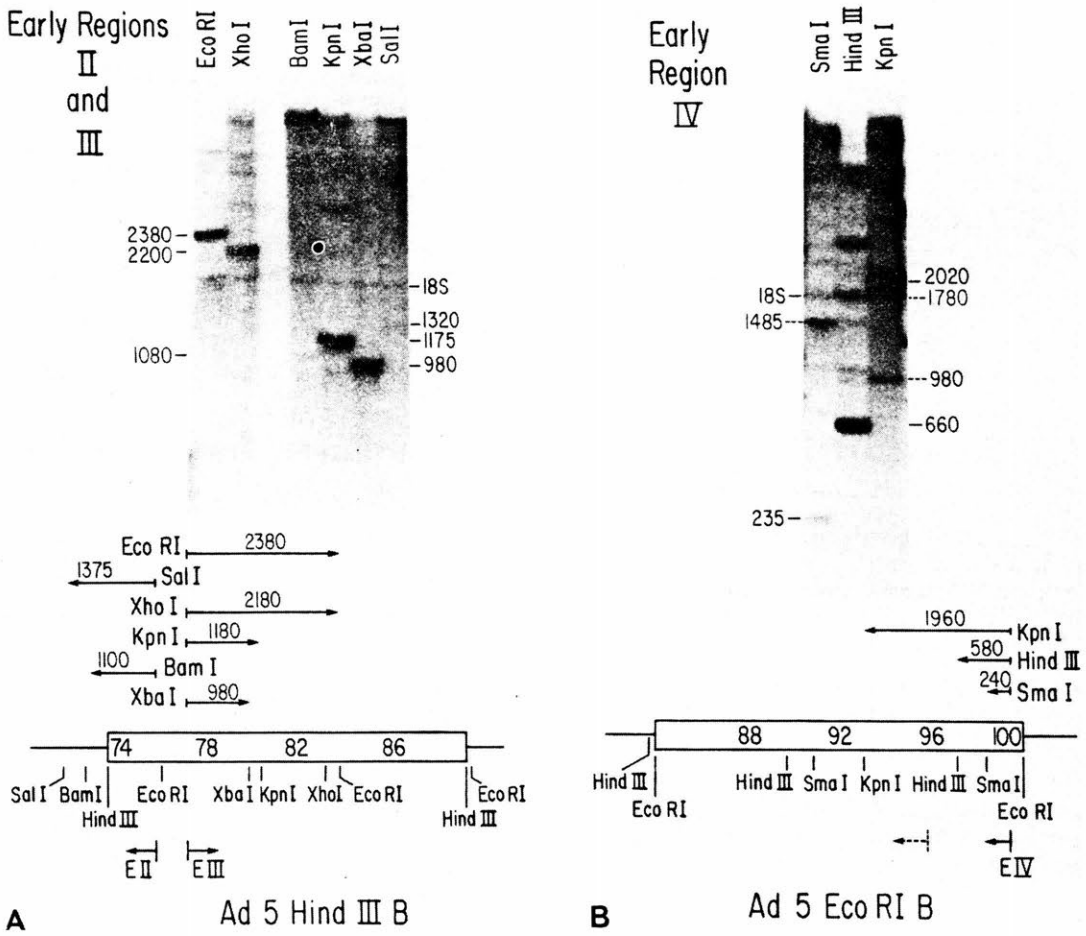


FIG. 2. *Transcription of EII and EIII region. Conditions of transcription were as in Fig. 1. A, Digests of Ad5 HindIII B recombinant show EII and EIII transcription. Bands of 2,380, 2,200, 1,175, and 980 n in the lanes marked EcoRI, XhoI, KpnI, and XbaI, respectively, correspond to EIII. Faint bands of 1,320 n and 1,080 n in lanes marked SalI and BamHI correspond to EII. B, Digests of Ad5 EcoRI B fragment show transcription of EIV and a minor initiation site (dotted lines). Expected runoff products from EIV are given in the diagram (34, 35). The lower molar yield of the 235-n runoff transcript in the first lane is due to a higher DNA optimum for very short transcripts (see text). Initiation at site near 96.3 m.u. generates bands of 1,485, 1,780, and 980 n on SmaI-, HindIII-, and KpnI-cleaved templates, respectively. Bands of 1,760 n in the HindIII lane and those migrating between 18S and 28S in all other lanes are pBR322 transcripts.*

this is artifactual recognition by the in vitro system or an undetected in vivo promoter.

Two intermediate stage promoters, those for IVa_2 and the late form of EII mRNA (EIIL) are not utilized in the in vitro system. Transcription of a fragment containing the latter promoter will be described below. Transcription of the IVa_2 promoter was assessed by using the *Bal* I E fragment of Ad2 (14.7 to 21.5 m.u.) cloned with *Bam*I linkers into the *Bam*I site of pBR322. The fragment was cut either with *Bam*I or with *Bam*I and *Hin*dIII. In both cases, a 460-n runoff would be expected from the IVa_2 promoter (Baker and Ziff, in press), in addition to late promoter transcripts of 1,750 and 196 n from the two templates, respectively. Though the late promoter bands were very strong, the expected comigrating IVa_2 runoffs were not seen (0.5% of the late promoter level would have been detected; data not shown).

Fidelity of in vitro initiation at the EIV promoter. Baker and Ziff (3, 3a) and Hashimoto and Green (19) have shown that the EIV promoter site generates microheterogeneous 5′ termini. As indicated in Table 1, the capped nucleotide forming the 5′ end of mRNA's from

TABLE 1. *DNA sequences preceding mRNA cap sites*[a]

"TATA" BOX / CAPS	Promoter (Coordinate)	Relative Efficiency: Mock	Relative Efficiency: Late	DNA Titration
GTG[TATTTATA]CCCGGTGAGTTCCTCAAGAGGCCACTCTTGAGTG	Ad5 EIa (1.4)	0.2	0.03	E
GGG[TATATAAT]GCGCCGTGGGCTAATCTTGGTTACATCTGACCTC	Ad5 EIb (4.7)	0.35	0.05	E
GAA[TATATAA]GGTGGGGGTCTTATGTAGTTTTGTATCTGTTTTGC	Ad5 Protein IX (9.8)	0.35	0.6	L
GGC[TATAAAA]GGGGGTGGGGGCGCGTTCGTCCTCACTCTCTTCCG	Ad2 Major Late (16.4)	1.0	1.0	L
TCCTTCGTGCTGGCCTGGACGCGAGCCTTCGTCICAGAGTGGTCC	Ad2 IVa$_2$ (15.9)	<0.005	<0.005	—
TAGTCCTTAAGAGTCAGCGCGCAGTATTTGCTGAAGAGAGCCTCC	Ad2 E II (75)	0.04	0.005	E
AGGTACAAATTTGCGAAGGTAAGCCGACGTCCACAGCCCCGGAGT	Ad2 E IIa Late (72)	<0.005	<0.005	—
Not Sequenced	Ad5 E II (75)	0.02	0.005	E
GGG[TATAA]CTCACCTGAAAATCAGAGGGCGAGGTATTCAGCTCAA	Ad2 E III (76.6)	0.3	0.05	E
Not Sequenced	Ad5 E III (76.6)	0.3	0.05	E
TCC[TATATATA]CTCGCTCTGCACTTGGCCCTTTTTTACACTGTGA	Ad5 E IV (99.1)	0.3	0.05	E

[a] Sites for initiation of in vivo transcription on adenovirus DNA and their relative activities in vitro. At left are the sequences surrounding initiation sites (1–3a, 22, 23, 37). The Goldberg-Hogness or TATA consensus sequence is boxed. Cap sites in vivo for each region are underlined. The G-string homology between the PL and PIX promoters is indicated. At right are the activities of the different promoters normalized in vitro to an activity of 1.0 for the late promoter. The far-right column shows the behavior of each promoter in a DNA titration curve. E indicates behavior similar to that of PEIa in Fig. 7, and L indicates behavior similar to that of PL in Fig. 7.

this region can be either an adenosine or any one of the six adjacent uridines (3a). To determine whether the heterogeneity was reproduced in vitro, RNA transcribed from this region was analyzed in detail. Short 250-n runoff RNAs from the EIV promoter site (PEIV) were prepared by transcription of the *Sma*I-cleaved *Eco*RI B fragment of Ad5 (Fig. 2B). These RNAs were labeled in vitro by incorporation of either α-^{32}P-labeled UTP, ATP, or GTP, and the products were resolved by electrophoresis in urea-acrylamide gels. The runoff bands were identified by autoradiography and electroeluted. After digestion with RNase T_1, the oligonucleotides were analyzed by two-dimensional fingerprinting (3a).

Figure 3A is an autoradiograph of a T_1 fingerprint of ^{32}P-labeled in vivo Ad2 nuclear RNA selected by hybridization to the Ad2 *Sma*K fragment (98.3 to 100 m.u.) (3a). Most of the T_1 oligonucleotides predicted from the EIV sequence were identifiable in the two-dimensional pattern (Fig. 3). Panels B, C, and D of Fig. 3 show the equivalent T_1 fingerprints of 250-n runoff transcripts labeled in vitro by [α-^{32}P] UTP, -ATP, and -GTP, respectively. Secondary analyses of these T_1 oligonucleotides with RNases T_2 and A are presented in Table 2. Spots a, b, c and 8 gave T_2-resistant radioactivity and thus contained cap structures (32). In addition, spot 8 gave products on secondary analysis expected from the major A cap oligonucleotide. Also of interest is the fact that no T_1 oligonucleotides were detected from sequences upstream of or spanning the in vivo initiation site. At this limit of analysis the in vitro and in vivo 5′ termini are identical.

A more precise means of examining the 5′ termini of in vitro-transcribed RNA is to uniquely label the 5′-terminal phosphate. This can be accomplished by removal of the 7-methyl guanosine by periodate oxidation and β elimination. The 5′ end is exposed by treatment with alkaline phosphatase, allowing end labeling with [γ-^{32}P]ATP and polynucleotide kinase (3a). Since the in vitro RNA products were diluted into a mixture of endogenous RNA, the 5′-labeled RNA was selected on filters containing the *Sma*I K fragment. Figure 4 shows such a comparison of the T_1 fingerprints of in vivo and in vitro Ad5 EIV RNA labeled at their 5′ termini. The series of labeled T_1 oligonucleotides is identical in the two cases. Some 60% of the initiations are at the adenine position, whereas the remaining 40% are distributed between the six adjacent uridine positions. The identification of the la-

beled 5′ termini as cap sites was further confirmed by the presence of labeled dinucleotides protected from T_2 digestion by 2′-*O*-methylation. A small fraction of in vitro-capped termini was unmodified by 2′-*O*-methylation as is shown by the presence of type 0 caps in the T_2 digest of the major A cap oligonucleotide in Fig. 3 (spot 8, Table 2). About 20% of the in vitro-synthesized RNA from PEIV did not have any cap structure at all, since about this fraction of the in vitro 5′ termini (and none of the in vivo termini) can be labeled with [γ-^{32}P]ATP by simple treatment with phosphatase and polynucleotide kinase. This subset of in vitro RNA chains also lack 2′-*O*-methylation on their terminal nucleotides. Whether this suggests that only a fraction of the newly initiated chains is modified by capping, and thus that cap formation is not a necessary component of transcription, or that only a fraction of the newly in vitro-synthesized caps is methylated and thus stable, cannot be determined with this methodology. The addition of (2 mM) *S*-adenosylmethionine to the WCE did not affect methylation levels or the fraction of capped 5′ termini (data not shown).

Heterogeneously capped 5′ termini could be generated either by endonuclease cleavage or by initiation of transcription at multiple sites. The frequency of initiation with a particular nucleotide would be expected to be dependent on the concentration of its triphosphate precursor. Conversely, the frequency of endonuclease cleavage should be independent of specific nucleotide concentrations. Panels C, D, and E of Fig. 4 show a comparison of 5′ termini synthesized in the presence of normal UTP and ATP concentrations, limiting UTP concentrations, and limiting ATP concentrations, respectively. The visible demonstration that changes in the ratio of ATP to UTP concentrations are specifically reflected in the ratio of adenine to uridine caps indicates that the microheterogeneous EIV termini are derived from initiation at each cap site.

In vitro transcription of early region II. Early region II is unique among early regions of adenovirus for several reasons: (i) it is transcribed at the early and late stages of infection by initiation at different promoter sites (7, 8); (ii) the early promoter site PEII (75 m.u.) does not have a TATAAA consensus sequence positioned 25 to 31 n upstream from the cap site (Table 1); (iii) the late promoter site for early region II (PEIIL) at 72 m.u. has a TACAAA sequence at the expected consensus sequence positions (3a); and (iv) both EII and EIIL promoter sites are efficient in vivo. It is therefore somewhat surprising that neither of these promoters is efficiently recognized in the WCE. The Ad2 *Eco*RI F (70.7 to 75.9 m.u.) fragment encompasses both EII and EIIL promoter sites (Fig. 5). In vitro transcription of this fragment after cleavage with *Eco*RI, *Kpn*I, and *Sst*I generated low levels of runoff transcripts of 1,460, 1,240, and 880 n, respectively, those expected from initiation at PEII. The PEII activity detected is about 0.04 of that observed from the late promoter of Ad2 (Fig. 5; a myriad of other minor bands appear on this overexposure, some of which correspond to the initiation sites within pBR322). Even this low level of runoff transcription was not observed for initiation at PEIIL (72 m.u.). An RNA of 320 n would be predicted for an *Eco*RI-cleaved template (see Fig. 5), and an RNA of 1,200 n would be predicted from *Bgl*I cleavage (data not shown).

To show that the in vitro system actually initiates correctly at PEII, RNA transcribed in the WCE was subjected to 5′ terminus analysis as described previously for PEIV. In this case, RNA was extracted from a very large reaction mix (2 ml) and 5′ labeled as described above. The 5′-labeled RNA was selected by hybridization to a *Hin*dIII-*Eco*RI fragment spanning from 72.8 to 75.9 m.u., digested with RNase A, and fingerprinted. The two large oligonucleotides (spots 1 and 2 of Fig. 6) were shown to carry the 2′-*O*-methylation and behaved identically to spots derived from the in vivo termini on secondary analysis shown in Fig. 5B and C of reference 2. Thus, the faint bands from PEII are initiated at the same sites as in vivo mRNA's. Interestingly, many of the smaller oligonucleotides in Fig. 6 also have 2′-*O*-methylated bases. Since the in vitro reaction mix was scaled up 10-fold over that used in Fig. 4, these capped oligonucleotides probably represent other in vitro initiation sites within the 1,100-n DNA fragment used to select hybrids for fingerprinting. This suggests that RNAs initiated at nonpromoter sites are also modified by cap synthesis.

A similar labeled 5′ terminus analysis with a hybridization probe spanning the EIIL promoter site selected a number of 5′-terminal 2′-*O*-methylated oligonucleotides, none of which corresponds to the in vivo PEIIL initiation sequence. This demonstrates that PEIIL is very infrequently utilized as a promoter site in vitro.

Relative transcriptional activities of the different promoters. To determine the relative rates of transcription of different viral promoters under a variety of conditions, an equimolar mixture of template fragments (see Fig. 8A) was constructed. Each fragment generated runoff transcripts of unique length. Since the in vitro transcripts are labeled with [α-^{32}P]UTP,

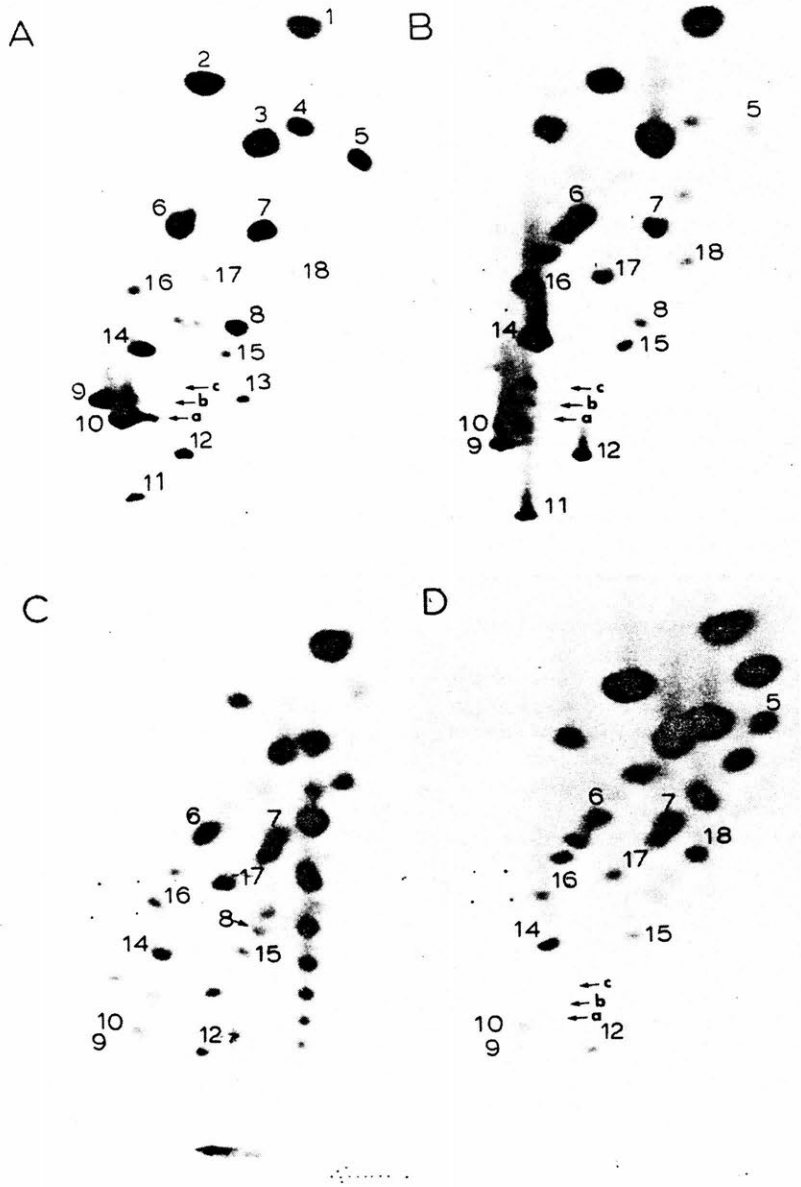


FIG. 3. *Fingerprints of 250-n runoff transcripts from PEIV. A, Ad2 early nuclear RNA was selected on filters by hybridization to Ad2 SmaI K fragment (98.3 to 100 m.u.). RNA was labeled with $^{32}PO_4$ in vivo as described previously (2). Selected RNA was treated with RNase T_1 on filters, the RNA was eluted, and an RNase T_1 digest was resolved by the two-dimensional fingerprint analysis of Brownlee et al. (6, 31). B, Ad5 EcoRI B recombinant was cut with SmaI and transcribed in a reaction mix as noted in the text. Nucleotide concentrations were 50 μM each ATP, CTP, and GTP and 100 μM [α-^{32}P]UTP at 450 Ci/mmol. The 250-n runoff band was electroeluted from urea acrylamide gel and fingerprinted with RNase T_1 as above. Spot 8 is the major A cap, and spots a, b, and c are three of the minor U caps. C, The same template as in B was transcribed with 75 μM [α-^{32}P]ATP (specific activity, 304 Ci/mmol), 50 μM CTP and GTP, and 5 μM UTP. The series of spots along the right side appears to be oligoadenylic acid. D, The same template as in B was transcribed with [α-^{32}P]GTP (specific activity, 600 Ci/mmol); 50 μM ATP, CTP, and UTP; and 29 μM GTP. Diagram, Sequence flanking EIV promoter site of Ad5 (34) shows aligments for T_1 spots of panels A through D on the basis of secondary analyses presented in Table 2. Oligonucleotides upstream or traversing the cap and not found in in vitro runoff products are marked NP. Spot 13, found in nuclear RNA, is absent from the in vitro runoff products since the template is truncated in this oligonucleotide. Note the conformation of the major A cap (spot 8) by identification of RNase A products. Similar analysis of the in vivo spots has been shown (3a).*

NP
NP
NP
r 5′ CGTAGGTTCGCGTGCGGTTTTCTGGGTGTTTTTTTGTGGACTTTAACCGTTACGTCATTTTTTAGTCCTATATATACTCGCTCTGCACTTG
l 3′ GCATCCAAGCGCACGCCAAAAGACCCACAAAAAAACACCTGAAATTGGCAATGCAGTAAAAAATCAGGATATATATGAGCGAGACGTGAAC
-100
-80
-60
-40
-20
7mGpppACACUG---
7mGpppUACACUG---
7mGpppUUACACUG---
7mGpppUUUACACUG---
7mGpppUUUUACACUG---
7mGpppUUUUUACACUG---
7mGpppUUUUUUACACUG---
TRANSCRIPTION
GCCCTTTTTTACACTGTGACTGATTGAGCTGGTGCCGTGTCGAGTGGTGTTTTTTTTAATAGGTTTTCTTTTTTTACTGGTAAGGCTGACTG
CGGGAAAAAATGTGACACTGACTAACTCGACCACGGCACAGCTCACCACAAAAAAATTATCCAAAAGAAAAAATGACCATTCCGACTGAC
+1
+20
+40
+60
+80
TTAGGCTGCCGCTGTGAAGCGCTGTATGTTGTTCTGGAGCGGGAGGGTGCTATTTTGCCTAGGCAGGAGGGTTTTTCAGGTGTTTATGTG
AATCCGACGGCGACACTTCGCGACATACAACAAGACCTCGCCCTCCCACGATAAAACGGATCCGTCCTCCCAAAAAGTCCACAAATACAC
+100
+120
+140
+160
TTTTTCTCTCCTATTAATTTTGTTATACCTCCTATGGGGGCTGTAATGTTGTCTCTACGCCTGCGGGTATGTATTCCCCCGGG 3′ r
AAAAAGAGAGGATAATTAAAACAATATGGAGGATACCCCCGACATTACAACAGAGATGCGGACGCCCATACATAAGGGGGCCC 5′ l
+180
+200
+220
+240
Sma I

TABLE 2. *Analysis of RNase T_1 oligonucleotides from Ad5 EIV RNA transcribed in vitro*

Spot no.[a]	RNase A products[b]		
	[α-^{32}P]UTP label[c]	[α-^{32}P]ATP label[c]	[α-^{32}P]GTP label[c]
1	NA	NA	NA
2	NA	NA	NA
3	NA	NA	NA
4	NA	NA	NA
5	Gp	NA	Cp
6	AUp, Gp	Up, Gp	AUp, Up
7	ACp, Gp	Gp	[AAGp, AGp, Gp, (AUp, Cp, Up)][d]
8	ACp, Gp	[Type O and I cap cores][e]	NP
9	ACp, Cp, Up	Up	Up, Gp
10	AAUp, Up, (ACp, Cp)[f]	AAUp, Up	AGp
11	AAUp, AUp, Cp, Up, (Gp)	NA	NA
12	AUp, Cp, (Up)	AUp, Up	AUp, Gp
13	NP	NP	NP
14	AUp, Up, Cp	AAUp,[d] Cp, Up	Up, AGp
15	Cp	Up	ACp
16	AUp, Gp, Up	Up	AUp, (Up)[d]
17	AAUp, Gp	AAUp, Up	AAUp
18	Cp	NA	AGp
a	[Cp, Up, cap I core][e]	NP	[cap I core, Up][e]
b	[Cp, Up, cap I core][e]	NP	[cap I core, Up][e]
c	[Cp, Up, cap I core][e]	NP	[cap I core, Up][e]

[a] Numbers correspond to the spot numbers of Fig. 3.

[b] For secondary analysis, RNase T_1 oligonucleotides were redigested with RNase A. Products from the secondary digests were fractionated by electrophoresis on DEAE paper at pH 3.5.

[c] These are the products resulting from RNase A redigestion of the RNase T_1 oligonucleotides from the fingerprints, shown in Fig. 3B, C, and D, of RNA labeled in vitro with [α-^{32}P]UTP, [α-^{32}P]ATP, or [α-^{32}P]GTP. NA indicates that the corresponding spot was not analyzed; NP indicates that the spot was not present in the fingerprint. Products in parentheses were barely detectable. Abbreviations: A, adenosine; C, cytidine; U, uridine; G, guanosine; P, phosphate.

[d] This product(s) was not predicted by the Ad5 sequence in Fig. 3.

[e] These products are the result of redigestion with RNase T_2.

[f] Spot 10 was contaminated with spot 9.

the amount of radioactivity incorporated per mole of product is proportional to length. The late promoter site (PL) was the most active template in the mix, but only generated a faint band of 225 n under these conditions. In addition to the adenovirus promoters, transcripts from the various vector sequences were observed, one migrating slightly below 18S and others migrating well above the EIb band at 2,575 n. The relative strengths of various promoter sites can be approximated by densitometry of gels similar to that in Fig. 8A or of gels containing a single runoff transcript such as those in Fig. 1 and 2 (Table 1). The two methods yield comparable results under similar conditions.

It is well known that the in vitro transcription activity of different promoters can vary markedly with changes in total DNA concentration. It has been reported (29; U. Hansen and P. A. Sharp, personal communication) that relative transcriptional activities of different simian virus 40 promoters vary with DNA concentration. A similar observation with adenovirus promoters is shown in Fig. 7. An equimolar mixture of template DNAs for PEIa (1,220-n runoff) and PL (924-n runoff) was prepared, and a DNA-titration curve was performed. As shown in Fig. 7A, both early and late promoters gave qualitatively similar titration curves, with a threshold value below which no transcription was seen, and inhibition by high DNA concentration (see above). A graph of the molar ratio of early and late transcripts is shown in Fig. 7B. There was a 20-fold decrease in the ratio over a DNA range of 7 to 56 μg/ml. That this is a function of bulk DNA concentration is shown by a similar curve when total promoter DNA was held constant and DNA concentration was titrated up with increasing concentrations of duplex copolymer poly(dIC:dIC). Adenovirus promoters fall into two general classes in the DNA titration curve: all early promoters respond like pEIa, whereas the polypeptide IX promoter behaves like the late promoter (data not shown).

A comparison of the relative efficiencies of the different promoters each at their optimal DNA

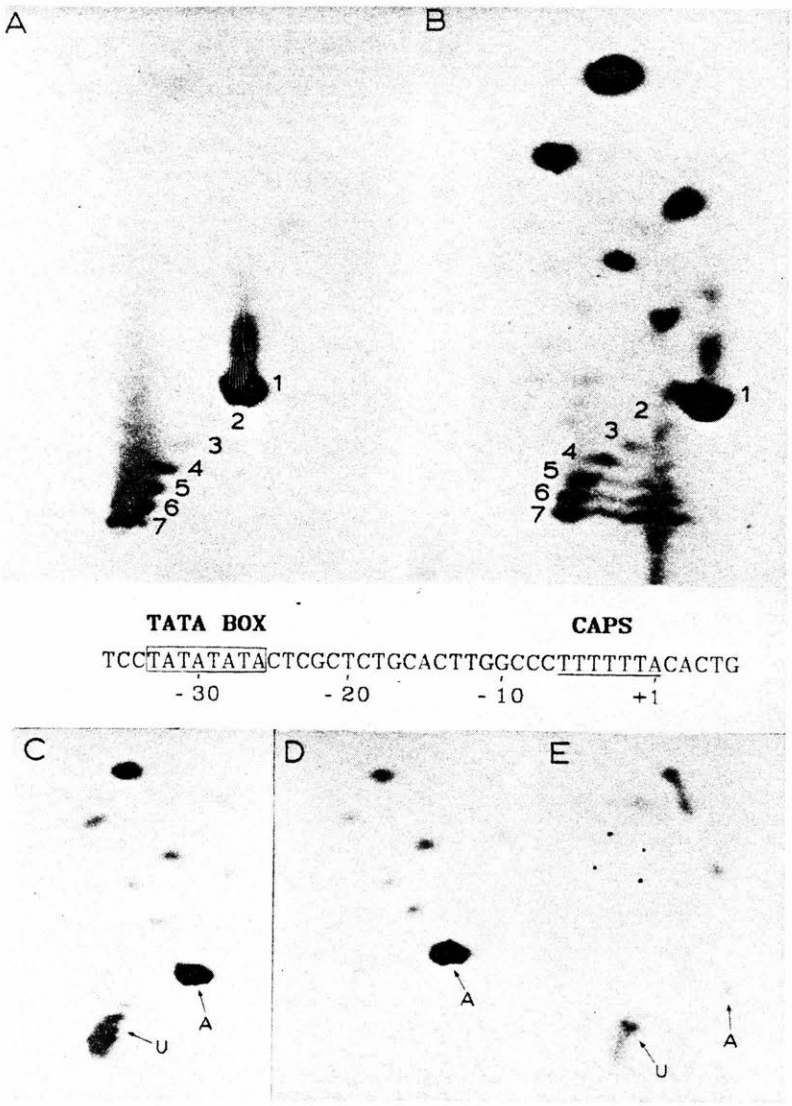


FIG. 4. *Localization of 5′ termini from the EIV region. RNAs were 5′ labeled by decapping and kinasing using previously described procedures (3a). The RNA was selected and fingerprinted as in Fig. 3A. In both A and B, spot 1 and spots 2 through 7 derive from the major A terminus and minor U termini, respectively. In each case, identity of the spots was confirmed by redigestion with nuclease P1, RNases T_2 and A, and chromatography on 540 or DEAE paper (3a). A, Analysis of cytoplasmic RNA prepared from cycloheximide-treated, Ad5-infected HeLa cells at 5 h postinfection. B, Analysis of in vitro-transcribed RNA; 6.4 μg of BglII-cleaved EcoRI B recombinant was transcribed in a 180-μl reaction mix under preparative conditions (see text). The prominent spots above the A terminus are not 2′-O-methylated and thus probably do not correspond to capped 5′ termini. A number of the very minor spots were shown to have 2′-O-methylation and probably represent minor initiation events. The sequence of EIV showing seven capped termini and the TATA box is shown in the diagram. C, D, and E, Variability of relative levels of A and U caps with nucleotide concentration. Transcriptions in 0.2 ml with 4 μg of HindIII-cleaved EcoRI B were analyzed in A. Spots derived from A caps and U caps are indicated. C, Nucleotide concentrations were 50 μM ATP, GTP, CTP, and UTP. D, Nucleotide concentrations were 50 μM GTP and CTP, 500 μM ATP, and 1 μM UTP (endogenous pool). E, Nucleotide concentrations were 50 μM GTP and CTP, 500 μM UTP, and 10 μM ATP. Incorporation was much lower in this sample, due to limiting ATP.*

concentration is shown in Table 1. It should be noted that whereas the trailing and leading peak shapes differ, the actual optimal DNA concentrations for the different promoters under the conditions used were quite similar.

Regulation of transcription is the major level

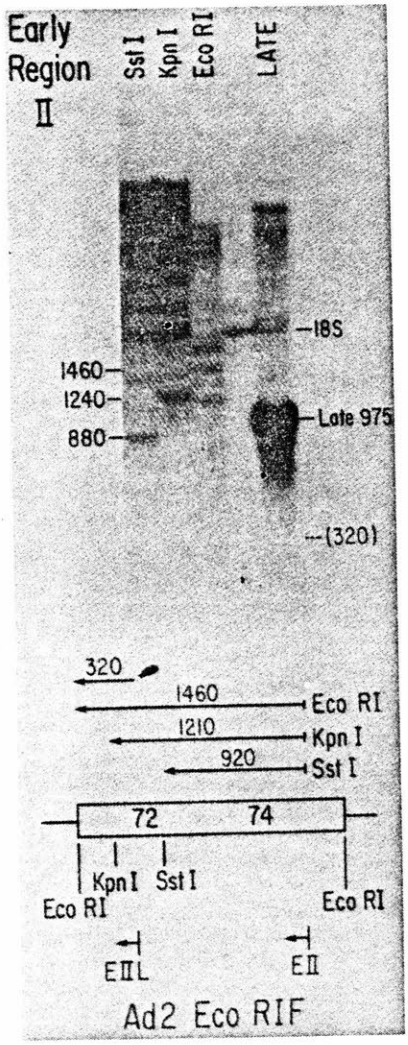


FIG. 5. *Runoff mapping of EII. Conditions were as in Fig. 1, except that autoradiography was for 150 h. Bands of 880, 1,240, and 1,460 n in the SstI, KpnI, and EcoRI lanes, respectively, correspond to EII transcription (2, 13). Bands of 2,300, 1,960, and 1,740 n correspond to transcripts from pBR322. The other minor bands have not been characterized. The lack of a band migrating at 320 n in the EcoRI-digested lane indicates that transcription from the EIIL promoter at 72 m.u. does not occur under these conditions. In addition, transcription of a BglI-cleaved template which would generate EIIL runoff of 1,200 n failed to yield such a band (data not shown). A major late promoter incorporation done under the same conditions is shown.*

of gene control during the adenovirus lytic cycle. Factors responsible for this regulation must appear or disappear (or both) during various phases of the lytic cycle. Extracts were prepared from cells at different stages of infection in an attempt to detect such factors. HeLa cells were infected with Ad2 at 50 PFU/cell and were harvested 6 h postinfection for early extracts or

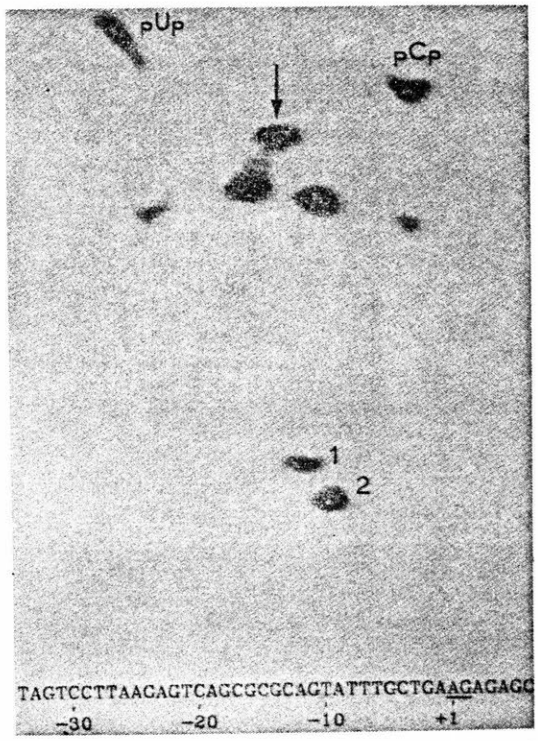


FIG. 6. *Early region II fingerprint. The EcoRI F recombinant shown in Fig. 3A was cut with EcoRI and HindIII; 50 μg of this template was transcribed in a 2-ml reaction mix under the conditions described in the legend to Fig. 5. The subsequent manipulations were as in Fig. 5, except that the HindIII-EcoRI fragment spanning 72.8 to 75.9 m.u. was immobilized on filters to select RNA. The selected RNA was digested with ribonuclease A (which cuts on the 3′ side of pyrimidines). The lower doublets of spots 1 and 2 were identified as GAGAGC and AGAGAGC, respectively, which correspond to the in vivo termini for EII as reported in Baker et al. (2). This was confirmed by secondary analysis with RNases T_2 and T_1 and nuclease P1. These spots in addition carry the 2′-O-methylation indicative of capped termini. Because the reaction is scaled up 10-fold from that in Fig. 5, 5′-labeled termini from background initiation have been proportionally enhanced. The CU spot indicated by an arrow and a number of the other oligonucleotide spots carry 2′-O-methylation, indicating that they were derived from capped 5′ termini encoded within the fragment. These other termini were not seen in vivo (Fig. 3). The diagram below shows the sequence around the promoter site of EII (75.0 m.u.) with A and G 5′ termini at positions +1 and +2, respectively.*

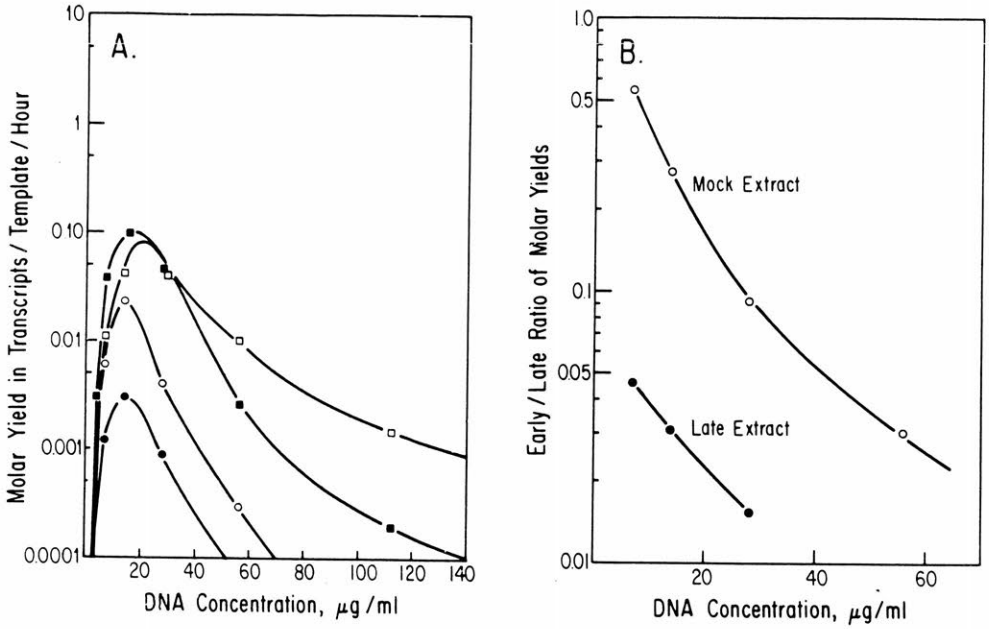


FIG. 7. *DNA titration curves for EIa and the late promoter in uninfected and infected extracts. Cleaved plasmids containing PEIa and the late promoter were mixed in an equimolar ratio for template. The EIa promoter generated a runoff of 1,220 n, and the Ad2 late promoter generated a runoff of 974 n. Similar results were obtained with a 4,000-base-pair runoff of the late promoter. Transcription conditions were as in Fig. 1, except that nucleotide concentrations were 1/10 of those given. Quantitation was by scanning with a laser densitometer. A, Specific activity of promoters expressed in moles of transcript per mole of template in a 1-h reaction. Symbols: □, PL promoter in mock extract; ○, PEIa in mock extract; ■, PL in late extract; ●, PEIa in late extract. B, Early/late ratio as a function of DNA concentration.*

21 h postinfection for late extracts. In parallel, a mock-infected extract was prepared from the same batch of HeLa cells. These extracts had roughly similar total transcriptional activities.

In light of variation in relative activity of different promoters with changes in DNA concentration it is important to ensure that any effects seen in infected extracts are not simply the result of endogenous viral DNA contaminating those extracts. Extracts were assayed for viral DNA by two methods. (i) Total nucleic acid corresponding to several reaction mixes was electrophoresed on neutral agarose gels with and without *Hin*dIII digestion and stained with acridine orange (27). No viral DNA was observed. (ii) The level of endogenous VA RNA synthesis in infected cell extracts was measured and compared with a standard of added viral DNA. Less than 0.02 μg of template viral DNA was observed per 20 μl of reaction mix. The addition of three times this concentration of viral DNA sequences to the transcription shown in Fig. 8A did not change the ratio of the runoff transcripts (data not shown). All of the extracts also gave similar bell-shaped DNA titration curves for the various promoters (Fig. 7A), indicating that a certain bulk of exogenous DNA was still needed for transcription to occur.

Since the strong in vivo promoter for EII was only marginally active in extracts from uninfected cells, it was interesting to test its activity in extracts from early and late-infected cells. The Ad5 *Hin*dIII B fragment cloned into pBR322 was cleaved with *Xho*I and *Sal*I, which would generate runoffs of 2,200 and 1,375 n from the EIII and EII promoter sites, respectively. Only low levels of the 1,375-n EII runoff were generated with mock and early extracts; this runoff was undetectable in a late extract (data not shown). The EIII promoter site was fully active in both mock and early extracts, but had less activity in late extracts. Similar experiments have been done with DNA fragments spanning the EIIL or the IVa_2 promoter sites. In neither case was transcription detected from early or late extracts. Thus, infected extracts seem to lack components which allow transcription from these promoters.

When transcription from the equimolar mixture of efficient promoters was compared in extracts from mock, early, and late stages in infection, a shift was observed (Fig. 8B). Although

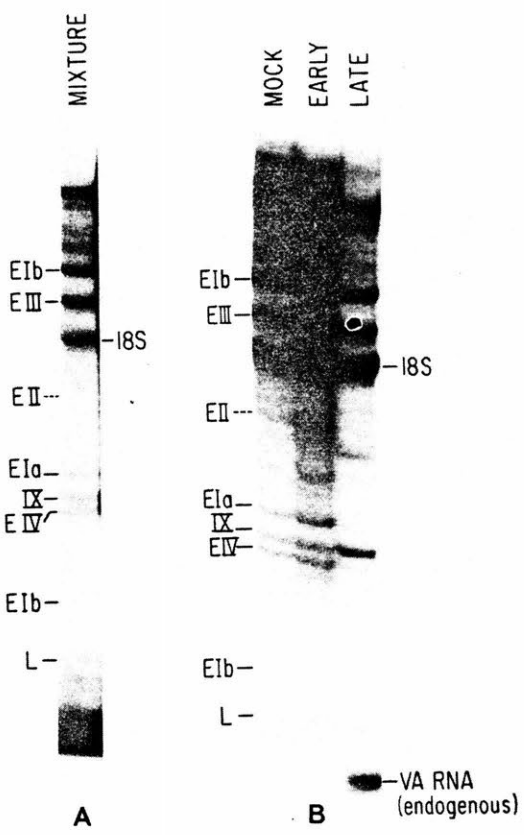


FIG. 8. *Mixture of templates from different regions. An equimolar mixture of the following cleaved plasmids was prepared: Ad5 HindIII G recombinant cut with XbaI and KpnI, which generates runoffs of 340 n from PEIb, 840 n from PEIa, and 1,165 n from the initiation site at 0.7 m.u. (Fig. 1A); Ad5 SmaI F recombinant cut with BamHI, which generates runoffs of 2,575 n from PEIb and 715 n from PIX (Fig. 1B); Ad2 BalI (14.7)-HindIII (17.0) cloned from BamHI to the HindIII site of pBR322—this recombinant cut with EcoRI, which generates a runoff of 225 n from the major late promoter; Ad5 HindIII B recombinant cut with XhoI and SalI, which generates runoffs of 1,375 n from PEII and 2,180 n from PEIII (Fig. 2A); Ad5 EcoRI B recombinant cut with HindIII, which generates a 680-n runoff from PEIV (Fig. 2B). The equimolar mixture (10 μg/ml) was supplemented with 15 μg of poly(dIC:dIC) per ml of carrier as noted in the text. Omission of any of the above DNAs removed the corresponding runoff bands from the labeled RNA pattern. A, Mixture transcribed in an uninfected extract. Due to differences in runoff length and promoter strength, it was often necessary to examine several exposures to accurately compare activity. B, Comparison of extracts from mock-, early-, and late-infected cells. Cells were harvested, and extracts were prepared from mock-infected cells and cells at 6 and 21 h postinfection with Ad2. Activities were determined for a variety of extract concentrations, and the reactions giving peak transcription are shown. The relative ratios of bands from an extract did not change over the range of the titrations. Protein concentrations were: mock, 6.3 mg/ml; early, 6.1 mg/ml; late, 5.8 mg/ml. A 150-n band in the late extract resulted from endogenous VA RNA (virus-associated small RNA) synthesis. The level of VA transcription is comparable to that obtained with 1.0 μg of exogenously added Ad2 DNA per ml. Two runoff bands of 2,575 and 340 n from EIb have been included. Lighter and darker exposures allow comparisons of these two bands in the different extracts. The ratio of radioactivity in these two EIb runoffs did not vary. The band of 1,140 n, which was strongly enhanced in the late extract, corresponds to the rightward promoter at 0.7 m.u.*

mock and early extracts produced essentially identical responses, the late extract showed preferential transcription of the intermediate PIX and the late (PL) promoters relative to the early promoter sites. In Fig. 8B, it is particularly striking to compare PIX with the flanking PEIa and PEIV runoffs. This shift in transcription specificity has been observed repeatedly with different mock and late extracts. The observation of relative enhancement of transcription of the PIX and PL promoter sites in late extracts did not change with variations in time of incubation, ionic strength, or nucleotide concentration, suggesting that simple differences in preparation do not account for the shift. Figure 7 shows a comparison of the DNA titration curves for the two extracts. Note that the DNA concentrations for optimal transcription of the early and late promoters in the two extracts are similar, but that the PEIa/PL ratio of the late extract remains 10-fold lower than that in an uninfected extract throughout the course of the DNA titration curve. A more complete description of the comparison of the transcription capacity in mock and late extracts is presented by Fire et al. (11).

DISCUSSION

Adenovirus mRNA synthesis results from initiation of transcription at nine sites on viral DNA (35, 42). Baker and Ziff (3a) have defined the set of capped nucleotides at each of these positions (Table 1). Six of these nine sequences are efficiently recognized by RNA polymerase II and factors for initiation of transcription in the WCE system. Each of these six sites has an obvious TATA or Goldberg-Hogness consensus sequences between −25 to −31 n. Two previous studies have shown that the EIa, EIb, and IX promoter sites are active in in vitro systems (24, 39). Three sites which encode RNA 5′ ends, the promoter sites for IVa_2 (15.9 m.u.) and the early and late promoter sites for region II (PEII [75 m.u.] and PEIIL [72.0 m.u.]), do not have consensus TATA sequences and are poorly recog-

nized in the WCE reaction mix. Very sensitive analysis of capped termini shows that transcription is specifically initiated at the PEII position, but at 1/25 the frequency observed from the Ad2 late promoter. With the same assay, transcription from the late promoter site for region II (PEIIL) is not detected (less than 0.5% of PL [16.4 m.u.]). This site has TACAAA at the position of the consensus sequence. The promoter site for the rabbit α-uteroglobin has an identical consensus sequence and has been reported to be efficiently transcribed in this system (S. Woo, personal communication). Thus, factors in the WCE must be discriminating between these promoters on the basis of sequences other than the TACAAA.

Several other results also suggest that RNA polymerase II and factors in the extract interact with sequences in addition to the TATA. Many perfect consensus sequences in viral or vector DNAs are not recognized for transcription initiation. In fact, in all of the adenovirus DNA sequences surveyed in this study, only two sites were detectably utilized for initiation that were not previously known to be in vivo initiation sites; these sites mapped on the *r* strand at approximately 0.7 m.u. and on the *l* strand at 96.3 m.u. The former may account for minor mRNA's detected from Ad2-, Ad5-, Ad7-, and Ad12-infected and -transformed cells (5, 12, 30, 41). No Ad2 mRNA has been mapped with a 5′ terminus at 96.3 m.u. (5).

A direct test of the importance of the TATA sequences for in vitro transcriptional activity emerges from studies in which these sequences have been modified by deletion or mutation. Hu and Manley (20) concluded that the TATAAA sequence in the late Ad2 promoter was essential for in vitro transcriptional activity. Studies of similar deletions in the conalbumin (9), ovalbumin (36), early simian virus 40 (26), and late Ad2 (9) promoters have given similar results. In fact, Wasylyk et al. (38) showed that conversion of the third base, T, to a G in the conalbumin promoter sequence of TATAAA almost abolishes the in vitro reaction. Additionally, Hu and Manley's results indicated that deletion of sequences from either −51 n upstream or +5 n downstream of the late cap site also affected the efficiency of transcription, suggesting the in vitro reaction senses 60 n of sequence (20). Thus, the picture emerges that the TATA consensus sequence is of central importance for in vitro recognition, and that sequences lying 25 n to either side of this site can affect the efficiency of initiation.

There are reasons to suspect that the TATA consensus sequence is of secondary importance for in vivo transcription initiation. Where it has been studied, deletion of the TATA sequence does not significantly reduce in vivo initiation of transcription from a region, but does generally render an RNA product with widely dispersed 5′ termini (4, 10, 14, 15). Hence, the TATA consensus sequence seems to have a positioning role for RNA polymerase II. In addition, there seems to be a marked in vivo dependence on sequences beyond −60 n (4, 10, 15–17), whereas deletion of these sequences has no effect in vitro. The in vitro reaction as it is now constituted is only a shadow of the in vivo process; it probably is only responsive to the higher affinity of RNA polymerase II and factors for sequences involved in positioning the complex for initiation.

One of the more remarkable features of the in vitro system is the fidelity of its reproduction of the nucleotide specificity of the in vivo initiation. This was shown previously for the late promoter of Ad2 (24, 40) and is vividly seen here in the microheterogeneity of in vitro initiation at the EIV promoter site. Both in vivo (3a) and in vitro RNA initiation at the Ad5 PEIV occur at any one of seven adjacent nucleotides, a string of six thymidine bases, or the adjacent adenosine (Table 1). This suggests that the biochemical complex specifying the in vitro initiation is identical to that specifying the actual initiation event in vivo. Although the relative ratios of PEIV caps under standard conditions in vitro are the same as those in vivo, the distribution of sites of in vitro initiation is sensitive to the concentration of free nucleotide triphosphates. Reduction of relative UTP or ATP concentrations can drastically reduce the fraction of chains initiated with uridine or adenine, respectively. Thus, initiation at multiple sites rather than RNA processing by 5′ cleavage must account for the heterogeneous termini.

The lytic cycle of adenovirus evolves through a series of temporal stages where subsets of the nine promoters listed in Table 1 are optimally active. Either factors, RNA polymerase II or the viral template, must be modified during the course of the cycle to enhance transcription of various sites. The finding that WCEs from uninfected cells initiate at either early, intermediate, or late promoter sites with roughly comparable efficiency suggests that the in vivo transcription regulation is not fully reproduced in the in vitro system. This is further suggested by the lack of observed differences between extracts from mock- and early-infected cells as well as the lack of stimulation of the early PEII or intermediate PEIIL and PIVa_2 promoters in any of the infected extracts. Thus, some of the factors that play a role in adenovirus transcription in vivo are either missing or nonfunctional in WCEs.

One striking feature of the lytic cycle of adenovirus is the shift in rates of transcription between early and late promoters at the time of viral DNA replication. A shift in the ratio of transcription of late versus early promoters was also observed in vitro. An increase in the total DNA concentration enhanced transcription of both PL and PIX promoters relative to any early promoter in extracts prepared from either uninfected or late-infected cells. PL and PIX were also distinguished from the early promoters in their relatively enhanced activity in extracts prepared from late-infected cells. The ratio of transcription from either PL or PIX relative to an early promoter was 10-fold higher in late extracts than in mock extracts (Table 1). Thus, soluble factors in uninfected extracts distinguish late promoters from early promoters. Shifts in the level of these factors in late extracts could account for the enhanced transcription of late promoters. It should be possible in the future to identify these factors. Similarly, it is likely that some common sequence feature of PL and PIX promoters mediates their common enhanced recognition. Perhaps this feature is the string of guanine bases immediately 3′ to their TATA consensus sequences. In any case, manipulation of DNA sequences around early and late promoter sites should permit identification of important features.

ACKNOWLEDGMENTS

We thank K. Berkner for the gifts of all terminal-region recombinants, G. Chu for the *Hin*dIII B recombinant, U. Hansen, F. Laski, R. Kaufman, and S. Woo for communicating results before publication, C. Cepko for fluorescent antibodies and techniques, L. Spencer and S. Huang for technical assistance, and M. Siafaca for preparing the manuscript. We also thank M. Gefter and members of the Sharp and Ziff labs for illuminating discussions and M. Samuels for helpful comments on the manuscript.

This work was supported by grant PCM78-23230 from the National Science Foundation, by Public Health Service grants CA26717 (Program Project Grant) to P.A.S. and GM21779 from the National Institutes of Health, by grant MV75 to E.B.Z. from the American Cancer Society, and by Public Health Service grant GM28983 to J.L.M. from the National Institutes of Health. A.F. is a National Science Foundation pregraduate fellow, and C.C.B. is a National Institutes of Health trainee. This work was partially supported by Public Health Service grant CA14051 from the National Cancer Institute.

LITERATURE CITED

1. **Alestrom, P., G. Akusjarvi, M. Perricaudet, M. Mathews, D. Klessig, and U. Pettersson.** 1980. The gene for polypeptide IX of adenovirus type 2 and its unspliced messenger RNA. Cell **19**:671-681.
2. **Baker, C., J. Herisse, G. Courtois, R. Galibert, and E. Ziff** 1979. Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5′ end. Cell **18**:569-580.
3. **Baker, C., and E. Ziff.** 1980. Biogenesis, structures, and sites of encoding of the 5′ termini of adenovirus 2 mRNAs. Cold Spring Harbor Symp. Quant. Biol. **44**: 415-428.

3a. **Baker, C., and E. Ziff.** 1981. Promoters and heterogeneous 5′ termini of the messenger RNA's of adenovirus serotype 2. J. Mol. Biol. **149**:189-221.

4. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304-310.
5. **Berk, A., F. Lee, T. Harrison, J. Williams, and P. A. Sharp.** 1979. Pre-early Adenovirus 5 gene product regulates synthesis of early viral messenger RNAs. Cell **17**: 935-944.
6. **Brownlee, G.** 1972. Determination of sequences in RNA, p. 1-265. *In* S. Work and E. Work (ed.), Laboratory techniques in biochemistry and molecular biology, 3, part 1. Elsevier-North Holland Publishing Co., Amsterdam.
7. **Chow, L., T. Broker, and J. Lewis.** 1979. Complex splicing patterns of RNA from the early regions of Adenovirus 2 messenger RNA. J. Mol. Biol. **134**:265-303.
8. **Chow, L., J. Lewis, and T. Broker.** 1980. RNA transcription and splicing at early and intermediate times after adenovirus 2 infection. Cold Spring Harbor Symp. Quant. Biol. **44**:401-414.
9. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406-1414.
10. **Dierks, P., A. van Ooyen, N. Mantei, and C. Weissman.** 1981. DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5′ terminus. Proc. Natl. Acad. Sci. U.S.A. **78**:1411-1415.
11. **Fire, A., C. Baker, E. Ziff, and P. A. Sharp.** 1981. Transcription of adenovirus DNA in infected cell extracts, p. 387-400. *In* D. Brown and C. Fox (ed.), Developmental biology using purified genes, ICN-UCLA Symposia, vol. 23. Academic Press, Inc., New York.
12. **Fujinaga, K., Y. Sawada, Y. Uemizu, T. Yamashita, H. Shimojo, K. Shiroki, H. Sugisaki, K. Sugimoto, and M. Takanami.** 1980. Nucleotide sequences, integration, and transcription of the adenovirus 12 transforming genes. Cold Spring Harbor Symp. Quant. Biol. **44**:519-532.
13. **Galibert, F., J. Herisse, and F. Courtois.** 1979. Nucleotide sequence of the Eco RIF fragment of adenovirus 2 genome. Gene **6**:1-22.
14. **Ghosh, P., P. Lebowitz, J. Frisque, and Y. Gluzman.** 1981. Identification of a promoter component involved in positioning the 5′ termini of Simian Virus 40 early mRNAs. Proc. Natl. Acad. Sci. U.S.A. **78**:100-104.
15. **Grosschedl, R., and M. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:1432-1436.
16. **Grosschedl, R., and M. Birnstiel.** 1980. Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:7102-7106.
17. **Gruss, D., R. Dhar, and G. Khory.** 1981. Simian virus 40 random repeated sequences as an element of the early promoters. Proc. Natl. Acad. Sci. U.S.A. **78**:943-947.
18. **Handa, H., R. Kaufman, J. Manley, M. Gefter, and P. A. Sharp.** 1981. Transcription of simian virus 40 DNA in a HeLa whole cell extract. J. Biol. Chem. **256**:478-482.
19. **Hashimoto, S., and M. Green.** 1980. Adenovirus 2 early messenger RNA-genome mapping of 5′ terminal RNase T1 oligonucleotides and heterogeneity of 5′ termini. J. Biol. Chem. **255**:6780-6788.

20. **Hu, S.-L., and J. Manley.** 1981. DNA sequence required for transcription *in vitro* from the major late promoter of adenovirus 2. Proc. Natl. Acad. Sci. U.S.A. **78**:820–824.
21. **Lewis, J., and M. Mathews.** 1980. Control of adenovirus early gene expression: a class of immediate early products. Cell **21**:303–313.
22. **Maat, J., C. P. Van Beveren, and H. Van Ormondt.** 1980. The nucleotide sequence of adenovirus type 5 early region El: the region between map positions 8.0 (Hind III site) and 11.8 (Sma I site). Gene **10**:27–38.
23. **Maat, J., and H. Van Ormondt.** 1979. The nucleotide sequences of the transforming Hind III-G fragment of adenovirus type 5 DNA. The region between positions 4.5 (Hpa I site) and 8.0 (Hind III site). Gene **6**:75–90.
24. **Manley, J., A. Fire, A. Cano, P. A. Sharp, and M. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. U.S.A. **77**:3855–3859.
25. **Manley, J., P. A. Sharp, and M. Gefter.** 1979. RNA synthesis in isolated nuclei: identification and comparison of adenovirus 2 encoded transcripts synthesized *in vitro* and *in vivo*. J. Mol. Biol. **135**:171–197.
26. **Mathis, D., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature (London) **290**:310–315.
27. **McMasters, G., and G. Carmichael.** 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. U.S.A. **79**:4835–4838.
28. **Nevins, J., and M. Wilson.** 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. Nature (London) **290**:113–118.
29. **Rio, D., A. Robbins, R. Myers, and R. Tijan.** 1980. Regulation of simian virus 40 early transcription *in vitro* by a purified tumor antigen. Proc. Natl. Acad. Sci. U.S.A. **77**:5706–5710.
30. **Sambrook, J., R. Greene, J. Stringer, T. Mitchison, S.-L. Hu, and M. Botchan.** 1980. Analysis of the sites of integration of viral DNA sequences in rat cells transformed by adenovirus 2 or SV40. Cold Spring Harbor Symp. Quant. Biol. **44**:569–584.
31. **Sanger, F., G. Brownlee, and B. Barrell.** 1965. A two-dimensional fractionation procedure for radioactive nucleotides. J. Mol. Biol. **131**:373–398.
32. **Shatkin, A.** 1976. Capping of eukaryotic RNAs. Cell **9**: 645–653.
33. **Shaw, A., and E. Ziff.** 1980. Transcripts from the adenovirus-2 major late promoter yield a single family of 3′ coterminal mRNAs during early infection and five families at late times. Cell **22**:905–916.
34. **Steenbergh, P., and J. Sussenbach.** 1979. The nucleotide sequence of the right-hand terminus of adenovirus type 5 DNA: implications for the mechanism of DNA replication. Gene **6**:307–318.
35. **Tooze, J.** 1980. DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Tsai, S., M. J. Tsai, and B. O'Malley.** 1981. Specific 5′ flanking sequences are required for faithful initiation of *in vitro* transcription of the ovalbumin gene. Proc. Natl. Acad. Sci. U.S.A. **78**:879–883.
37. **Van Ormondt, H., J. Maat, A. DeWaasd, and A. van der Eb.** 1978. The nucleotide sequence of the transforming Hpa I E fragment of Ad5 DNA. Gene **4**:309–328.
38. **Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon.** 1980. Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci. U.S.A. **77**: 7024–7028.
39. **Wasylyk, B., C. Kedinger, J. Corden, O. Brison, and P. Chambon.** 1980. Specific *in vitro* initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus 2 early and late genes. Nature (London) **285**:367–373.
40. **Weil, P. A., D. Luse, J. Segall, and R. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484.
41. **Yoshida, K., and K. Fujinaga.** 1980. Unique species of mRNA from adenovirus type 7 early region 1 in cells transformed by adenovirus type 7 DNA fragment. J. Virol. **36**:337–352.
42. **Ziff, E.** 1980. Transcription and RNA processing by the DNA tumor viruses. Nature (London) **287**:491–499.
43. **Ziff, E., and R. Evans.** 1978. Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus-2 major late transcription unit. Cell **15**:1463–1475.

Chapter IV

SEPARATION AND CHARACTERIZATON OF FACTORS MEDIATING ACCURATE TRANSCRIPTION BY RNA POLYMERASE II*

*This Chapter describes the results of an equal collaboration with Mark Samuels

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 257, No. 23, Issue of December 10, pp. 14419-14427, 1982
Printed in U.S.A.

Separation and Characterization of Factors Mediating Accurate Transcription by RNA Polymerase II*

(Received for publication, July 16, 1982)

Mark Samuels‡§, Andrew Fire§, and Phillip A. Sharp¶

From the Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

A whole cell extract of HeLa cells was resolved through two successive chromatographic steps using an extension of the procedure of Matsui *et al.* (Matsui, T., Segall, J., Weil, P. A., and Roeder, R. G. (1980) *J. Biol. Chem.* 255, 11992-11996). RNA polymerase II and three of the resulting fractions were necessary and sufficient for accurate transcription of the adenovirus major late promoter. This accurate transcription was quantitated as a function of each of the required fractions, polymerase, and DNA. A linear range of response was observed in each case. Using the linear ranges for assay, it was possible to calculate net purifications and yields for each of the required transcriptional activities after chromatography. These activities were each shown to sediment with a distinct peak on sucrose gradients. The effects of variations in salt concentration, magnesium concentration, temperature, and reaction time were determined. High resolution analysis of runoff transcripts showed that the reconstituted system initiated transcription precisely at the adenovirus major late and early region IV promoters.

Regulation at the level of transcription is the predominant means by which prokaryotic cells control gene expression. In order to understand the molecular basis of transcriptional regulation it was necessary to elucidate the mechanisms by which RNA polymerase recognizes and transcribes a given DNA segment. These studies have proven difficult to duplicate in eukaryotic systems. The enzyme catalyzing transcription of eukaryotic messenger RNA *in vivo*, RNA polymerase II, has been identified and purified (1, 2). In addition, recombinant DNA technology has provided a variety of templates which have well defined transcriptional properties *in vivo*. Unfortunately, purified RNA polymerase II does not accurately initiate transcription on any of these templates. It has been demonstrated, however, that crude extracts of cells can produce *in vitro* the precise 5′ termini seen *in vivo* (3, 4). Matsui *et al.* (5) showed that such a crude lysate contained multiple fractions that were required to direct accurate transcription by RNA polymerase II. A number of other groups have repeated these observations using different fractionation protocols (6, 7).

The requirement for several factors in addition to the polymerase suggests an analogy to the complex process of DNA replication in bacteria. A combination of genetics and biochemistry was used to analyze the replication apparatus and ultimately to purify many of the components involved in the reaction (8). The limitations on genetics in mammalian systems leave only the straightforward enzymological approach to purification of the relevant factors. As a prerequisite to purification and characterization in the RNA polymerase II system, it will be necessary to construct a framework in which the activity in each of the fractions can be assayed individually.

Such a framework entails two criteria: first, a reproducible procedure to cleanly separate fractions from each other in order that one can use a mixture devoid of a single fraction as an assay for the activity in that fraction; second, an understanding of the dose-response relationship for each of the fractions—in the best possible case, one could work in the linear range of a titration curve. These criteria are sufficient to allow one to monitor quantitatively the purification of each fraction. For further characterization at this stage it is very useful to understand the response of the reconstituted system to the biochemical parameters of the transcription reaction: substrate concentrations, solution conditions, time, and temperature.

EXPERIMENTAL PROCEDURES

Materials

Unlabeled ribonucleoside triphosphates purified by high pressure liquid chromatography, and [α-^{32}P]UTP (450 Ci/mmol) were purchased from ICN. α-Amanitin was purchased from Calbiochem, crystallized BSA[1] from Miles Biochemicals, creatine phosphokinase from Sigma, and purified human placental ribonuclease inhibitor was purchased from Biotec. The polyacrylamide gel silver stain kit and protein assay dye reagent were purchased from Bio-Rad.

Methods

Buffer A contained 20 mM Hepes-NaOH, pH 7.9, 20% glycerol, 1 mM EDTA, 1 mM dithiothreitol. Buffer B contained 20 mM Hepes-NaOH, pH 7.9, 17% glycerol, 1 mM EDTA, 1 mM dithiothreitol, 12.5 mM $MgCl_2$. Buffer C contained 20 mM Hepes-NaOH, pH 7.9, 20% glycerol, 1 mM EDTA, 1 mM dithiothreitol, 5 mM $MgCl_2$. Buffer D contained 20 mM Hepes-NaOH, pH 7.9, 5% glycerol, 1 mM EDTA, 1 mM dithiothreitol.

Preparation of Ion Exchange Resins—Phosphocellulose (Whatman P11) was extensively washed according to the manufacturer's instructions, and was subsequently equilibrated in buffer A + 0.1 M

* These studies were supported by National Science Foundation Grant PCM 7823230 (currently PCM 8200309) and National Institutes of Health Grant CA 26717 and in part by National Institutes of Health Center for Cancer Biology at MIT Grant CA14051. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by a predoctoral fellowship from the National Science Foundation.

§ This work resulted from an equal contribution by the first two authors.

¶ To whom reprint requests should be addressed.

[1] The abbreviations used are: BSA, bovine serum albumin; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; EIV, adenovirus early region four.

KCl. DEAE-Sephacel (Pharmacia) was equilibrated directly with buffer A + 0.1 M KCl. DNA-cellulose was prepared essentially according to Alberts and Herrick (9). Cellex 410 (Bio-Rad) was extensively washed with boiling ethanol, then successively with H_2O, 0.1 M NaOH 1 mM EDTA, H_2O, 10 mM HCl, and H_2O until its pH reached neutrality. The washed cellulose was mixed with denatured herring sperm DNA, dried, and washed extensively to yield a resin with 1.4 mg of DNA/ml of cellulose. DNA-cellulose was equilibrated with buffer A + 0.1 M KCl and stored at 4 °C. The resin was not used more than twice.

Determination of Protein Concentration—The protein concentration was measured according to the Coomassie blue binding method of Bradford (10).

Cells and Preparation of Extract—Whole cell extracts from HeLa cells were prepared according to Manley *et al.* (4). Routine preparations began with 50 liters of suspension cultures. The resulting extract (about 100 ml) was dialyzed against 2 × 20 volumes of buffer B + 0.1 M KCl for a total of 24 h. Following clarification at 10,000 × *g* for 20 min, supernatants were further dialyzed against 2 × 20 volumes of buffer A + 0.1 M KCl for a total of 24 h. The whole cell extracts had protein concentrations of approximately 20 mg/ml. Extracts were quick-frozen in liquid nitrogen and stored at −80 °C.

Ion Exchange Chromatography—The preparation of transcriptional fractions routinely began with 100 ml of whole cell extract (2,000 mg of protein). The chromatographic procedure is summarized in Fig. 1.

A phosphocellulose column washed with three volumes of buffer A + 0.1 M KCl + 0.2 mg/ml of BSA and three volumes of buffer A + 0.1 M KCl was equilibrated with buffer A + 0.04 M KCl. Whole cell extract was thawed, diluted with buffer A to a final KCl concentration of 0.04 M, and applied to the phosphocellulose column (8 mg of protein/ml of bed volume). After washing with one column volume of buffer A + 0.04 M KCl, three successive step elutions were performed with 0.35 M KCl (five column volumes), 0.6 M KCl (three column volumes), and 1 M KCl + 0.2 mg/ml of BSA (two column volumes) in buffer A. Aliquots of the eluate were assayed for protein; 65–90% of the total protein that eluted in the flowthrough and in each step was combined as fractions [A], [B], [C], and [D]. Fractions [B], [C], and [D] were dialyzed against buffer A + 0.1 M KCl.

A column of DEAE-Sephacel was washed with three column volumes each of buffer A + 0.1 M KCl + 0.2 mg/ml of BSA and buffer A + 0.1 M KCl, and was equilibrated with buffer A + 0.15 M KCl. Fraction [A] was adjusted to 0.15 M KCl and applied to the column at 8.5 mg of protein/ml of bed volume. The column was washed with A + 0.15 M KCl, and was eluted with two steps of 0.35 M KCl and 1 M KCl in buffer A (three column volumes each). 65–90% of the protein that eluted in the flowthrough and in each step was combined as fractions [AA], [AB], and [AC].

A column of single stranded DNA-cellulose was washed with three volumes of buffer C + 0.1 M KCl + 0.2 mg/ml of BSA and was equilibrated with buffer C + 0.1 M KCl. Dialyzed fraction [C] was adjusted to 5 mM $MgCl_2$ and applied to the column (2 mg of protein/ml of bed volume). The column was washed with three volumes each of 0.1 M KCl, 0.3 M KCl, 0.6 M KCl, and 1 M KCl in buffer C. In some preparations, 0.2 mg/ml of BSA was included in column elution buffers with no apparent effect. Of the protein eluted in the flowthrough, 0.3 M KCl, and 0.6 M KCl steps, 50–100% was combined as fractions [CA], [CB], and [CC], respectively. Fraction [CD] consisted of those fractions encompassing the increase in conductivity from 0.6 M KCl to 1 M KCl.

A DEAE-Sephacel column was washed successively with three volumes each of buffer A + 0.1 M KCl + 0.2 mg/ml of BSA and buffer A + 0.1 M KCl, and was equilibrated with buffer A + 0.05 M KCl.

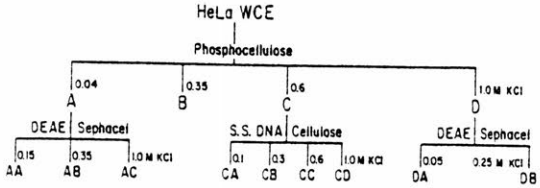


FIG. 1. **Scheme for the resolution of transcription factors contained in solubilized HeLa cell extracts.** The details of chromatography are described under "Materials and Methods."

Dialyzed fraction [D], diluted to 0.05 M KCl with buffer A, was applied to the column (4 mg of protein/ml of bed volume). The column was washed with 0.05 M KCl (five column volumes) and 0.25 M KCl (two column volumes) in buffer A. The 0.05 M KCl wash contained no measurable protein; the flowthrough as determined by volume measurement was pooled as fraction [DA]. Of the 0.25 M KCl eluate, 60–75% of the protein was combined as fraction [DB].

Chromatographic fractions were dialyzed against buffer A + 0.1 M KCl, quick-frozen in small aliquots, and stored at −80 °C; fractions [A], [AA], [CA], and [DA] were frozen directly. No significant losses in activity were observed over a three-month period, or with one cycle of thawing and freezing.

RNA Polymerase II—Polymerase was prepared from fresh calf thymus according to Hodo and Blatti (11), and from HeLa cells as described (12). Except where otherwise noted, all transcriptions were performed using the most highly purified (glycerol gradient) enzyme. The activity was measured by incorporation of labeled nucleotide into trichloroacetic acid precipitable material, using denatured salmon sperm DNA as a template. All incorporation by the purified enzyme was sensitive to α-amanitin at 0.5 μg/ml. One unit represented 1 pmol of UTP incorporated in 20 min at 37 °C. The glycerol gradient enzyme had a specific activity of 1.1×10^5 units/mg of protein, and was diluted to 20 units/μl with buffer A + 0.1 M KCl before use.

Sucrose Gradient Analysis—Aliquots of fractions [AB], [CB], and [DB] were dialyzed against buffer D + 0.1 M KCl. For centrifugation in a Beckman SW 41 rotor, 0.7 ml of sample was applied to 11-ml linear gradients of 5–20% sucrose in buffer D + 0.1 M KCl. Gradients were centrifuged at 4 °C for 12 to 24 h at 40,000 rpm. 15–20 fractions were collected from the bottom of each gradient. Standards of hemoglobin, catalase, and in some cases 18 S ribosomal RNA (applied as total HeLa cell cytoplasmic RNA) were sedimented in parallel gradients; hemoglobin and catalase were assayed by measuring protein concentration or absorbance at 450 nm, 18 S RNA by absorbance at 260 nm.

In Vitro Synthesis and Analysis of RNA—Analytical reactions were usually performed in 20 μl, and had final concentrations of 12 mM Hepes-NaOH, pH 7.9, 12% glycerol, 1 mM EDTA, 0.6 mM dithiothreitol, 60 mM KCl, 5 mM $MgCl_2$, 5 mM creatine phosphate, and 0.2 mg/ml of creatine phosphokinase. Nucleotide concentrations were 60 μM ATP, GTP, and CTP, 10 μM [α-^{32}P]UTP (10 Ci/mmol). Concentrations of extract, protein fractions, and DNA were as noted in figure legends. Unless otherwise noted, reactions were performed at 30 °C for 90 min.

Workup of the RNA products and agarose gel electrophoresis followed a standard protocol (13), with the following exceptions. Transcriptions performed with ion exchange column fractions received only one extraction with phenol/chloroform/isoamyl alcohol. For the assay of sucrose gradient fractions, reactions were performed in 10 μl and stopped by the addition of 10 μl of 17 mM sodium phosphate, pH 6.8, 15 mM EDTA, 2% sodium dodecyl sulfate, 2.5 mg/ml of tRNA. After the addition of 50 μl of glyoxal mix (70% dimethyl sulfoxide, 1.4 M deionized glyoxal, 10 mM sodium phosphate, pH 6.8, 1 mM EDTA, 0.01% bromophenol blue) and incubation at 50 °C for 60 min, agarose gel electrophoresis was performed in 10 mM sodium phosphate, 1 mM EDTA as described (13, 14). For the analysis of very short runoff transcripts, reaction products were extracted as usual, precipitated twice with ethanol, resuspended in 10 μl of 80% deionized formamide, 0.1 M Tris-borate, pH 8.3, 2.5 mM EDTA, and electrophoresed in 0.2-mm thick 8% polyacrylamide-urea gels (15).

Quantitation of Specific Transcription—XAR (Kodak) film was pre-flashed using an electronic photographic flash with an orange filter so that the background absorbance was 0.15 at 540 nm after developing, relative to untrated film (16). After the autoradiography of dried agarose gels, the developed films were scanned using a Zeineh soft laser scanning densitometer (Biomed Instruments). Baseline ranges were drawn on duplicate scans, and peak areas were integrated with a model 246-117 Numonics electronics graphics calculator. Reconstructions employing serial dilutions of a standard transcription reaction product confirmed that the signal as measured by this method was proportional to the input radioactivity in the specific runoff transcript. Because the actual peak area measurement depended upon the time of autoradiography and upon the densitometer gain setting, ordinate values were normalized to a value of 10 for an analytical reaction under standard conditions employing 0.12 μg of template DNA, 4.4 units of RNA polymerase II, 0.5 μl of fraction [AB], 3 μl of fraction [CB], 3 μl of fraction [CD], and 3 μl of fraction [DB].

RESULTS

Separation of Transcription Components—Because RNA polymerase II, either in crude form or supplemented with a crude lysate, does not specifically terminate transcription *in vitro*, the standard assay for specific transcription uses a template DNA cleaved at a unique site downstream from a known promoter (3, 4). Incorporation of labeled ribonucleoside triphosphates into RNA whose length corresponds to that of the expected "runoff" transcript indicates that accurate transcription has occurred.

A crude whole cell extract of HeLa cells provided the starting material for partial purification of the transcription factors. Fig. 1 summarizes the separation protocol. The use of phosphocellulose chromatography for the initial separation was suggested by the success of the procedure with an "S-100" extract described by Matsui *et al.* (5). Indeed, the behavior of the whole cell extract transcriptional activity on phosphocellulose appeared identical with that reported for the S-100. Removal of magnesium from the whole cell extract by a second dialysis step was found to be essential for the recovery of activity.

Four fractions were recovered from the phosphocellulose chromatography: [A] (flowthrough at 0.04 M KCl), [B] (eluate of 0.35 M KCl wash), [C] (eluate of 0.6 M KCl wash), and [D] (eluate of 1.0 M KCl wash). The capacity of these factors for accurate transcription was tested by the runoff assay, using a pBR322-adenovirus recombinant (Fig. 2*A*) carrying the adenovirus major late promoter (for a review of adenovirus transcription see Ref. 17). Initiation at the promoter with subsequent elongation to the *Pst*I cleavage site would generate a 974 nucleotide band. As a control, the α-amanitin-sensitive transcription of this template in the whole cell extract is shown in Fig. 2*B*, *lanes 1* and *2*.

The phosphocellulose fractions [A], [C], and [D], together with RNA polymerase II, reconstituted accurate transcription (Fig. 2*B*, *lane 10*). When only one or two of these three phosphocellulose fractions was mixed with polymerase, no accurate transcription was observed (Fig. 2*B*, *lanes 3–9*). In addition, the accurate transcription observed upon reconstitution of phosphocellulose fractions was completely inhibited by the addition of α-amanitin to 0.5 μg/ml (Fig. 2*B*, *lane 12*), indicating that the synthesis was indeed catalyzed by RNA polymerase II (1, 2). The omission of exogenous polymerase significantly reduced, but did not eliminate, specific transcription (Fig. 2*B*, *lane 13*). The residual activity was due to endogenous polymerase detectable in fraction [C] (data not shown). Specific transcription did not require fraction [B]. Low concentrations of [B] had little effect on activity (Fig. 2*B*, *lane 11*), while high concentrations were inhibitory (data not shown).

Each of the first column fractions was further chromatographed to obtain more concentrated reagents with fewer impurities. Fraction [A] was chromatographed on DEAE-Sephacel; a flowthrough at 0.15 M KCl (fraction [AA]), and two subsequent eluates at 0.35 M KCl (fraction [AB]) and 1.0 M KCl (fraction [AC]) were collected. These fractions were tested for their ability to replace [A] in the transcription reaction. *Lanes 1–3* of Fig. 2*C* show that [AB] contained the activity required for accurate transcription.

Matsui et al. (5) further fractionated a 0.35–0.6 M KCl wash of phosphocellulose by chromatography on DEAE- and DNA-cellulose resins. We have used a similar procedure, omitting the DEAE fractionation because of poor yields. Thus, fraction [C] was chromatographed on single stranded DNA-cellulose; a flowthrough at 0.1 M KCl (fraction [CA]), and three subsequent eluates at 0.3 M KCl (fraction [CB]), 0.6 M KCl (fraction [CC]), and 1.0 M KCl (fraction [CD]) were collected. Of these

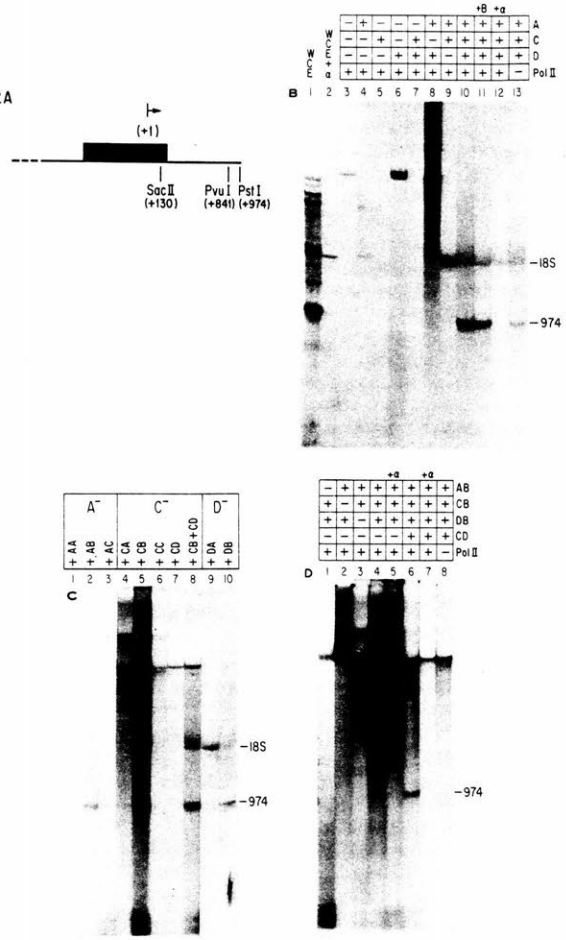


FIG. 2. **Identification of fractions required for transcription.** Analytical transcriptions were performed under standard conditions (see "Materials and Methods"). All reactions received 0.15 μg of pFLBH digested with *Pst*I (*A*). Reaction products were analyzed by agarose gel electrophoresis. An *arrow* indicates the position of the 974 nucleotide runoff transcript; an *arrow* at 18 S indicates the position of radioactively labeled ribosomal RNA. *A*, the pFLBH recombinant used for transcription. Adenovirus 2 sequences (14.7–17.0 map units), denoted by a *solid bar*, were inserted between the *Bam*HI and *Hin*dIII sites in pBR322. The adenovirus major late promoter, denoted by an *arrow*, defines position +1 of the plasmid; relevant restriction enzyme sites are shown. This plasmid was a kind gift of J. Manley and F. Laski. *B*, reconstitution of first column fractions. 9 μl of whole cell extract, 2 μl of phosphocellulose fraction [A], 1 μl of [B], 4 μl of [C], 4 μl of [D], and 4.4 units of RNA polymerase II were combined as indicated in the figure. 0.5 μg of poly{d(I-C)}:poly{d(I-C)} (18) was added as carrier DNA to reactions 1 and 2. α-Amanitin was added to reactions 2 and 12 to 0.5 μg/ml. *C*, substitution of second column fractions. *Lanes 1–3*, 1 μl of DEAE-Sephacel fractions [AA], [AB], or [AC] was added as noted in the figure to an assay mix containing 4.4 units of polymerase, 4 μl of [C], and 4 μl of [D]. *Lanes 4–8*, 2.5 μl of DNA-cellulose fractions [CA], [CB], [CC], or 3.5 μl of [CD], were added to an assay mix containing 4.4 units of polymerase, 2 μl of [A], and 4 μl of [D]. *Lanes 9* and *10*, 2.5 μl of DEAE-Sephacel fractions [DA] or [DB] were added to an assay mix containing 4.4 units of polymerase, 2 μl of [A], and 4 μl of [C]. *D*, reconstitution of second column fractions. 4.4 units of polymerase, 0.5 μl of [AB], 2.5 μl of [CB], 3.5 μl of [CD], and 2.5 μl of [DB] were combined as noted in the figure. Reactions 5 and 7 received α-amanitin to 0.5 μg/ml.

four fractions, only fraction [CB] substituted for fraction [C] in reconstituting accurate transcription (Fig. 2*C*, *lanes 4–7*). Fraction [CD], and to a lesser extent [CC], reduced the general background incorporation without producing accurate transcription. Indeed, the high background of nonspecific incorporation observed when fraction [C] was replaced by [CB] was strikingly reduced if [CD] was also included in the reaction (Fig. 2*C*, *lane 8*). Fraction [CD] contained one major polypeptide of M_r = 110,000 (90% of the protein by sodium dodecyl sulfate-polyacrylamide gel electrophoretic analysis). We have purified this protein to homogeneity by a procedure similar to that of Slattery *et al.*[2] The pure protein retains the ability to suppress background incorporation.

Chromatography of fraction [D] on DEAE-Sephacel was performed to generate two fractions, a flowthrough at 0.05 M KCl (fraction [DA]) and a 0.25 M KCl eluate (fraction [DB]). Of the two fractions, only [DB] substituted for fraction [D] in reconstituting accurate transcription (Fig. 2*C*, *lanes 9* and *10*).

The second column fractions [AB], [CB], and [DB] reconstituted accurate transcription with exogenous RNA polymerase II in a mutually dependent reaction which was sensitive to low concentrations of α-amanitin (Fig. 2*D*, *lanes 1–4*, *7*, and *8*). Again, the high level of background incorporation was suppressed by the addition of fraction [CD], or of the purified M_r = 110,000 protein, while the level of specific transcription remained unchanged. Therefore, either [CD] or purified M_r = 110,000 protein was routinely included in reactions using second column fractions.

Occasionally, the omission of [AB] resulted in the disappearance of the nonspecific background as well as the promoter runoff RNA (Fig. 2*D*, *lane 1*). In these cases, a smear of low molecular weight material was observed, suggesting that a ribonuclease might be present. When a commercial preparation of pure human placental ribonuclease inhibitor (19) was added, the nonspecific background and certain high molecular weight labeled bands were restored. However, this addition did not restore the specific runoff transcript (data not shown).

The labeling of 18 S ribosomal RNA, observed with whole cell extracts and first column fractions, did not occur when second column fractions were reconstituted. This resulted from the removal of ribosomal RNA during chromatography, and from the concentration of an end-labeling inhibitor in [DB] (data not shown).

Second Column Fractions Initiate Transcription Precisely—The reconstituted fractions produced runoff RNAs of the appropriate length when pFLBH (see Fig. 2*A*) cleaved with *Pvu*I (+841), *Pst*I (+974), or *Tth*111I (+2370) was used as template. These products were all resolved on agarose gels, with a maximum resolution of 25 nucleotides. To examine the reconstitution reaction at high resolution, very short runoff RNAs were generated. Cleavage of the late promoter recombinant pFLBH with *Sac*II (130 bases downstream from the promoter; Ref. 20), and of the adenovirus EIV recombinant pECORIB5 with *Sma*I (250 bases downstream from the promoter; Refs. 21, 22), generated templates for these experiments. We have previously shown that the whole cell extract yields 5′ termini identical with those observed *in vivo* from these two promoters (4, 23).

Both the whole cell extract and the reconstituted system synthesized late promoter runoff RNA of the correct length from the *Sac*II-truncated template (Fig. 3*A*, *lanes 5* and *7*); bands migrating at 135 nucleotides were observed in each case. (The discrepancy between the measured and predicted molecular weights of the RNA probably resulted from the use of DNA size markers in this gel.) Moreover, a comparison of the reaction products from the EIV template reveals that the whole cell extract and the reconstituted system generated an identical cluster of RNA transcripts of the correct length (Fig. 3*B*, *lanes 1* and *3*). In all cases, the synthesis of the runoff RNAs was inhibited by α-amanitin at 0.5 μg/ml (Fig. 3*A*, *lanes 6* and *8*; Fig. 3*B*, *lanes 2* and *4*). These experiments confirm that the reconstituted system retained the specificity for initiation present in the whole cell extract.

Initiation at the EIV promoter *in vivo* and in the whole cell extract is heterogenous over a 6 base range (21, 23, 24). This heterogeneity may account for some of the complexity observed in the runoff transcript. The whole cell extract is known to initiate very precisely at a single residue on the late promoter (4), hence it was surprising to observe a multiplet of bands in this assay (Fig. 3*A*, *lane 5*). The apparent heterogeneity of a few nucleotides could be due to variability in the 3′ end of the transcript, or to variable capped and uncapped

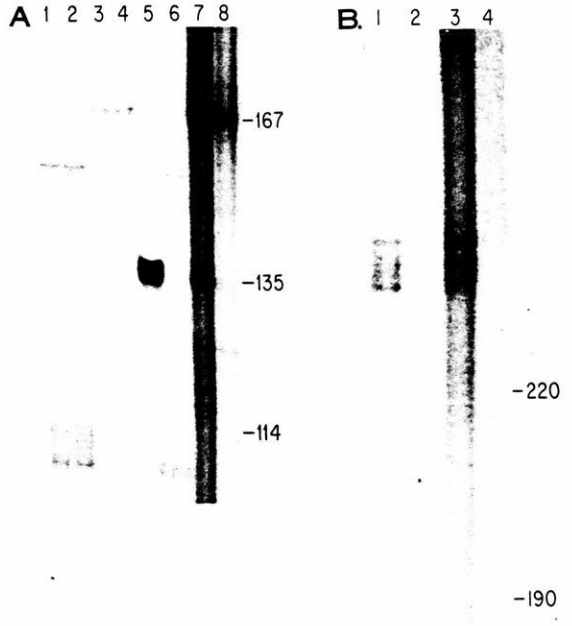


FIG. 3. **Short runoff analysis.** Analytical transcriptions were performed under standard conditions. Reactions with whole cell extract received 10 μCi of [α-^{32}P]UTP, 10 μl of whole cell extract, and 0.5 μg of poly{d(I-C)}:poly{d(I-C)}. Reactions with the reconstituted system received 20 μCi of [α-^{32}P]UTP, 4.4 units of polymerase, 0.5 μl of [AB], 4 μl of [CB], 4 μl of [CD], and 3 μl of [DB]. Reaction products were analyzed on denaturing polyacrylamide gels. The *arrows* indicate the positions of DNA size markers. *A*, reactions 1–4 received no template DNA; reactions 5–8 received 0.12 μg of pFLBH digested with *Sac*II (+130 nucleotides). α-Amanitin was added where noted to 0.5 μg/ml. *Lane 1*, whole cell extract. *Lane 2*, whole cell extract + α-amanitin. *Lane 3*, reconstituted system. *Lane 4*, reconstituted system + α-amanitin. *Lane 5*, whole cell extract. *Lane 6*, whole cell extract + α-amanitin. *Lane 7*, reconstituted system. *Lane 8*, reconstituted system + α-amanitin. *B*, all reactions received 0.12 μg of pECORIB5 digested with *Sma*I (+250 nucleotides). This plasmid (a kind gift of Kathleen Berkner) is an adenovirus-pBR322 recombinant with the right terminal 16% of adenovirus 5 inserted into the *Eco*RI site of pBR322 (17). The whole cell extract was used in reactions 1 and 2; the reconstituted system was used in reactions 3 and 4. α-Amanitin was added to reactions 2 and 4 to 0.5 μg/ml.

[2] E. Slattery, D. Dingnam, T. Matsui, and R. G. Roeder, personal communication.

structures at the 5′ end (23). In either case, the reconstituted system did not show this heterogeneity.

Both the whole cell extract and the second column reconstitutions produced some template-independent products, whose synthesis was totally resistant to α-amanitin at a concentration of 0.5 μg/ml (Fig. 3*A*, *lanes 1–4*).

The Transcription Components Are Cleanly Resolved—The extent of dependence on each fraction was determined by quantitating accurate transcription in the presence and absence of that fraction. When no signal was detected in the absence of a fraction, reconstructions set a lower limit on the dependence. Reconstructions were performed by serially diluting products from the complete reaction with products from the reaction omitting a particular fraction. The dependence was determined from the highest dilution at which the runoff RNA was detectable. Table I shows that the reaction was stimulated at least 100-fold by fractions [AB] and [CB]; no accurate transcription was detected in the absence of these fractions. RNA polymerase II and [DB] both increased the signal 20-fold above a very low but detectable basal level.

The Transcription Components Are Heat Labile—In addition to much of the soluble cell protein, the whole cell extract contains various amounts of RNA, nucleotides and cofactors, DNA, and possibly other cellular components. Each of the second column fractions was, therefore, tested for its sensitivity to treatment at 60 °C for 8 min. Inactivation is taken as evidence (though by no means conclusive) that the active agent in a fraction has a protein component. As shown in Table II, the transcriptional activities of all of the HeLa cell fractions and the polymerase were significantly reduced by the heating step. Fractions [CB] and [DB] were inactivated to the limit of the assay. Fraction [AB] retained 30% of its activity, suggesting the presence of a heat-resistant component.

Dose-Response Curves for Transcription Factors—Purification of an enzyme requires some reliable quantitation of activity. To assay a given transcription component, a specific assay mix was constructed which lacked only that component. These reaction mixes utilized purified polymerase and the second column fractions [AB], [CB], and [DB]. Titrations of each of the protein fractions were performed in their respective assays. In titrations of [AB], commercial ribonuclease inhibitor was added to substitute for the HeLa cell RNAase inhibitor present in that fraction. The use of densitometry allowed the radioactivity in promoter-specific runoff RNA to be accurately quantitated from autoradiographs of the agarose gels, as nonspecific background could be subtracted from the specific signal.

TABLE I

Dependence on individual fractions

Assays were performed with reaction conditions and protein fraction concentrations as described under "Methods;" the indicated fractions were omitted from the reactions. The template DNA consisted of an equimolar mixture of pFLBH digested with either *Pvu*I (+841 nucleotides) or *Tth*111I (+2370 nucleotides); 0.12 μg of the mixture was used in each assay. Reaction products were analyzed and in all experiments the amount of runoff RNA in the *Pvu*I transcript was quantitated, as described under "Methods."

Fraction omitted	Specific incorporation
	%
None	100
[AB][a]	<2[b]
[CB][c]	<1[b]
[DB][d]	5.4
RNA polymerase II	8.7

[a] The control reaction received 1 μl of [AB]. The addition of 22 units of placental ribonuclease inhibitor had no effect on the control reaction, but shifted the background incorporation to higher molecular weight material (without restoring the specific signal) in the absence of [AB].

[b] Determined by reconstruction.

[c] The control reaction received 3 μl of [CB].

[d] The control reaction received 2.2 μl of [DB].

TABLE II

Heat lability of individual fractions

Assays were performed with reaction conditions and protein concentrations as described under "Methods." Template DNA and quantitation were as in Table I. The indicated fractions were heated for 8 min at 60 °C, then cooled on ice prior to assay.

Fraction heat-inactivated	Specific incorporation
	%
None	100
[AB][a]	30
[CB]	<1[b]
[DB][c]	5.5
RNA polymerase II	12

[a] Reactions received 1 μl of untreated or heat-inactivated [AB].

[b] Determined by reconstruction. The control reaction received 3 μl of [CB].

[c] Reactions received 2.2 μl of untreated or heat-inactivated [DB].

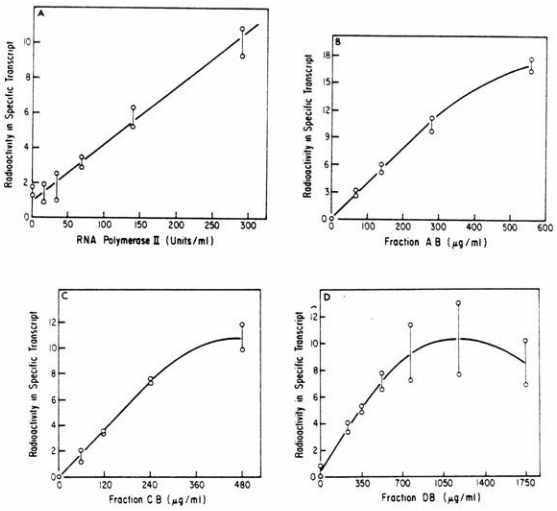


FIG. 4. **Dose-response curves for transcription components.** Reactions were performed under standard conditions with second column fractions. The template DNA and quantitation were as in Table I. All points were measured in duplicate transcription reactions. *A*, titration of RNA polymerase II. 5.6, 2.8, 1.4, 0.7, 0.35, or 0 units of polymerase were added to an assay mix containing 0.06 μg of template DNA, 0.5 μl of [AB], 3 μl of [CB], 3 μl of [CD], and 2.5 μl of [DB]. The normalized amount of specific transcript was plotted against the final RNA polymerase II concentration. *B*, titration of [AB]. 1.3, 0.65, 0.32, 0.16, or 0 μl of [AB] was added to an assay mix containing 0.12 μg of template DNA, 4.4 units of polymerase, 3 μl of [CB], 3 μl of [CD], 2.6 μl of [DB], and 11 units of placental ribonuclease inhibitor. The normalized amount of specific transcript was plotted against the final concentration of [AB] protein. *C*, titration of [CB]. 4, 2, 1, 0.5, or 0 μl of [CB] was added to an assay mix containing 0.12 μg of template DNA, 4.4 units of polymerase, 0.5 μl of [AB], 3 μl of [CD], and 3 μl of [DB]. The normalized amount of specific transcript was plotted against the final concentration of [CB] protein. *D*, titration of [DB]. 5, 3.3, 2.2, 1.5, 1, 0.67, or 0 μl of [DB] was added to an assay mix containing 0.12 μl of template DNA, 4.4 units of polymerase, 0.5 μl of [AB], 3 μl of [CB], and 3 μl of [CD]. The normalized amount of specific transcript was plotted against the final concentration of [DB] protein (including BSA from the chromatographic elution).

When RNA polymerase II was titrated, accurate transcription increased linearly above the nonzero basal level (Fig. 4*A*). At the highest concentration tested in this experiment, a 10-fold stimulation was observed. In more extensive titrations, saturation was observed with a maximal stimulation of 20-fold above the basal level (data not shown). Likewise, titrations of [AB], [CB], and [DB] each showed a linear range in response to concentration (Fig. 4, *B–D*). At concentrations of [AB] and [DB] higher than those shown in Fig. 4, accurate transcription was inhibited (data not shown). High concentrations of [CB] apparently saturated the system without inhibition.

The phosphocellulose column fractions [A], [C], and [D] also showed linear ranges in response to protein concentration. Because the absolute amount of runoff transcription depended simultaneously on the concentrations of several components, the activity of an individual component had to be defined relative to some standard state. Thus, the assignment of units was somewhat arbitrary. Nonetheless, the activities of the second column fractions could be measured with respect to those of the equivalent first column fractions. From the measurements of activity and protein concentration, the net purifications and yields were calculated for the second round of chromatographies (see Table III).

TABLE III

Purification of second column fractions

Titrations of first and second column fractions were performed as described in the text and as in Fig. 4. From the slopes of the activity curves at low concentrations of the fractions, and from the results of protein assays, specific activities in units per mg of protein (see "Methods") and total activities in units were calculated. The yields and purifications were determined relative to first column fractions.

Fraction	Total protein	Volume	Yield	Purification
	mg	*ml*	*%*	*–fold*
Whole cell extract	2100	98		
[A]	710	245	(100)[a]	(1)
[AB]	180	20	45	1.7
[C]	115	180	(100)	(1)
[CB]	22	10	15	0.77
[D]	85	145	(100)	(1)
[DB]	<9[b]	6.2	25	>2.5[b]

[a] Numbers in parentheses are defined values.

[b] The amount of BSA added to [DB] during chromatography was estimated at >80% of the total protein in [DB] by gel analysis, and was discounted in calculating the specific activity.

Sucrose Gradient Analysis of Transcription Fractions—As a preliminary step in the physical characterization of transcription factors, the second column fractions were each sedimented through a sucrose gradient. The transcriptional activity present in [AB] migrated near the top of the gradient, as a discrete peak at 3.5 S (Fig. 5*A*, average of three determinations); the ribonuclease inhibitor also present in [AB] appeared as a broad peak in the upper part of the gradient. An estimated 50–75% of the activity in [AB] was recovered in the gradient peak. The activity in [CB] was recovered as a peak at 5 S (Fig. 5*B*, average of four determinations), with estimated yields of 10–40%.

The sedimentation data for fraction [DB] was somewhat more complex. As seen in Fig. 5*C*, activity above the basal level was recovered across much of the gradient, with a reproducible peak at 17 S (average of five determinations). The activity in the peak was approximately 5-fold higher than in other fractions. Of the activity applied to the gradient, approximately 20–40% was present in the peak at 17 S.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis showed the presence of a large number of polypeptides in the second column fractions (Fig. 6, *lanes AB*, *CB*, and *DB*). The 3.5 S gradient fraction of [AB] contained most of the polypeptides present in [AB]. The 5 S gradient fraction of [CB] contained approximately 20% of the polypeptides present in [CB] (see Fig. 6).

All of the major polypeptides observed in [DB] migrated near the top of the sucrose gradient; this included a number of small polypeptides prominent in [DB]. To detect polypeptides around the 17 S peak in the gradient, it was necessary to

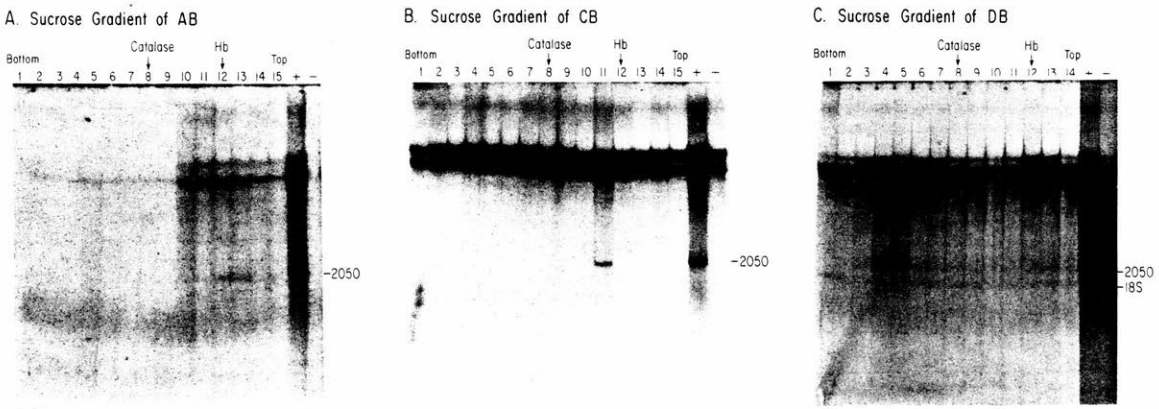


FIG. 5. **Sucrose gradient analyses of second column fractions.** Aliquots of fractions [AB], [CB], and [DB] were sedimented through 5–20% sucrose gradients for 18 h at 36,000 rpm in an SW 41 rotor. The positions of hemoglobin (*Hb*, 4.2 S) and catalase (11.2 S) in parallel marker gradients are noted. Gradient fractions were assayed for transcription activity. 10-μl reactions received 2.5 units of purified HeLa cell RNA polymerase II and 0.1 μg of pBal E digested with *Eco*RI. This plasmid contains map units 14.7–21.5 of adenovirus 2 inserted into the *Bam*HI site of pBR322 (4). Cleavage with *Eco*RI should generate a 2050 nucleotide runoff. An *arrow* indicates the position of the specific runoff transcript. *A*, gradient analysis of [AB]. 2 μl of each fraction were assayed in the presence of polymerase, 1.5 μl of [CB], and 1.5 μl of [DB]. The (–) reaction received no additional protein; the (+) reaction received 2 μl of [AB] dialyzed against buffer D. *B*, gradient analysis of [CB]. 2 μl of each fraction were assayed in the presence of polymerase, 1.5 μl of [AB], and 1.5 μl of [DB]. The (–) reaction received no additional protein; the (+) reaction received 2 μl of [CB] dialyzed against buffer D. *C*, gradient analysis of [DB]. 2 μl of each fraction were assayed in the presence of polymerase, 1.5 μl of [AB], and 1.5 μl of [CB]. The (–) reaction received no additional protein; the (+) reaction received 2 μl of [DB] dialyzed against buffer D. The *arrow* at 18 S indicates the position of radioactively labeled ribosomal RNA.

concentrate the fractions by acid precipitation. Even in this concentrated material, only the very sensitive silver stain visualized proteins. A number of high molecular weight polypeptides were thus observed across the bottom half of the gradient (see Fig. 6, *lane cDB 17 S*).

Characterization of Transcription in the Reconstituted System—The dependence of accurate transcription using second column fractions on a variety of other reaction parameters was tested (Table IV). The reaction depended completely on exogenous template and on magnesium. No specific transcription was observed in the absence of GTP, CTP, or UTP (dependence on ATP could not be tested due to its presence as a contaminant in commercial preparations of radioactive nucleotides). Transcription in the whole cell extract is in general not dependent on addition of exogenous GTP, CTP, and UTP, due to endogenous nucleotides which do not readily dialyze (23). The reconstituted system was not stimulated by the components of the ATP regenerating system, creatine phosphate and creatine phosphokinase, indicating that the nucleotide pools were relatively stable.

The behavior of the second column reconstitution reaction in response to varying salt, temperature, and magnesium was also determined. Specific transcription was inhibited at concentrations of KCl above 90 mM, remaining relatively insensitive over a broad intermediate range (Fig. 7*A*). Dependence on magnesium showed a similarly broad optimum with a peak at 4 mM $MgCl_2$ (Fig. 7*B*). These responses are both similar to those of the polymerase II + crude S-100 system as reported

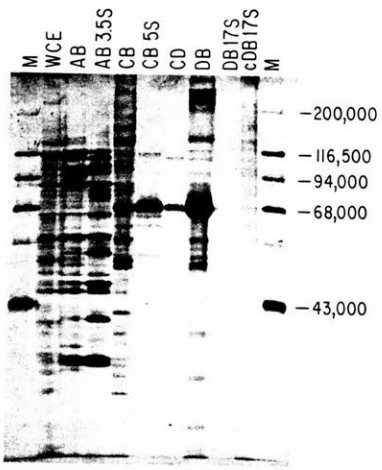


FIG. 6. **Sodium dodecyl sulfate-polyacrylamide gel electrophoretic analysis of transcriptionally active fractions.** 0.7-mm thick 10% polyacrylamide gels were run according to Laemmli, except that the acrylamide/bisacrylamide ratio was 75:1. Samples were electrophoresed, and the gel was stained with silver according to the method of Merril *et al.* (25). The lanes were loaded as follows: *M*, molecular weight markers; *WCE*, 0.25 μl of whole cell extract; *AB*, 0.33 μl of fraction [AB]; *AB 3.5 S*, 8 μl from the region of a sucrose gradient containing AB activity; *CB*, 3 μl of fraction [CB]; *CB 5 S*, 15 μl from the region of a sucrose gradient containing CB activity; *CD*, 7 μl of fraction [CD]; *DB*, 1.5 μl of fraction [DB]; *DB 17 S*, 15 μl from the region of a sucrose gradient containing DB activity; *cDB 17 S*, 105 μl from the region of a sucrose gradient containing DB activity, concentrated by precipitation in 10% trichloroacetic acid. The 68,000 molecular weight protein in fractions [CB], [CD], and [DB] is BSA which was added during chromatography.

TABLE IV

Dependence on reaction components

Assays were performed with reaction conditions and protein concentrations as described under "Methods." The template DNA and quantitation were as in Table I. The indicated components were omitted from the reactions.

Component omitted	Specific incorporation
	%
None	100
Template DNA	<1[a]
Mg^{2+}	<1[a]
Creatine phosphate	150
Creatine phosphokinase	146
GTP	<1[a]
CTP	<1[a]
UTP[b]	<1[a]

[a] The limit was estimated by comparison with reconstructions.

[b] The test reaction and a control were performed with [α-^{32}P]GTP (ICN).

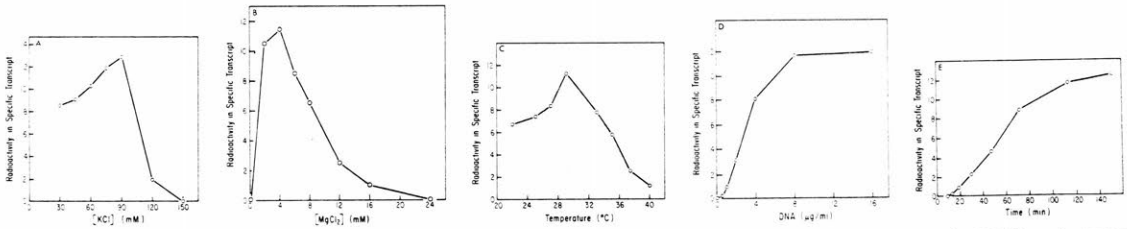


FIG. 7. **Dependence of accurate transcription on reaction parameters.** *A*, effect of salt concentration. Analytical reactions were performed under standard conditions, except that the KCl concentration was varied as noted in the figure. The template DNA and quantitation were as in Table I. Reactions received 0.12 μg of template DNA, 4.4 units of polymerase, 0.5 μl of [AB], 2.0 μl of [CB], 1.8 μl of [DB], and 75 ng of purified M_r = 110,000 protein. *B*, effect of $MgCl_2$ concentration. Analytical reactions were performed under standard conditions, except that the $MgCl_2$ concentration was varied as noted in the figure. The template DNA and quantitation were as in Table I. Reactions received 0.12 μg of template DNA, 4.4 units of polymerase, 0.5 μl of [AB], 4 μl of [CB], 3 μl of [CD], 3 μl of [DB], and 11 units of placental ribonuclease inhibitor. *C*, effect of temperature. Analytical reactions were performed under standard conditions, except that the temperature was varied as noted in the figure. The template DNA and quantitation were as in Table I. Reactions received 0.12 μg of template DNA, 4.4 units of polymerase, 0.5 μl of [AB], 4 μl of [CB], 3 μl of [CD], 3 μl of [DB], and 11 units of placental ribonuclease inhibitor. *D*, dose-response curve for template DNA concentration. Analytical reactions were performed under standard conditions. The template DNA consisted of pFLBH digested with *Pst*I; reactions received 0.32, 0.16, 0.08, 0.04, 0.02, or 0 μg of template DNA. Each reaction included 5.5 units of polymerase, 0.6 μl of [AB], 3.5 μl of [CB], 3 μl of [CD], and 3 μl of [DB]. The normalized amount of specific transcription was plotted against the final concentration of template DNA. *E*, time course. A large volume of standard reaction mix received, per reaction volume, 0.12 μg of template DNA (as in Table I), 4.4 units of polymerase, 0.6 μl of [AB], 3 μl of [CB], 3 μl of [CD], and 2.5 μl of [DB]. At the times indicated, 20-μl aliquots were removed and extracted. The normalized amount of specific transcription was plotted against the elapsed time of incubation of the reaction mix.

by Weil *et al.* (3). Dependence on temperature, however, was strikingly different for the reconstituted and crude systems. Degradation of RNA is apparent in the whole cell extract and S-100 systems above 30 °C, with a complete loss of detectable RNA at 37 °C. In contrast, the reconstituted system was able to generate RNA at temperatures up to 37.5 °C (Fig. 7*C*).

The effect of varying DNA concentrations on accurate transcription is shown in Fig. 7*D*. At extremely low concentrations of DNA (<0.5 μg/ml), no specific transcription was observed with the reconstituted system. Above this threshold, the reconstituted system responded linearly to template concentration over a 10-fold range, ultimately reaching a plateau in the amount of runoff RNA.

The time course of transcription with the reconstituted system was qualitatively similar to that of the whole cell extract (Fig. 7*E*). After a 15-min lag, transcription increased linearly for about 1 h. As in the whole cell extract, the lag appears to be due to two phenomena, a preliminary DNA-protein interaction (which does not require nucleotides), and a subsequent chain elongation step.[3]

DISCUSSION

Accurate transcription *in vitro* of genes encoding messenger RNA has been shown to involve multiple components in addition to purified RNA polymerase II. The key to purification and analysis of each of the factors involved is the definition and characterization of sensitive assays. Conversely, the features of the assay are of paramount importance in interpreting any experimental result.

The basic runoff assay used in these experiments requires initiation at the promoter, elongation to a unique cleavage site in the DNA, and sufficient stability for the full length runoff RNA to be detected. Fine structure analysis confirms that RNA synthesized in cell-free systems has precisely the same 5′ terminal sequences as *in vivo* messenger RNA; to this extent the biochemistry of soluble reaction systems must be similar to that occurring *in vivo*. However, the *in vitro* reactions is almost certainly not responsive to all of the machinery directing the *in vivo* process of transcription. The efficiency of *in vitro* transcription by either the whole cell extract or the reconstituted system in terms of initiation events per cell equivalent is much less than that *in vivo* (4). Moreover, the nucleotide sequences around promoter sites effecting the *in vitro* process are only a subset of the sequences involved *in vivo* (for a review see Ref. 26). These differences could result in part from the nature of the template DNA, which *in vivo* is chromatin but *in vitro* is a relatively short (5000 bp) linear molecule without obvious topological constraints. Alternatively, certain activities involved in transcription *in vivo* might be lost during solubilization. The runoff assay thus defines the minimum apparatus necessary for the utilization of promoters.

The use of the adenovirus major late promoter for the assay of accurate transcription could conceivably lead to the purification of a promoter-specific subset of cellular transcription factors. The reconstituted system at its current stage of purity also recognizes both the adenovirus EIV promoter (Fig. 3) and the EIa promoter (data not shown). Further fractionation and analysis using these and other promoters may, however, detect promoter-specific components.

In addition to RNA polymerase II, the reconstituted system consists of three mutually dependent fractions, each of which has been purified through two steps of chromatography. The transcriptional activities in the three required second column fractions demonstrated the following characteristics.

1) Each was sensitive to mild heat treatment.

2) Each sedimented with a peak of activity through sucrose gradients.

3) Each titrated linearly with concentration through a range of at least 10-fold.

Fraction [AB] stimulated the runoff transcription reaction at least 100-fold. We believe that [AB] contains two relevant activities: a ribonuclease inhibitor (for which a commercial preparation of pure human placental ribonuclease inhibitor can substitute), and a *bona fide* transcription factor. [AB] presumably corresponds to factor 2A of Matsui *et al.* (5).

The specific transcription reaction was also at least 100-fold dependent on fraction [CB]. This fraction has been separated with very low yields into two mutually dependent subfractions by chromatography on phosphocellulose or heparin-Sepharose (data not shown). However, neither of the procedures has proven satisfactory; possibly one of the subfractions is very labile after the separation. The co-sedimentation on sucrose gradients, cofractionation on a gradient salt elution of DNA-cellulose (data not shown), and linear titration curve are consistent with the two factors in [CB] existing as a complex. However, these results are not conclusive and require further study. Fraction [CB] presumably corresponds to factor 2B of Matsui *et al.* (5). Dingnam and Roeder have also reported the separation of this activity into two subfractions using DEAE-cellulose.[4]

Fraction [DB] stimulated a very low basal level of transcription by 20-fold. This basal activity in the absence of [DB] could be due to contamination in other fractions or could indicate that the activity in [DB] was not absolutely necessary for specific transcription. The inhibition of runoff RNA synthesis by high concentrations of [DB] was much more pronounced when observing a longer transcript, suggesting either interference with the elongation reaction or a decrease in RNA stability (data not shown). [DB] apparently corresponds to factor 2D of Matsui *et al.* (5).

One consequence of the purification so far achieved was the increased level of background incorporation with the reconstituted system. This background was effectively suppressed by [CD], a fraction not required for accurate transcription. The 110,000 molecular weight major polypeptide in this fraction has been identified[2] as poly (ADP-ribose) synthetase, a protein believed to bind nicks and free ends in DNA (27, 28). As RNA polymerase II initiates nonspecifically at nicks and ends (1, 2), the background suppression function of this M_r = 110,000 protein *in vitro* may be an artifact consequent on the use of naked DNA as template.

The reconstituted transcription system has already demonstrated distinct advantages over the initial whole cell extract. The whole cell extract behaves nonlinearly with protein concentration, precluding a reliable measurement of activity (4). Even more seriously, the whole cell extract does not respond simply to DNA concentration (4, 23). Typically, the whole cell extract produces no runoff transcripts below a concentration of 10–50 μg/ml of total DNA. The reconstituted system showed a similar threshold, but at a much lower concentration of DNA (0.5 μg/ml), and the response above this value was linear over a considerable range of concentration. This was probably because a number of inhibitory DNA-binding proteins were removed by chromatography. As a further advantage, the reconstituted system was highly responsive to exogenous RNA polymerase II, and was absolutely dependent on each of the nucleoside triphosphates.

One of the major limitations of all current *in vitro* systems for accurate transcription by RNA polymerase II is the small fraction (less than 10% per hour) of DNA templates utilized

[3] A. Fire and M. Samuels, unpublished observations.

[4] D. Dingnam and R. G. Roeder, personal communication.

(4). This prevents the application of a number of methodologies for studying DNA-protein interactions. At the moment, the reconstituted system is not more efficient than the initial whole cell extract, probably because the factors are present at about the same concentration in each system. Thus, the reconstitution reaction shows similar activity per template molecule to the whole cell extract, with 1/20 of the total protein.

The specificity (or lack thereof) of the reaction catalyzed by purified RNA polymerase II gives some hints as to possible functions for the factors required to observe accurate transcription. The purified polymerase initiates most efficiently on single-stranded DNA, or, as already noted, at nicks and ends in double-stranded DNA. The inability to initiate efficiently within double-stranded DNA may result from the absence of binding to double-stranded regions, from a failure to open the DNA helix, or from the failure to begin polymerizing an RNA chain. Auxiliary factors interacting directly with the DNA and/or with the polymerase could be required at any of these steps.

The elongation reaction carried out by purified polymerase after initiation at nicks, ends, or on single-stranded DNA apparently produces a heteroduplex product (29), displacing the non-coding strand of DNA if it is present. However, promoter runoff RNA synthesized in the whole cell extract is not hybridized to the DNA (data not shown). Kadesch and Chamberlin (29) have suggested that a factor might be involved in displacement of the nascent RNA chain from the template during specific transcription. Such a displacement reaction might be essential for elongation from internal initiation sites in duplex DNA. Pausing or premature termination caused by sequence elements or DNA-binding proteins might also necessitate anti-attenuation factors.

Finally, factors might interact with the RNA transcript to effect elongation and/or stability under assay conditions. Interactions of this type could involve formation of a capped 5′ terminus. It is worth noting, however, that at least 95% of the 68,000 molecular weight capping enzyme (assayed as in Venkatesan and Moss; Ref. 30) is present in the [B] fraction (data not shown). This fraction was not required for the observation of accurate transcription.

HeLa cells are not available in adequate quantities for bulk purification of the transcriptional factors. On the other hand, the use of more abundant animal tissue (*e.g.* calf thymus), is hampered by the presence of large amounts of ribonuclease in soluble extracts. Partly for this reason, primary extracts of tissue generally do not support accurate transcription. Given an inactive starting material, the co-purification of mutually dependent components would be extremely difficult. However, one can expect factors from tissue sources to substitute for the homologous HeLa transcription factors (indeed, this has been shown for hen oviduct; Ref. 6). The use of the linear assays constructed with fractions from HeLa cells should thus allow the identification and quantitation of each of the transcriptional activities upon fractionation of tissue extracts.

Acknowledgments—We are grateful to U. Hansen and R. Kingston for their advice and guidance; to K. Berkner, F. Laski, and J. Manley for recombinant plasmids; to N. Crawford, R. Padgett, M. Chow, and S. Desiderio for suggestions on fractionation; to L. Corboy for expert technical assistance; to N. Crawford, E. Slattery, D. Dingnam, T. Matsui, and R. Roeder for communicating results prior to publication; to the MIT Cell Culture Center for preparation of HeLa cells; to M. Siafaca for help in preparing the manuscript; and to members of the Sharp lab for illuminating discussions.

REFERENCES

1. Chambon, P. (1975) *Annu. Rev. Biochem.* **44,** 613–638
2. Roeder, R. G. (1976) in *RNA Polymerase* (Losick, R., and Chamberlin, M., eds) pp. 285–292, Cold Spring Harbor Press, New York
3. Weil, P. A., Luse, D. S., Segall, J., and Roeder, R. G. (1979) *Cell* **18,** 469–484
4. Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 3855–3859
5. Matsui, T., Segall, J., Weil, P. A., and Roeder, R. G. (1980) *J. Biol. Chem.* **255,** 11992–11996
6. Tsai, S. Y., Tsai, M.-J., Kops, L. E., Minghetti, P. P., and O'Malley, B. W. (1981) *J. Biol. Chem.* **256,** 13055–13059
7. Dynan, W., and Tjian, R. (1981) in *Developmental Biology Using Purified Genes* (Brown, D., ed) pp. 401–414, Academic Press, New York
8. Kornberg, A. (1980) in *DNA Replication*, pp. 347–570, W. H. Freeman, San Francisco
9. Alberts, B., and Herrick, G. (1971) *Methods Enzymol.* **21,** 198–217
10. Bradford, M. (1976) *Anal. Biochem.* **72,** 248–254
11. Hodo, H. G., and Blatti, S. P. (1977) *Biochemistry* **16,** 2334–2343
12. Crawford, N., Fire, A., Samuels, M., Sharp, P. A., and Baltimore, D. (1981) *Cell* **27,** 555–561
13. Manley, J. L., Fire, A., Samuels, M., and Sharp, P. A. (1982) *Methods Enzymol.*, in press
14. McMaster, G., and Carmichael, G. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 4835–4838
15. Maxam, A., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
16. Laskey, R. A., and Mills, A. D. (1975) *Eur. J. Biochem.* **56,** 335–341
17. Ziff, E. (1980) *Nature (Lond.)* **287,** 491–499
18. Hansen, U., Tenen, D. G., Livingston, D. M., and Sharp, P. A. (1981) *Cell* **27,** 603–612
19. Blackburn, P., Wilson, G., and Moore, S. (1977) *J. Biol. Chem.* **252,** 5904–5910
20. Ziff, E. B., and Evans, R. M. (1978) *Cell* **15,** 1463–1475
21. Baker, C., and Ziff, E. (1981) *J. Mol. Biol.* **149,** 189–221
22. Steenbergh, P., and Sussenbach, J. (1979) *Gene* **6,** 307–318
23. Fire, A., Baker, C. C., Manley, J. L., Ziff, E. B., and Sharp, P. A. (1981) *J. Virol.* **40,** 703–719
24. Lee, D. C., and Roeder, R. G. (1981) *Mol. Cell. Biol.* **1,** 635–651
25. Merril, C. R., Goldman, D., Sedman, S. A., and Ebert, M. H. (1981) *Science* **211,** 1437–1438
26. Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383
27. Ohgushi, H., Yoshihara, K., and Kamiya, T. (1980) *J. Biol. Chem.* **255,** 6205–6211
28. Benjamin, R. C., and Gill, D. M. (1980) *J. Biol. Chem.* **255,** 10502–10508
29. Kadesch, T. R., and Chamberlin, M. J. (1982) *J. Biol. Chem.* **257,** 5286–5295
30. Venkatesan, S., and Moss, B. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 340–344

Chapter V

KINETIC DISSECTION OF THE RNA POLYMERASE II REACTION

*This work was performed in collaboration with Mark Samuels

Summary

Accurate transcription by RNA polymerase II has been shown to require multiple factors in addition to the purified polymerase. In this study we use a reconstituted transcription system, consisting of purified RNA polymerase II and three essential HeLa cell chromatographic fractions, to study events leading to transcription from the adenovirus major late promoter. A preincubation-pulse-chase protocol resolves the reaction into events occurring before and after nucleotide addition. Preincubation of template with a mixture of RNA polymerase II and factors allows formation of "activated" complexes, which are defined by the ability to rapidly commence accurate transcription when presented nucleotides. Maximal activation requires that polymerase, template, and each of the three HeLa fractions be present during preincubation. The activated complexes are template associated, as shown by their inability to exchange onto a second template added during further preincubation. Similar protocols are used to define functional intermediates leading to the activated complex. A template associated functional complex is formed during preincubation of template with just two of the HeLa fractions. Polymerase can associate with this intermediate complex in the absence of the third HeLa fraction. In the accompanying paper, we describe a direct analysis of initiation by "activated" complexes.

INTRODUCTION

Synthesis of messenger RNA precursors in eukaryotic cells is catalyzed by RNA polymerase II (1). An understanding of site selection and transcription by this enzyme is crucial to the biochemical analysis of gene expression. Precise mapping of 5' termini for a variety of messenger RNAs *in vivo* has provided the criterion for defining faithful transcription *in vitro*. Purified RNA polymerase II does not faithfully initiate transcription on any gene so far tested. However, cellular extracts have been shown to carry out faithful transcription on a variety of genes (2-4).

Specific transcription requires several factors present in cellular extracts in addition to purified RNA polymerase II (5). These factors have been resolved by chromatography and can be used to constitute a reaction similar to that observed in the initial extract (5-8). At the moment, at least four factors have been detected; as yet none has been purified to homogeneity.

The quantitative assay for faithful transcription has been to measure total accumulation of a discrete length RNA product. Because efficient termination has not been observed *in vitro*, such studies generally use a cleaved DNA template. Accurate initiation followed by elongation to the end of the template produces a "runoff RNA" that migrates as a band upon electrophoresis. Even in the simplest conceivable scheme, many different events must occur to generate runoff RNA. Polymerase must bind to the initiation site; the strands of the helix must be locally separated to allow reading of the DNA; the first two nucleoside triphosphates must be properly positioned and joined to initiate transcription; finally, subsequent nucleotides must be added to complete synthesis of the runoff transcript (elongation). The requirement for multiple protein factors in addition to RNA polymerase II suggests that the transcription reaction could be considerably more intricate than this simple scheme.

Because runoff transcript accumulation depends on a complex series of events, a large change in rate for a step not near limiting would be undetectable. This is of particular importance in attempting to reconstruct regulatory effects and sequence dependences observed *in vivo*. For example, deletion of a DNA segment promoting polymerase

binding might have no detectable effect on total accumulation if the relevant binding step is not limiting under the assay conditions used. A kinetic analysis of different steps in the reaction would contribute to interpretation of such studies, and could provide the desired biochemical assays for regulation and for upstream promoter elements.

The pathway leading to transcription by E. coli RNA polymerase has been well studied (9). The initial [polymerase:promoter] complex, called the "closed" complex, must undergo an isomerization reaction to an "open" complex before initiating transcription. This isomerization is nucleotide independent and apparently involves an opening of the DNA helix around the initiation site (10). For some promoters, efficient formation of the open complex requires extra factors in addition to the polymerase holoenzyme. In the case of the lac promoter, an activator (catabolite activator protein) binds specifically to the template, facilitating subsequent transcription by polymerase (11,12). A similar phenomenon occurs in transcription of 5S genes by RNA polymerase III; in that case, the promoter recognition event appears to be mediated primarily by factor TFIIIA, a highly promoter-specific DNA binding protein (13-16). The "preprimosome" formed during synthesis of replication primers on ØX 174 provides a more intricate example of sequence specific events that can precede polymerase binding (17). In each of the above cases, a functional complex can be identified on the template in the absence of polymerase. In the 5S and catabolite activation examples, the interaction between factor and template plays a physiological role in regulating transcription.

We describe here a kinetic dissection of the RNA polymerase II transcription reaction. In particular, assays for events preceding initiation and an assay for subsequent elongation events are presented. These assays have been used to obtain information concerning the roles of template, polymerase, and the required protein fractions in events leading to transcription.

Materials and Methods

Unlabeled nucleoside triphosphates purified by high pressure liquid chromatography were purchased from ICN. [α-^{32}P] UTP and GTP were purchased from NEN. Purified human placental RNase inhibitor (18) (RNasin) was purchased from Biotec. Poly (dI-dC:dI-dC) was purchased from Collaborative Research. HeLa chromatographic fractions (7) and purified calf thymus RNA polymerase II (19) were prepared as described. Our preparations of RNA polymerase II had specific activities of $1\text{-}2\times10^5$ u/mg (7). Purity of the polymerase preparation (>95%) was determined by examination of denaturing polyacrylamide gels stained with silver nitrate.

Transcription reaction. The following solution conditions were maintained throughout each reaction: 12 mM Hepes -NaOH pH 7.9, 12% glycerol, 1 mM EDTA, 0.6 mM dithiothreitol, 60 mM KCl, 5 mM $MgCl_2$. To start preincubations, proteins (mixed at 4°C) were added to DNA; all incubations were at 30°C. A standard "complete" preincubation contained (except where noted) template DNA, RNA polymerase II (20 units), RNasin (15 units), and HeLa chromatographic fractions: 1 μl [AB], 3 μl [CB], 2 μl [DB], and 2 μl [CD]. Preincubations were performed in 20 μl; in experiments with sequential preincubations, the first preincubation was in 15 μl with the second preincubation bringing volume up to 20 μl. DNA concentrations in the text all refer to a 20 μl preincubation. Pulses and chases each added 5 μl to total volume. Pulse nucleotide concentrations were 2 M [α-^{32}P] UTP or GTP (400 Ci/mmol) and 30 μM of the three unlabeled nucleoside triphosphates. Chase nucleotide concentrations were 1 mM of each of the four nucleoside triphosphates, with equimolar $MgCl_2$ added.

RNA analysis. Extraction of RNA products, denaturation with glyoxal, and agarose gel electrophoresis followed a standard protocol (4). Procedures for quantitation of specific transcription by densitometry of preflashed film have been described in detail (7). Because relative peak area measurements from different experiments depended on autoradiography time and densitometer settings, ordinate values have been normalized to an (arbitrary) value of 10 for a "complete" reaction: a 60 minute preincubation with 60 ng of the specific template, a 4 minute pulse and a 10 minute chase.

RESULTS

The Reconstituted Transcription System

We have recently described a transcription system utilizing purified RNA polymerase II supplemented with partially purified transcription factors (7). Calf thymus RNA polymerase II was purified (>95% pure) by the method of Hodo and Blatti (19) and contained an equal mixture of IIA and IIB forms. A whole cell extract of HeLa cells (3), which is itself capable of directing faithful transcription by RNA polymerase II, was fractionated as shown in Figure 1a. The purified polymerase and three of the HeLa fractions (designated [AB], [CB], and [DB]) were necessary and sufficient for production of accurately initiated runoff transcripts from the adenovirus major late promoter. Transcriptional activities in these fractions exhibited the following properties: 1) each was sensitive to mild heat treatment, 2) each sedimented with a peak of activity through sucrose gradients, and 3) each titrated linearly with concentration through a range of at least ten fold (7). In the following experiments, concentrations of polymerase and the three required HeLa fractions were within the linear ranges previously determined.

Fraction [CD] (consisting primarily of a M_r=116,000 polypeptide which has been identified as ADP ribosyl transferase; 20) suppressed a high nonspecific background without apparent effect on the specific signal (7), and was added to facilitate quantitation. Fraction [AB] contains an RNase inhibitor in addition to the transcriptional factor (7). Purified human placental RNase inhibitor will substitute for the RNase inhibiting activity in [AB], and was routinely added to assays. The templates for these experiments were all derived from plasmid pFLBH, a pBR322 derivative containing a 874 bp segment spanning the adenovirus major late promoter (Figure 1b).

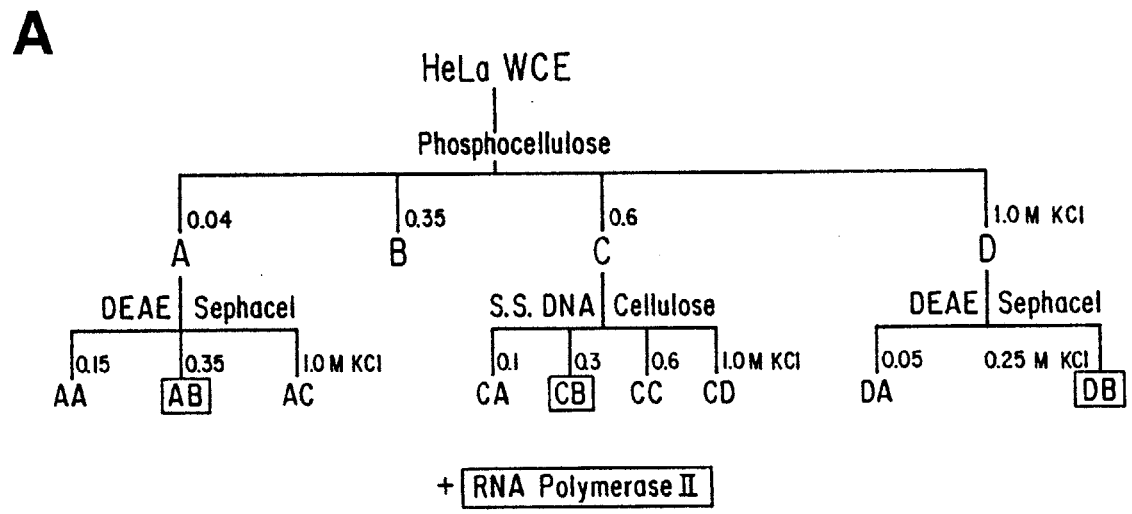
The Preincubation-Pulse-Chase Protocol

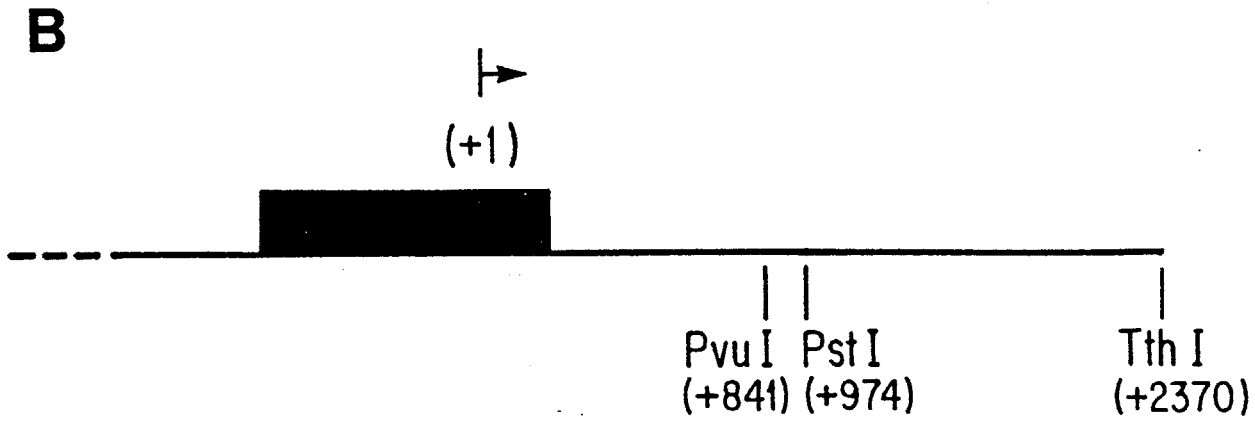
A three stage protocol was used to dissect the transcription reaction. I. Preincubation: DNA and protein were preincubated in the absence of nucleoside triphosphates. II. Pulse: The four ribonucleoside triphosphates were added to commence RNA synthesis. One nucleoside triphosphate was radioactive [α-^{32}P] and present at limiting concentration (2 M). III. Chase: Excess unlabeled nucleoside

Figure 1

Panel A: Scheme for the resolution of transcription factors contained in solubilized HeLa cell extracts. The details of the chromatography have been described (7).

Panel B: The adenovirus late promoter-pBR322 recombinant used as template in these studies. Adenovirus 2 sequences (14.7-17.1 map units), denoted by a solid bar, were inserted between the Bam HI and Hind III sites of pBR322. The adenovirus major late cap site (26) denoted by an arrow, defines position +1 of the plasmid; relevant restriction enzyme sites are shown.





triphosphates (1 mM) were added. This protocol was developed to allow pre-initiation complexes to form during the preincubation, initiation and early elongation of chains during the pulse, and elongation of RNAs to the end of the template during the chase. Such a protocol should be particularly useful in resolving pre- and post-initiation events.

It was conceivable that the fractions were contaminated with a sufficient concentration of nucleotides so that initiation could occur during the preincubation phase. However, this is not the case. As is shown in the accompanying paper, by manipulation of 5' termini all of the chains are initiated during the pulse (21). The signal observed (radioactive incorporation into specific runoff RNA) is therefore a product of:

A. The number of RNA polymerase molecules accurately initiating during the pulse.

B. The average number of ^{32}P-rNTP incorporated into each chain during the time of the pulse.

C. The fraction of these chains that becomes fully elongated during the chase to yield full length runoff RNA.

In order to characterize the elongation reaction, a time course of chase was performed with fixed periods for the preincubation (32 min) and pulse (2 min) (Fig. 2A). The template for this experiment was an equimolar mixture of PvuI-cut pFLBH and TthI-cut pFLBH (see Figure 1B). The elongation rate during the chase was measured by observing the differential appearance of the two resulting transcripts. The 841 n PvuI runoff transcript appeared between 1 min and 2 min, while the 2364 n TthI runoff transcript appeared between 4 min and 8 min. From this and similar experiments, an average elongation rate between 400 and 600 nucleotides per minute was calculated for the chase phase. This is similar to the values observed in pulse chase experiments with purified RNA polymerase II elongating from the ends of a phage T7 template (22).

During the course of the chase, a number of bands shorter than the full length runoffs appeared and subsequently decreased. These are evidently pause sites during elongation. These pause sites could be intrinsic to RNA polymerase II (such pause sites were observed by Kadesch and Chamberlin; 22) or could represent binding sites for proteins that interfere with elongation.

Figure 2. Time courses for preincubation, pulse and chase.

Preincubations contained the "complete" protein mix described in Methods and the indicated template. Pulse and chase conditions are described under Methods. Graphs show the radioactive incorporation into the specific runoff RNA, quantitated by densitometry as described (7).

Panel A: Preincubation was for 60 minutes with a mixture of 60 ng PvuI-cut pFLBH and 60 ng TthI-cut pFLBH. After a 4 min pulse, chases of the indicated length were performed.

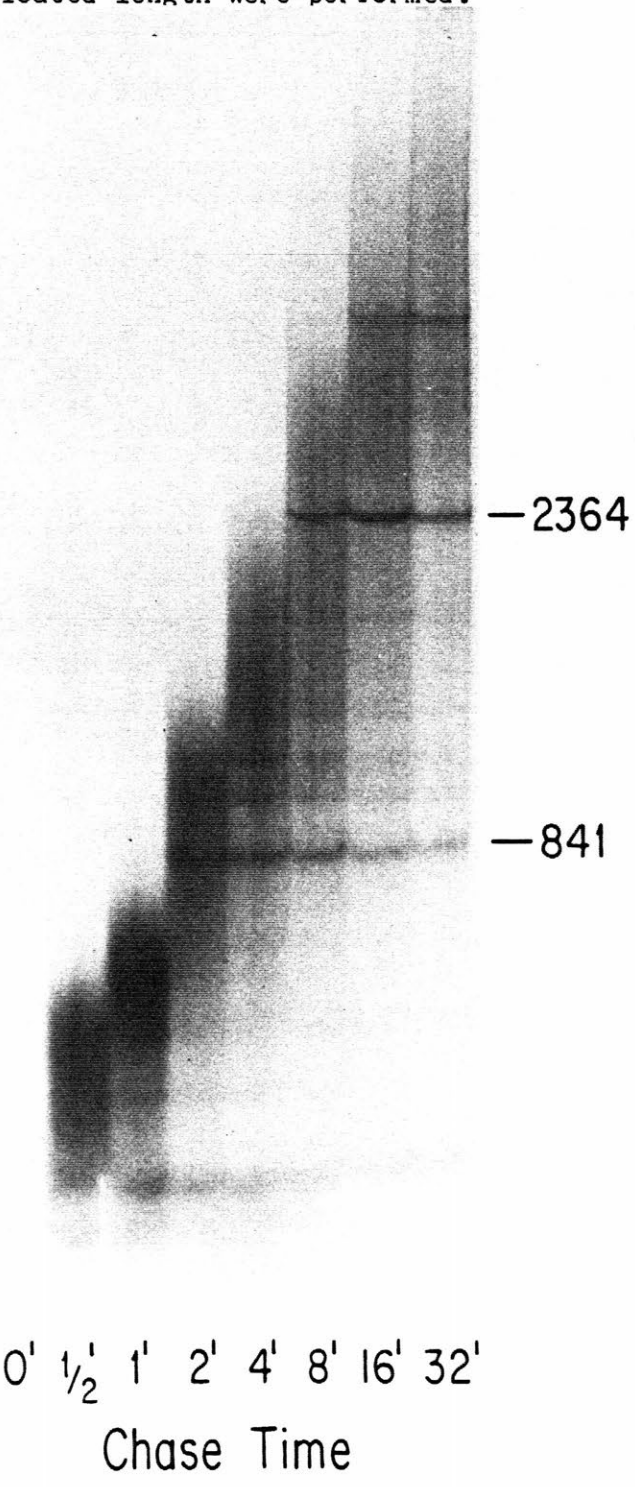
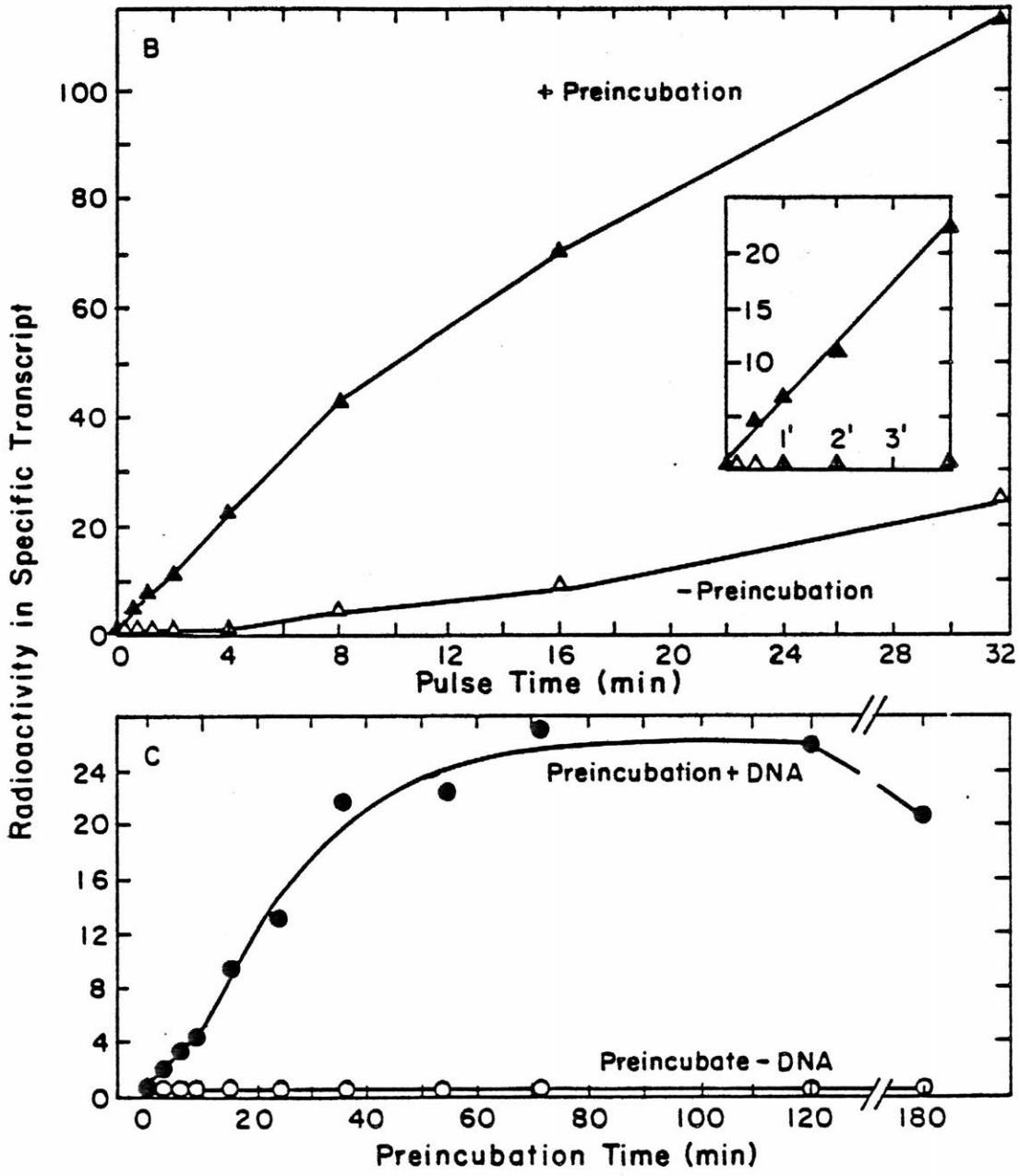


Figure 2(continued)

Panel B: (-▲-) Preincubation for 60 minutes with 120 ng PstI-cut pFLBH was followed by a variable pulse and constant 10 min chase. (-△-) is same except without preincubation. Inset shows expansion of early time points.

Panel C: (-●-) A variable (0-180 min) preincubation with 120 ng of PstI cut pFLBH a 4 minute pulse and 10 minute chase. (-o-) Same with template omitted from preincubation and added to pulse.



Using the above pulse-chase protocol, the specific activity (in counts per transcript) should be similar for the two different length runoff transcripts, since all radioactivity is incorporated near the 5' end. If elongation during the chase is efficient, and the transcripts are stable, then signals from an equimolar template mixture should reach plateaus of equal intensity after a long chase. This was the case for the 841 n and 2364 n late promoter runoffs. An 8 minute chase was sufficient for maximal accumulation of both runoffs (Fig. 2A).

Figure 2B shows the time course for the pulse phase of the reaction, with (60 min) preincubation and a constant (10 min) chase (-▲-). Each point represents radioactive incorporation into the specific runoff transcript, as measured by densitometer tracing of autoradiograms. With the 60 minute preincubation, the observed signal increased linearly for pulse times up to 32 min. This would be expected if a population of polymerase molecules initiated and began elongation at the beginning of the pulse. Two points from this experiment are worth special note. First, no signal was observed when the pulse was for zero time (i.e. labeled pulse and excess unlabeled chase nucleotides added together). This indicates that label was effectively diluted during the chase phase. Second, even with the longest (32 min) pulse, synthesis of full length runoff RNA still depended upon a chase (data not shown). This is the result of a slow elongation rate (<20 nucleotides per minute) under the limiting nucleotide conditions of the pulse (unpublished observations).

A different time course of pulse labeling was observed when the preincubation was omitted. In this case incorporation was not observed at early time points, but eventually occurred after a lag [Fig. 1B (-Δ-)]. Hence, events in the preincubation must be required for transcription to begin immediately when nucleotides are added.

A time course of preincubation with constant pulse (4 min) and chase (10 min) phases is shown in Figure 2C (-●-). The observed signal intensified as preincubation time was increased, reaching a plateau with approximately one hour of preincubation and staying relatively constant over the next several hours. A very small amount of the specific runoff transcript was observed with no preincubation (<1% of maximum). By keeping the pulse phase short this specific background

was minimized, thus maximizing the dependence on the preincubation.

DNA Requirement During Preincubation

Preincubation of proteins in the absence of the DNA template did not stimulate subsequent incorporation [Fig. 2C (-o-)]. Thus, DNA-protein interactions are crucial in the preincubation. This experiment did not address the nature of the preincubation DNA requirement. Indeed the role of DNA could either be the formation of a specific pre-transcription complex ("activating" the DNA as template), or a less direct role: modifying the transcriptional apparatus or removing inhibitors so that transcription could begin more rapidly upon nucleotide addition. In order to examine these possibilities, experiments involving two DNA templates were performed. Selection of a specific template during preincubation would suggest the formation of a protein-template complex.

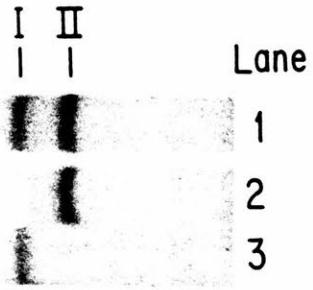
Plasmid pFLBH, cleaved uniquely by either PvuI or PstI, generates two templates yielding runoff transcripts of 841 and 974 n respectively. After preincubating a mixture of these two templates (3 μg/ml each) with polymerase and factors for 60 min, a standard 4 min pulse and 10 min chase yielded the two expected runoff RNAs (Fig. 3, lane 1). If one of the two templates was added in the preincubation and the other was added at the beginning of the pulse, only RNA originating from the preincubated DNA was observed (Figure 3, lanes 2,3). Thus the preincubated DNA had been assembled into some transcriptionally activated structure, distinguished from the second template.

If the preincubation dependence represents time required to form an "activated" complex on the template, then the second added template should be subject to the same lag (see Fig. 2C) as was observed in single template experiments. It is also conceivable, however, that preincubation with the first template could have rendered the protein mix unable to transcribe any subsequently added template. To address this possibility of "interference" the second template was allowed a further preincubation before pulse and chase. After preincubation for 60 min with template I (PvuI pFLBH; 3 μg/ml), the second template (PstI pFLBH; 3 μg/ml) was added and allowed an additional preincubation. Incorporation into the two runoff RNAs after the standard short pulse

Figure 3. Activation is specific to the preincubated template.

Plasmid pFLBH was cleaved with PvuI to generate template I and PstI to generate template II. Lane 1 shows the product of a complete (60 min) preincubation with 60 ng of each template followed by a pulse (4 min) and chase (10 min). Lane 2 is the same, except that only template 2 is preincubated and template I is added at the start of the pulse. Lane 3 shows the converse, with template I preincubated and template II added at the start of the pulse. The runoff transcripts from the two templates, RNA_I (841 n) and RNA_{II} (974 n) are indicated.

Preincubate Factors + Pol +		Pulse NTP +
$+DNA_I + DNA_{II}$		
$+DNA_{II}$		$+DNA_I$
$+DNA_I$		$+DNA_{II}$



(4 min) and chase (10 min) was plotted versus time of the second preincubation (Fig. 4A). Under these conditions, utilization of template II increased with increasing time of further preincubation, ultimately reaching levels comparable to that of template I. Thus, prior preincubation with low concentrations of the first template did not significantly alter the transcriptional competence of the protein mix.

Higher concentrations of the first template, however, did interfere with transcription of the second template. As shown in Figure 4B, a prior preincubation with 10 μg/ml of template I blocked template II utilization, even with 60 min of further preincubation. Template I remained activated throughout this further preincubation. In the control reactions (where both templates were added together to the second preincubation) the two resulting runoffs appeared in parallel (Figure 4C, D). Hence the observed interference depended on prior preincubation of the interfering DNA (template I). A similar interference with template II was observed if template I was replaced by pBR322 (data not shown); thus late promoter sequences were not absolutely required in the interfering DNA.

The above results define two assays for activities involved in template specific events preceding transcription: I. Activation. The ability of a template complex to rapidly begin transcription when presented nucleotides defines activation (see Fig. 2B). Operationally, the activation of a preincubated template is quantitated by measuring incorporation into the corresponding runoff transcript after a short pulse and chase. II. Interference. The key to such an assay is to find conditions which allow a previously preincubated template to remain activated but which will not allow activation of a second added template during further preincubation (see Fig. 4B). A positive result (utilization of the first but not the second template) indicates the following: 1) during the first preincubation, some stable functional complex must have formed on the first template and 2) such a complex can no longer form on (and will not exchange onto) new template added during the further preincubation. Assays for interference can detect stable complexes or structures whose formation may precede full template activation.

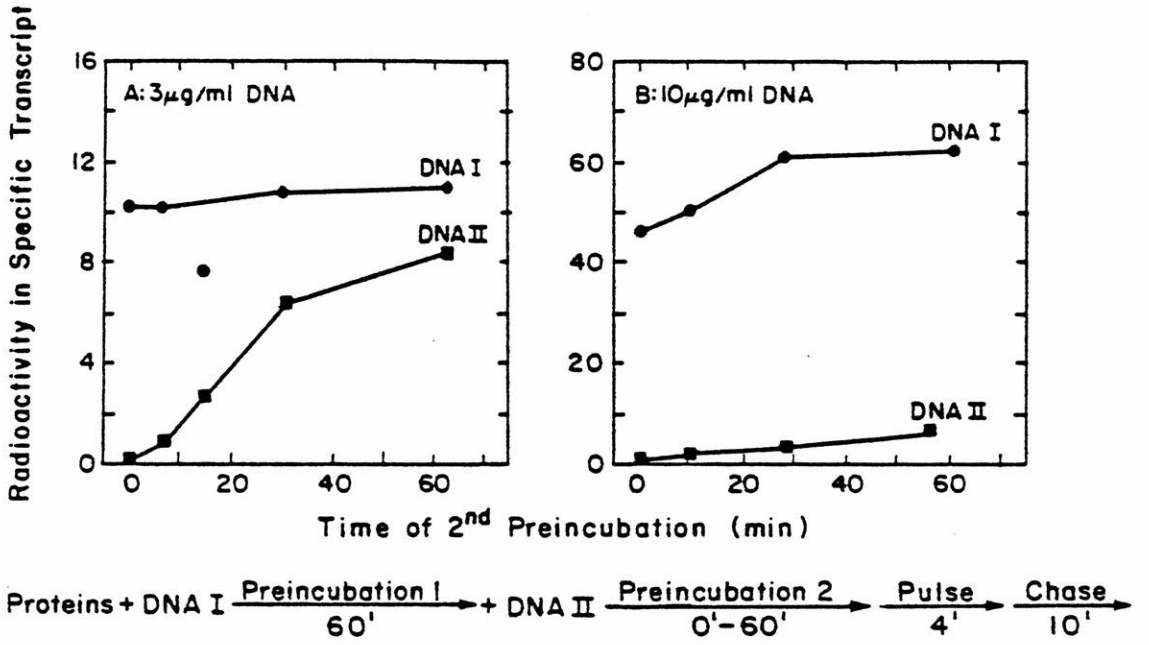
Figure 4. Sequential activation of two templates

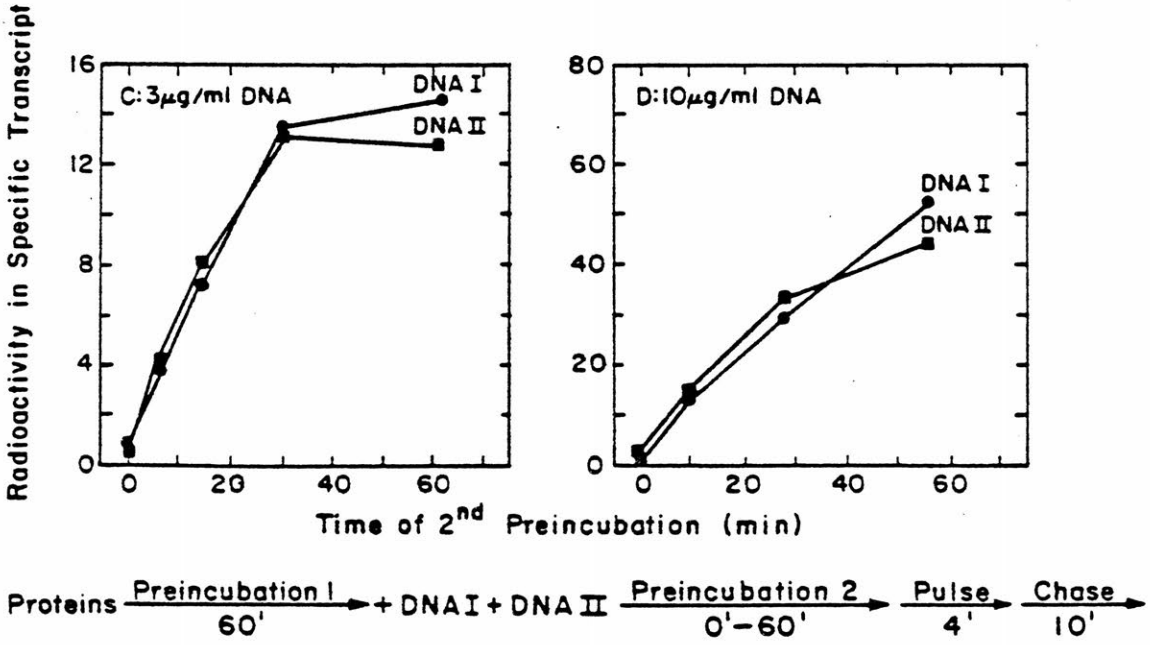
Panel A: 60 ng of PvuI pFLBH (template I) was allowed complete preincubation (60 min) after which 60 ng of PstI pFLBH (template II) was added and allowed a further (second) preincubation of 0-60 minutes. Radioactive incorporation into the two resulting runoff transcripts (DNA_I, 841n [-●-] ; DNA_{II}, 974n [-■-]) after a 4 min pulse and 10 min chase was quantitated as described in Methods.

Panel B: Same as A except with 200 ng of each template.

Panel C: Same as A except that the first preincubation contains neither template; both templates are added to the second preincubation.

Panel D: Same as C with 200 ng of each template.





Role of RNA Polymerase II and Factors During Preincubation

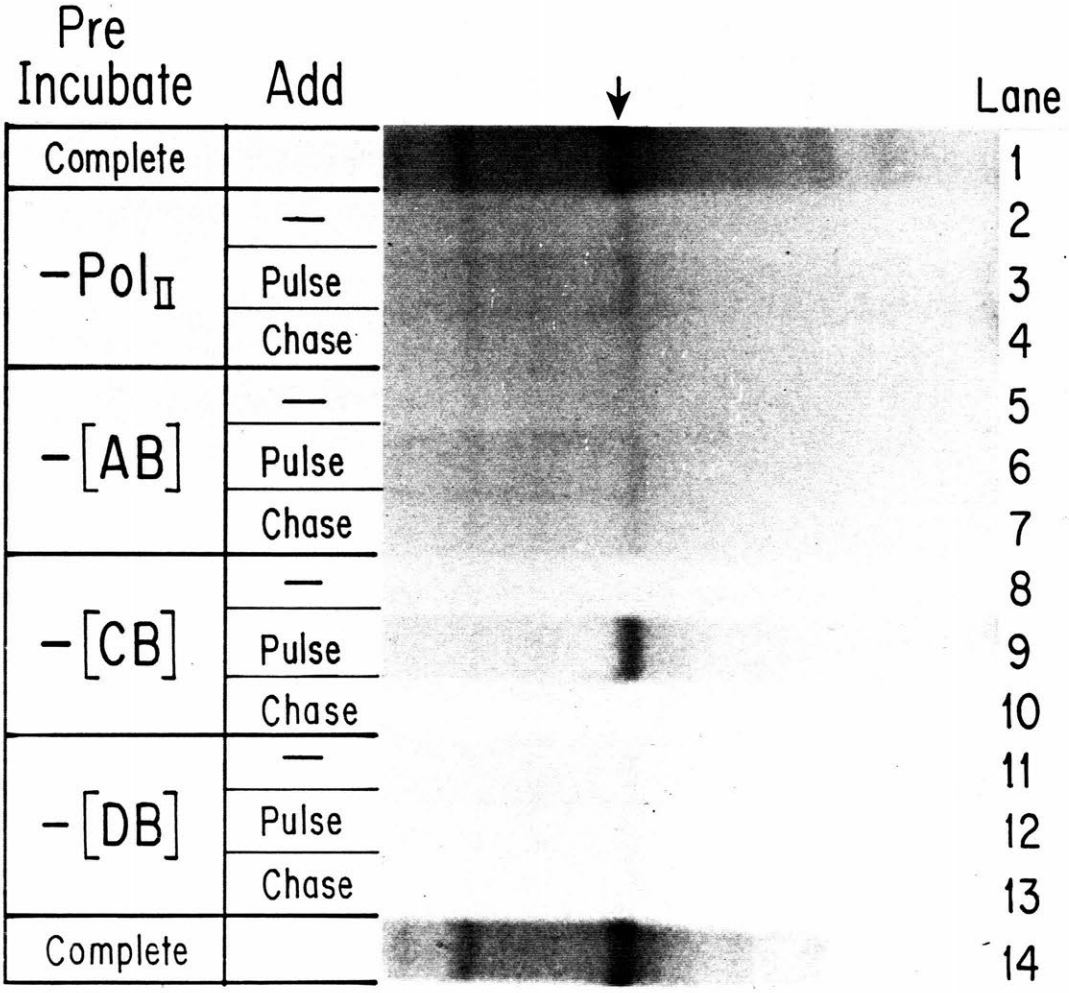
An analysis of requirements for the observed activation and interference phenomena proved suggestive concerning the roles of the different fractions.

The contribution of each fraction to template activation was examined using parallel sets of preincubation-pulse-chase reactions; in each set one of the required protein fractions was omitted from the preincubation and was added to the subsequent pulse phase, to the chase phase, or not at all (Fig. 5). Lane 1 shows a positive control with all components present in the preincubation. A background of 10% was observed in the absence of any added polymerase II (lane 2) and probably resulted from polymerase contamination in other fractions. No significant stimulation above this background was observed if polymerase was added during the pulse or chase phase (lanes 3,4). Similar results were obtained with fractions [AB] and [DB]: to significantly stimulate the signal above a low background, each of these fractions had to be present during the preincubation (lanes 5-7, 11-13). Strikingly different results were obtained when the same experiment was performed for fraction [CB]. Addition of [CB] during the pulse resulted in 20-50% of the positive control signal; this should be compared to a background of less than 1% observed in the absence of [CB] (lanes 8,9). Addition of [CB] during the chase failed to stimulate transcription above this background (lane 10). The unique level of stmulation obtained when [CB] was added to the pulse suggested that [CB] could act late relative to the other components.

To extend the results of figure 5 and obtain a time course of action for each of the protein fractions in the observed activation, a two-period preincubation protocol was used. The first template (I) was preincubated for one hour with three of the four necessary protein components. A second template (II) was then added with the omitted component. The reactions were allowed an additional preincubation of 0-60 min, after which activation of the two templates was measured using the standard short pulse (4 min) and chase (8 min). Activation of each template was plotted versus the time of additional preincubation (solid lines; Fig. 6). Negative controls (never adding the indicated fraction) are shown with dotted lines; Fig. 4 provides a

Figure 5. Factor requirements during preincubation

Lanes 1 and 14 show a complete preincubation (60 min; with 60 ng PstI cleaved pFLBH) followed by a 2 min pulse and 8 min chase. Lanes 2, 5, 8, and 11 show the reaction with the indicated component omitted. Lanes 3, 6, 9, and 12 show equivalent reactions with the indicated component added back to the pulse; lanes 4, 7, 10, and 13 show equivalent reactions with the indicated component added back to the chase. The arrow indicates the position of the 974n late promoter runoff.



positive control (complete preincubations).

Initially these experiments were done at 3 μg/ml of each DNA (Fig. 6, Panels A-D). In the case of polymerase and [CB], the results were straightforward extensions of those in Figure 5. When the [CB] was added in the second preincubation, both templates were activated indicating that 1) the first template was still accessible and 2) preincubation of the other components with DNA I had not interfered with subsequent activation of a second template (Fig. 6B). The more rapid activation of the first template indicates that factors present in the first preincubation selectively expedited activation of template I. When RNA polymerase II was added to the second preincubation, both templates were again activated (Fig. 6D); in this case, however, the two templates were activated in parallel.

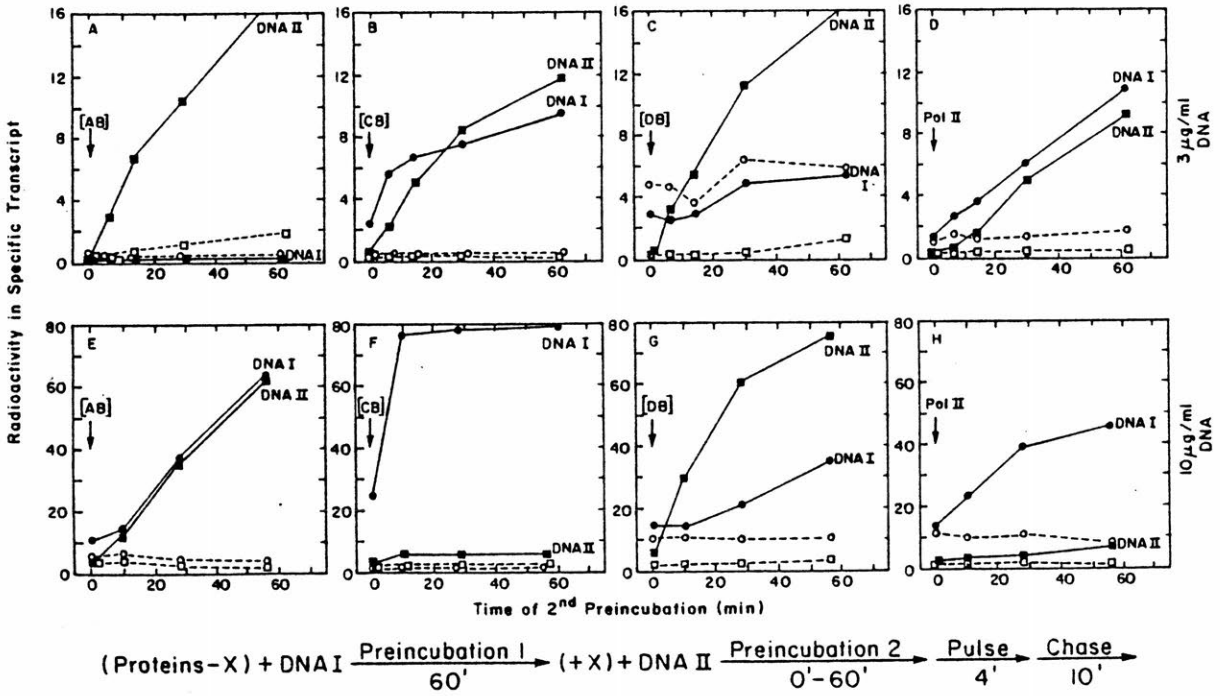
The results with [AB] and [DB] were unanticipated: in both cases activation of the first template was never observed while activation of the second template occurred with the standard time course (Fig. 6A,C). Thus, at 3 μg/ml of DNA, prior preincubation in the absence of [AB] or [DB] rendered the first template inaccessible. This effect could be titrated out by increasing the DNA concentration in the first preincubation: prior preincubation of template I at 40 μg/ml in the absence of [AB] or [DB] did not prevent (or significantly reduce) activation of that template in a subsequent complete preincubation (data not shown).

To investigate the behavior of the system under high template (interference) conditions, a similar series of experiments was performed with 10 μg/ml of each template (Fig. 6, panels E-H). One should first note that under these conditions, a time dependent activation of template I was observed when each fraction was added. Again the activation occurred most rapidly in the case of [CB]. At this intermediate template concentration, activation of template I in the case of [DB] was only partial (panel G).

As was shown in Fig. 4B, a complete preincubation with 10 μg/ml of template I interfered with the ability of the mix to activate a second added template. This interference was also observed when either [CB] or PolII were omitted from the first preincubation (Fig. 6F,H). However, no interference was observed when [AB] or [DB] was omitted

Figure 6. Template recovery after incomplete preincubation

In each panel, template I (PvuI cleaved pFLBH) was incubated in the absence of one component for 60 min, after which the second template was added with the omitted component, and a further (second) preincubation of 0-60 min allowed. Radioactive incorporation into the two resulting runoff transcripts (DNA_I, 841n [-●-] ; DNA_{II}, 974n [-■-]) after a 4 min pulse and 8 min chase was quantitated as described in Methods. The dotted lines represent identical protocols except that the indicated component was never added. Panels A-D were performed with 60 ng of each template; panels E-H were performed with 200 ng of each template.



from the first preincubation (Fig. 6E,G). This suggests that [AB], [DB] and DNA are necessary to establish interference. Indeed these components also were sufficient to establish interference. Preincubation of template I (10 μg/ml) with just [AB] and [DB], followed by a second preincubation with template II and all other components, led to activation of the first, but not the second template (data not shown).

These interference results indicate that a functional complex is formed during preincubation of [AB] and [DB] with template. Interference by excess template is relatively slow; the first template must be preincubated with the protein components in order to interfere with subsequent utilization of template (Fig. 4B,D). The use of a more rapid inhibitor not dependent on prior preincubation might provide a more applicable assay for complex formation. For such experiments, templates I and II are kept at low concentrations (3 μg/ml) so that template mediated interference was not observed (cf Fig. 4A).

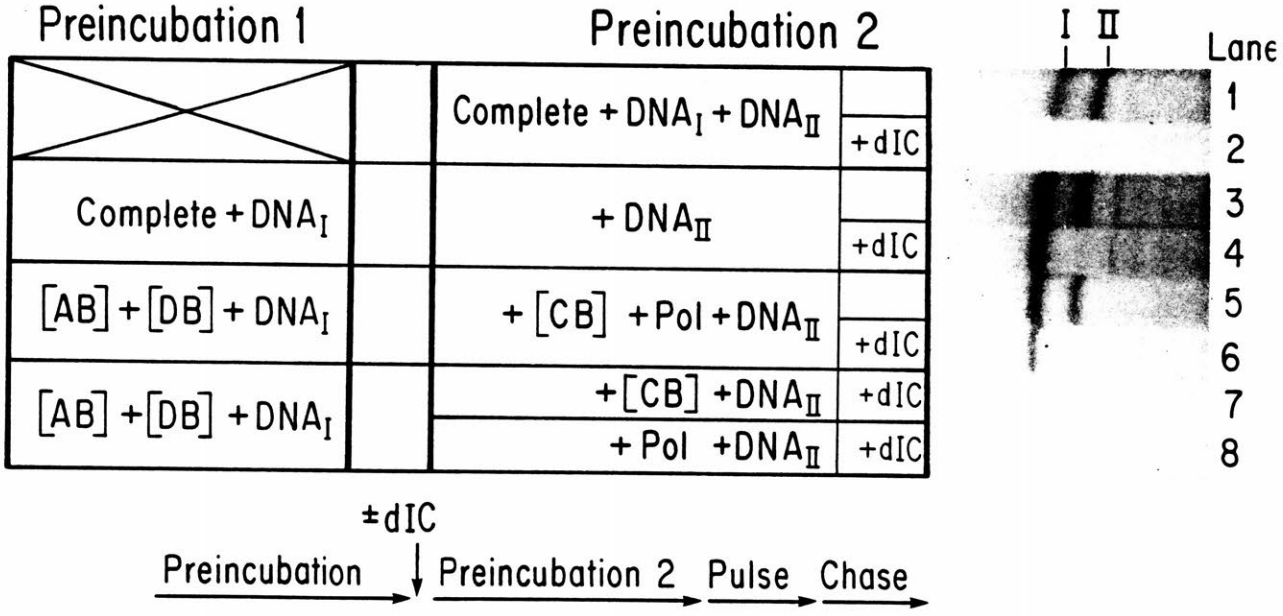
Poly (dI-dC:dI-dC) is a rapid and potent inhibitor of transcription in the reconstituted system. The resistance of preincubated templates to this inhibition can be used to define an early reaction intermediate. As shown in Figure 7A (lanes 1,2), addition of poly (dI-dC:dI-dC) during the preincubation phase totally inhibited transcription. Preincubation of proteins with template I followed by addition of poly (dI-dC:dI-dC) with template II resulted in activation of the first but not the second template (lane 4). If poly (dI-dC:dI-dC) was not added, signals from both templates were observed (lane 3). Prior preincubation of template I with just [AB] and [DB] was sufficient to render this template resistant to poly (dI-dC:dI-dC) (lanes 5,6). This demonstrates that formation of a functional template associated complex occurs in the presence of just [AB] and [DB]. If either [AB] or [DB] was omitted from the first preincubation and added with poly (dI-dC:dI-dC) to the second preincubation, no transcription was observed (data not shown). Note that poly (dI-dC:dI-dC) did not block the action of polymerase II or [CB]; these two fractions were added simultaneously with inhibitor in lane 6; their omission from the reaction (lanes 7,8) resulted in a loss of signal.

The time courses in figure 6 suggested that polymerase might

Figure 7.

Panel A: Inhibition by poly (dI-dC:dI-dC) defines a stable pre-initiation complex.

Lane 1: complete preincubation (60 min) with 60 ng of template I (PvuI-cut pFLBH) and 60 ng of template II (PstI-cleaved pFLBH) followed by a 5 min pulse and 10 min chase. Lane 2: same as lane 1 with 200 ng of poly (dI-dC:dI-dC) added at the beginning of the preincubation. Lanes 3-8 were all two stage preincubations with template I present in the first (60 min; 15 μl) preincubation and template II added at the start of a second (60 min; 25 μl) preincubation. These were followed by a 5 min pulse and 10 min chase. 200 ng of poly (dI-dC:dI-dC) was added with the second template in lanes 4, 6, 7, and 8. Lanes 3, 4: all protein components present during the first preincubation. Lanes 5, 6: only [AB] and [DB] present during the first preincubation; [CB], [CD], RNA polymerase II, and RNasin added with the second template. Lanes 7, 8 same as lane 6 except that polymerase (lane 7) or CB (lane 8) was never added.



associate directly with template complexes formed by incubation of template with [AB] and [DB]. Experiments to test this directly involved two parallel preincubations, differing only in the point of polymerase addition (Fig. 7B). Two mixes containing different templates were preincubated separately for one hour, each mix containing [AB] and [DB]. At the beginning of the pulse the two separate preincubations were mixed and [CB] was added. Polymerase was added to either of the two preincubations (lanes 3-4), to the pulse (lane 1), or not at all (lane 2). Addition of polymerase to the pulse produced a low level signal from both templates (lane 1). This signal was stimulated 3 fold when polymerase was present during preincubation; only the [template:AB:DB] mixture preincubated with polymerase was subject to this stimulation (lanes 3,4). This suggests that polymerase forms a functional association with a template complex in the absence of [CB].

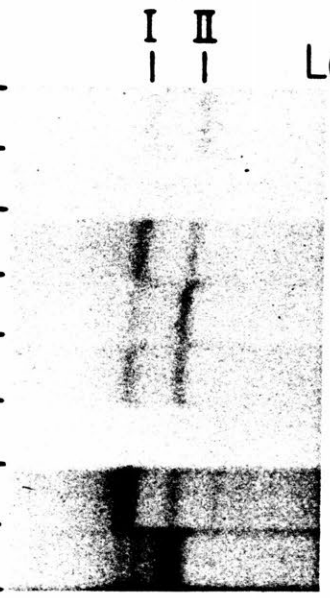
A similar protocol was used to test whether the stimulation by [CB] in the preincubation was template specific. Each template was preincubated with a mix of RNA polymerase II, [AB] and [DB]. Again, the two preincubations were mixed before pulse and chase. CB was added to either of the two preincubations (Fig. 7B, lanes 7,8) in the pulse (lane 5) or not at all (lane 6). With [CB] in the pulse, both templates were utilized comparably. Addition of [CB] to either preincubation specifically stimulated activation of the corresponding template.

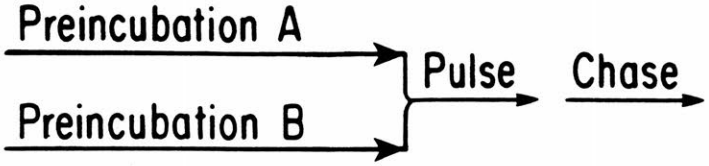
Figure 7.

Panel B: Polymerase associates with template in the absence of CB.

For each lane, templates I and II (200 ng each) were preincubated separately (60 min; 15 μl), each with [AB], [DB], [CD] and RNasin. The two preincubations were mixed at the start of the pulse (5 min) and chased (10 min). Polymerase (40 units) and [CB] were added as follows: lanes 1-5, [CB] in pulse; lane 6, no [CB]; lane 7, [CB] in template I preincubation only; lane 8, [CB] in template II preincubation only; lane 1, polymerase in pulse; lane 2, no polymerase; lane 3, polymerase in template I preincubation only; lane 4, polymerase in template II preincubation only; lanes 5, 6, 7, 8, polymerase in both preincubations. (Addition of polymerase during the pulse stimulated signal above the negative control with no polymerase to 5% of maximum. This experiment differs from that in figure 4, lanes 11, 12 in that the pulse was longer, more polymerase was used, and the [CB] was added during the pulse.)

Preincubation A [AB] + [DB] + DNA_I +	Preincubation B [AB] + [DB] + DNA_{II} +	Pulse NTP +	Lane
		Pol + [CB]	1
		[CB]	2
Pol		[CB]	3
	Pol	[CB]	4
Pol	Pol	[CB]	5
Pol	Pol		6
Pol + [CB]	Pol		7
Pol	Pol + [CB]		8





DISCUSSION

An immediate goal in the resolution and purification of the factors in a multi-component enzyme system is to determine the sequential steps in which each factor participates. Specific transcription by RNA polymerase II, assayed on the adenovirus major late promoter, has been shown to involve at least five separable components, including the polymerase (5-8). Because the purified polymerase carries out nonspecific initiation and elongation on denatured DNA templates (1), this enzyme is probably responsible for initiation and elongation during specific transcription as well. To date, no role has been assigned to the factors. The inability of purified polymerase to initiate accurately and its low activity on double stranded templates suggest that some of the factors could play a role in directing the polymerase to initiate at the proper site, or in allowing transcription from double stranded regions.

In order to address these possibilities, some knowledge of the sequence of events leading to initiation is needed. Using a three stage transcription protocol, we have identified an "activated" DNA-protein complex whose formation is nucleotide independent and precedes initiation of transcription. The criterion defining this "activated" complex was its ability to rapidly commence faithful transcription when presented nucleotide precursors. Formation of the "activated" complex required preincubation of both polymerase and factors with the DNA template. Controls involving two DNA templates were used to show that the observed activation was both DNA dependent and specific to the preincubated template. In addition, the "activated" complex was resistant to concentrations of a synthetic DNA which would completely inhibit the reaction when added simultaneously with factors and polymerase.

A number of intermediate complexes can also be detected using activation and interference assays. First, neither polymerase nor fraction [CB] (containing two of the identified transcription factors) is necessary for the formation on template of a complex resistant to inhibition by excess DNA. The other two fractions ([AB] and [DB]) are sufficient to mediate this step, forming an intermediate complex which is stably associated with the preincubated template. Second,

polymerase II will associate with this intermediate complex in the absence of the required activities in fraction [CB]. Third, some component in [CB] interacts with the template complex in the absence of nuleotides to allow maximal activation. Thus, [CB] activities can apparently act late in the formation of an "activated" complex. These data suggest the model pathway shown below.

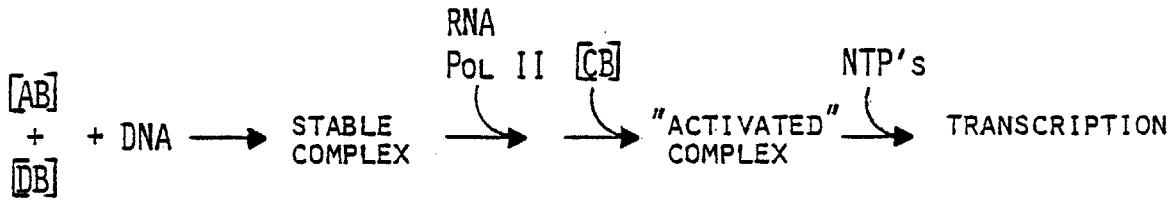


Figure 8. A model for kinetic steps in accurate transcription by RNA polymerase II.

Several important questions are not addressed by these experiments. First, the fractions used in these experiments are not purified activities and thus results can be complicated by the presence of extraneous proteins. Several transcriptional factors that act at different stages of the reaction may be present in any one fraction and thus complicate the analysis. However, the general conclusions that template selection can occur in the absence of polymerase and that at least one factor can act after polymerase has associated with the complex, would not be negated by this complication. Second, although various fractions are required for formation of intermediate and activated complexes, components from these fractions may not be present in the complexes. In the case of RNA polymerase II, presence in the activated complex is strongly indicated by association of enzymatic activity with the DNA template. Finally, the reaction order suggested in figure 8 is a permitted order: these results do not establish that a given step need depend mechanistically on previous steps.

One can speculate that stable complex formation by [AB] and [DB]

represents promoter recognition. Although our interference results demonstrate that this complex is template associated, there is no direct evidence for binding at the promoter site. Using a similar protocol, Davison et al. (25) have tested interference by a series of specific DNA fragments around the conalbumin promoter. Their results define a complex which appears equivalent to that defined by resistance to high concentrations of DNA in these studies. The correlation, although not perfect, between transcription and interference activities of the different conalbumin fragments suggests that an element near the "TATA" sequence upstream of the initiation site is involved in forming resistant complexes.

At this stage of analysis, the multi component processes leading to specific transcription by RNA polymerases II and III appear similar. Transcription of 5S genes requires three fractions in addition to RNA polymerase III; these factors are distinct from the RNA polymerase II factors (15,5). A 5S specific transcription factor has been purified and shown to bind to internal sequences in the gene (13,14). A similar analysis of sequence specific recognition events for the RNA polymerase II system will be possible with further purification of factors. Our present hypothesis is that the relevant factors are present in fractions [AB] and [DB].

ACKNOWLEDGEMENTS

We are grateful to R. Kingston, U. Hansen, N. Crawford, R. Padgett and P. Jat for critical discussions; to L. Corboy and M. Esteve for expert technical assistance; to the MIT Cell Culture Center for preparation of HeLa cells; to M. Siafaca for preparation of the manuscript; and to members of the Sharp lab for their continued advice and support.

REFERENCES

1. Roeder, R. G. (1976) in RNA Polymerase (Losick, R. and Chamberlin, M., eds.), pp 285-292, Cold Spring Harbor Press, New York.
2. Weil, P. A., Luse, D. S., Segall, J., and Roeder, R. G. (1979) Cell 18, 469-484.
3. Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980) Proc. Natl. Acad. Sci. USA 77, 3855-3559.
4. Manley, J. L., Fire, A., Samuels, M., and Sharp, P. A. (1983) Methods in Enzymology, in press.
5. Matsui, T., Segall, J., Weil, P. A., and Roeder, R. G. (1980) J. Biol. Chem. 255, 11992-11996.
6. Tsai, S. Y., Tsai, M.-J., Kops, L. E., Minghetti, P. P., and O'Malley, B. W. (1981) J. Biol. Chem. 256, 13055-13059.
7. Samuels, M., Fire, A., and Sharp, P. A. (1982) J. Biol. Chem. 257, 14419-14427.
8. Dynan, W. and Tjian, R. (1983) Cell 32, 669-680.
9. Chamberlin, M. (1976) in RNA Polymerase (Losick, R., and Chamberlin, M., eds.), pp. 17-67, Cold Spring Harbor Press, New York.
10. Wang, J. C., Jacobsen, J. H., and Saucier, J.-M. (1977) Nucleic Acids Res. 4, 1225-1241.
11. Eron, L. and Block, R. (1971) Proc. Natl. Acad. Sci. USA 68, 1828-1832.
12. Majors, J. (1975) Nature (London) 256, 672-674.
13. Engelke, D. R., Ng, S.-Y., Shastry, B. S., and Roeder, R. G. (1980) Cell 19, 717-728.
14. Pelham, H. R. B. and Brown, D. D. (1980) Proc. Natl. Acad. Sci. USA 77, 4170-4174.
15. Segall, J., Matsui, T., and Roeder, R. G. (1980) J. Biol. Chem. 255 11986-11991.
16. Sakonju, S., Brown, D. D., Engelke, D., Ng, S.-Y., Shastry, B. S. and Roeder, R. G. (1981) Cell 23, 665-669.
17. Weiner, J. H., McMacken, R. and Kornberg, A. (1976) Proc. Natl. Acad. Sci. USA 73, 752-756.
18. Blackburn, P., Wilson, G., and Moore, S. (1977) J. Biol. Chem. 252, 5904-5910.
19. Hodo, H. G., and Blatti, S. P. (1977) Biochem. 16, 2334-2343.
20. Slattery, E., Dignam, J. D., Matsui, T., and Roeder, R. G. (1982). J. Biol. Chem. 258, 5955-5959.
21. Samuels, M., Fire, A., and Sharp, P. A., submitted.
22. Kadesch, T. R. and Chamberlin, M. J. (1982) J. Biol. Chem. 257, 5286-5295.
23. Krakow, J. S., Rhodes, G., and Jovin, T. M. (1976) in RNA Polymerase (Losick, R. and Chamberlin, M., eds.), pp. 285-292, Cold Spring Harbor Press, New York.
24. Kornberg, A. (1980) in DNA Replication, pp. 347-570, W. H. Freeman, San Francisco.
25. Davison, B. L., Egly, J.-M., Mulvihill, E. R., and Chambon, P. (1983) Nature (London) 301, 680-686.
26. Ziff, E. and Evans, R. (1978) Cell 15, 1463-1475.

Chapter VI

USE OF DINUCLEOTIDES TO PROBE INITIATION BY RNA POLYMERASE II*

*This work was performed in collaboration with Mark Samuels, who is the principal author.

Summary:

Mammalian RNA polymerase II was shown to utilize dinucleoside monophosphates for priming of promoter specific RNA's. In a reconstituted system containing purified polymerase and HeLa cell fractions, dinucleotides were incorporated by complementarity with template sequences at the in vivo cap sites of the adenovirus major late and EIV promoters. Incorporation was shown by label transfer experiments and by determining the size of 5' terminal RNase T1-resistant oligonucleotides. All 16 dinucleotides were tested for priming of RNA chains at the major late promoter. RNA polymerase II initiated with various primers over a contiguous region of 9 bases, centered around the in vivo initiation site. We suggest that the polymerase drifts or oscillates over this region. Using a dinucleotide challenge protocol, the rate of initiation at the major late promoter was measured following preincubation of the template DNA with RNA polymerase II and factors. Initiation with ATP was 90% complete within the first minute after addition of nucleoside triphosphates. Stimulation of transcription by dinucleotides was not observed, due to this rapid initiation. The 5' hydroxyl terminus of dinucleotide primed RNA's remained unmodified. Although transcripts initiated with ATP were rapidly capped in whole cell extracts, ATP-primed RNA synthesized in the reconstituted system retained free 5' terminal phosphates. Thus, capping was not essential for synthesis of long runoff RNA's.

INTRODUCTION

In recent years, transcription mediated by eukaryotic RNA polymerase II has been faithfully reproduced *in vitro* (1,2). The primary criterion for this reaction has been the formation of a correct 5' end, as determined by comparison with the sequence of *in vivo* RNA. Analysis of mutants has shown that many of the same sequences promote initiation of transcription both *in vivo* and *in vitro* (3-8). However, little is known of the details of the reaction. Several factors are required in addition to purified RNA polymerase II to obtain accurate transcription (9-12), suggesting that the complete reaction pathway is complex.

We have begun an analysis of the transcription reaction, using a reconstituted system consisting of purified RNA polymerase II and partially purified HeLa cell protein fractions (12). The transcriptional activities in these fractions each titrate linearly with protein concentration and yield single peaks of activity after sedimentation through sucrose gradients. Employing a preincubation-pulse-chase protocol, several steps in the transcription reaction have been resolved (13). Components of the transcriptional apparatus can associate with the DNA to form template-specific "activated" complexes during preincubation in the absence of nucleotides. During the pulse, these complexes rapidly incorporate radioactive nucleotides into RNA chains. In addition, the activated complexes are resistant to a challenge by inhibitory concentrations of DNA. Stable protein-DNA complexes have also been observed in template competition experiments with eukaryotic RNA polymerases II and III (14-16).

A variety of physical and kinetic studies of *E. coli* RNA polymerase have revealed a multi-step pathway leading from free enzyme to initiation (17). Several intermediates occur following binding of RNA polymerase to template DNA (18,19); the final "open" complex appears to involve invasion of the double helix by the enzyme (20-22). The mechanism of initiation by E. coli RNA polymerase has been studied by the use of special primers in place of nucleoside triphosphates. The bacterial RNA polymerase can utilize a variety of short oligonucleotides to start RNA transcripts (23,24). In particular,

dinucleotides specifically stimulate initiation at promoter sites having complementary sequences; different dinucleotides can be used to alter the 5' termini of transcripts from a given promoter (25-28).

Eukaryotic RNA polymerases I and III have recently been shown to use dinucleotides as RNA chain initiators (29,30)[1]. As yet, incorporation of dinucleotides by RNA polymerase II at promoter sites has not been reported. However, the apparent stuttering of this enzyme during transcription of the polyoma early region suggests that RNA polymerase II can carry out transcription with oligonucleotide primers (31).

Transcription by RNA polymerase II differs from transcription by the other eukaryotic RNA polymerases in that the RNA polymerase II transcripts are modified by the addition of a cap (32). A majority of the promoter-specific RNA's synthesized in soluble cell extracts are also capped (1,2). RNA's synthesized in isolated cytoplasmic polyhedrosis virions are similarly capped; in this system transcription and capping appear to be mechanistically coupled (33). The rapid kinetics of capping heterogeneous nuclear RNA's *in vivo* (34), together with the observation that RNA polymerase II transcription is inhibited *in vitro* by S-adenosylhomocysteine (35), have led to the speculation that cap formation is obligatory for transcription.

We describe here the use of dinucleoside monophosphates to probe transcription initiation by RNA polymerase II.

Materials

Unlabeled and [α-^{32}P] labeled HPLC-pure nucleoside triphosphates were purchased from ICN and NEN, respectively. Dinucleotides were from Sigma, ICN or Boehringer-Mannheim. The identities of all dinucleotides were confirmed by thin-layer chromatography on polyethylene-imine-cellulose, both in 0.1 N acetic acid and in 0.16 M LiCl (36). RNases T1, T2, and U2 were purchased from Calbiochem, RNase A from Sigma, and Proteinase K from Boehringer-Mannheim. Nitrocellulose filters, with 25 mm diameter and 0.45 μ porosity, were from Schleicher and Schuell. *E. coli* infected with phage M13 mp10 or M13 mp11 were a kind gift of J. Vieira and J. Messing. End-labeled marker tRNA was a gift of Harold Drabkin.

Methods

Transcriptional proteins. Glycerol gradient purified calf thymus RNA polymerase II (37), HeLa whole cell extracts(2) and chromatographic fractions [AB], [CB] and [DB] were prepared as previously described. The polymerase had a specific activity of 1-2x10^5 U/mg (1 unit=1 pmol UTP incorporated in 20' at 37^{o}C with denatured calf thymus DNA template). Kadesch and Chamberlin have shown that RNA polymerase II purified by a similar method and having a comparable specific activity contains from 15-25% active molecules (38). Fraction [CD], a Mr=116,000 ADP-ribosyl transferase (39), reduced non-specific background without affecting signal strength, and was routinely included in reactions.

DNA templates. Plasmid pFLBH contains adenovirus type 2 sequences around the major late promoter cap site (14.7 to 17.1 map units) inserted between the Bam HI and Hind III sites of pBR322. pFLBH was cut with Pvu I (+841) or Pst I (+974) to generate templates for synthesis of runoff transcripts. pHinI[5] contains Ad5 DNA from 97.1 (Hind III) to 100 map units (Eco RI linker) around the EIV promoter[2], inserted between the Hind III and Eco RI sites of pBR322. Cleavage of pHinI[5] with Hind III yielded the template for 660 nucleotide EIV runoffs.

M13 probes. The Xho I-Hind III fragment of pFLBH (-261 to +197 of the major late promoter) was inserted between the Sal I and Hind III

sites of both M13 mp10 and mp11. Phage M13 XH11 (the mp11 recombinant) contains the strand complementary to major late promoter RNA, while M13 XH10 (the mp10 recombinant) contains the opposite strand. For hybridization to in vitro synthesized RNA, recombinant phage DNA was purified by CsCl-gradient ultracentrifugation of protease-digested phage particles. This density gradient step was important for reproducible hybridization by different phage preparations[3].

The Eco RI-Xma I fragment of pHinI[5] (-331 to +250 of the EIV promoter) was inserted between the Eco RI and Xma I (Sma I) sites of both M13 mp10 and mp11. The resulting recombinants (M13 XE10 and M13 XE11) were prepared as phage DNA's. M13 XE11 hybridizes to EIV RNA, while M13 XE10 contains the opposite DNA strand.

In vitro transcriptions. The 3-stage protocol is described in detail in the accompanying paper (13). Briefly, RNA polymerase II (20 units), fractions AB (1 μl), CB (3 μl), CD (3 μl), and DB (2 μl) were preincubated for one hour with template DNA (at 10 μg/ml) in 20 μl. 5 μl of pulse mix were then added, giving final concentrations of 30 M ATP, CTP, and UTP and 1 μM [α-^{32}P]GTP (3000 Ci/mmol). In some cases the pulse phase also contained one of the 16 dinucleotides, at 2 mM final concentration. After a short pulse of three or four minutes, 5 μl of chase nucleotides were added to give final concentrations of 1 mM for each of the four NTP's. Concentrations of glycerol and salts were maintained through each phase of the reaction. The chases were in all cases adequate for complete elongation of runoff transcripts.

Analysis of RNA. Reactions were stopped and the nucleic acids purified as described previously (40). Two rounds of ethanol precipitation were required to remove the large amount of unlabeled nucleotides added in the chase.

Short runoffs were analyzed on 0.2 mm thick, 8% polyacrylamide-urea sequencing type gels. Long runoffs were resuspended in 10 μl of 50 mM Hepes-NaOH pH 7.0, 1 mM EDTA for M13 analysis. M13 selection was performed according to U. Hansen et al.[4] (see Fig. 1B). To each reaction, 0.2 g of recombinant M13 phage DNA was added, and NaCl was added to 0.75 M. Hybridization mixtures were heated to 70°-80°C for 5 min and then incubated at 50°C for 2-5 hrs. Hybridizations were stopped with 200 μl of cold quench buffer (10 mM Hepes-NaOH pH 7.5, 1

mM EDTA, 0.2 M NaCl). Hybrids were treated with 5 units of RNase T1, a single-strand RNA specific endonuclease, for 30 minutes at 30°C; the T1 was subsequently destroyed by digestion with 50 μg of proteinase K for 60 min at 30°C.

The hybrids were filtered slowly (about 1 ml/min) through nitrocellulose filters. Filters were washed with 5-10 mls of quench buffer. The bound nucleic acids were eluted by boiling for 5 min in 1.5 mls of 2 mM EDTA pH 7.0 + 25 μg of carrier tRNA, and cooled rapidly in ice water. The nucleic acids were precipitated twice from ethanol, and resuspended either in 80% formamide, 0.05 M Tris-borate pH 8.3, 1.25 mM EDTA for direct acrylamide gel electrophoresis, or else in 10 mM Tris pH 7.4, 1 mM EDTA for further analysis.

For preparation of T1 oligonucleotides, the M13 selected RNA dissolved in 2 mM EDTA pH 7.0 was heated for 5 min at 90°C, quick chilled in ice water and digested with 5-10 units of RNase T1 for 2 hours at 30°C. 50 μl of water freshly saturated with diethyl pyrocarbonate was added to inactivate the RNase, and reactions were lyophilized three times with 50 μl water washes. The resulting oligonucleotides were resuspended in 80% formamide, 0.05 M Tris-borate pH 8.3, 1.25 mM EDTA and electrophoresed on 0.2 mm thick, 15% polyacrylamide-urea gels.

For RNase U2 digestion, RNAs dissolved in 2 mM EDTA pH 7.0 were digested with RNase T1 as above. After the T1 incubation, sodium citrate pH 3.5 was added to 20 mM. After another round of heating and rapid cooling, 0.02 unit of RNase U2 was added for 30 minutes at 30°C. The digestion products were lyophilized, dissolved and electrophoresed as described above.

Decapping was performed by periodate oxidation and β-elimination with lysine (as described, 41) before M13-selection. Phosphatase treatments were performed after M13-selection, with 16 units of bacterial alkaline phosphatase (Worthington) for 45 min at 55°C in 10 mM Tris pH 7.9, 1 mM EDTA, 0.05% SDS. For unknown reasons, late promoter RNA was highly resistant to phosphatase at 30°C.

For label transfer analysis, oligonucleotides from complete T1 digestion of preincubation-pulse-chase RNA's were gel purified, and digested either with RNase T2 (1 unit, 37°C overnight in 20 mM ammonium

acetate pH 4.5) or with RNase A (0.4 μg, 37°C for 60 minutes in 10 mM Tris pH 7.4, 1 mM EDTA). The digestion products were resolved by chromatography on cellulose thin layers in one or two dimensions (42).

To confirm that the decapping reaction had worked, [α-^{32}P]UTP-labeled late promoter RNA was synthesized by a whole cell extract in a simple 3 hour reaction. The 5' terminal 41 nucleotide fragment was isolated as described (43). Aliquots were removed for decapping and/or phosphatase treatment. The products of complete RNase T2 digestion were resolved by two-dimensional thin-layer chromatography. From untreated, decapped, and decap + phosphatase reactions, spots were observed with the mobilities expected for GpppAmCp, pppAmCp, and AmCp, respectively (2'-O-methylation protected the A from T2 cleavage). A spot possibly consisting of m^6AmCp was also observed.

Quantitation. Autoradiograms obtained with response-linearized (preflashed) film were quantitated by densitometry and planimetry (12). For estimates of absolute transcriptional efficiency, T1 oligonucleotide bands were excised from acrylamide gels and counted for Cerenkov radiation. The number of RNA chains initiated in the pulse was calculated from the specific activity of the labeled nucleotide.

RESULTS

Transcription reactions were performed using a reconstituted system consisting of purified RNA polymerase II supplemented with HeLa cell fractions [AB], [CB], and [DB] (12, and see Fig. 1A). The 3-stage protocol described by Fire et al. (13) was used to resolve the reaction into several kinetic steps. As shown in Fig. 1B, RNA polymerase II, factors and template DNA were first preincubated, allowing formation of "activated" complexes. Radioactive nucleotides were then added for a brief pulse, during which elongation proceeded only a short distance. An ensuing chase with excess unlabeled nucleotides allowed the completion of long runoff RNAs.

To examine initiation by the activated complexes, pulse nucleotide concentrations were varied. The template for these experiments contained the Ad5 early region IV promoter (see Fig. 2B). Previous work with this promoter has shown that the site of initiation in vitro exhibits heterogeneity over a seven nucleotide region in the whole cell extract, thereby accurately reproducing the heterogeneity observed in vivo (44). Moreover, the relative amounts of the various termini depend on the nucleotide concentrations in the reaction. Thus, high UTP concentrations favor the use of U initiation sites (residues -6 to -1), while high ATP concentrations favor initiations at A (residue +1).

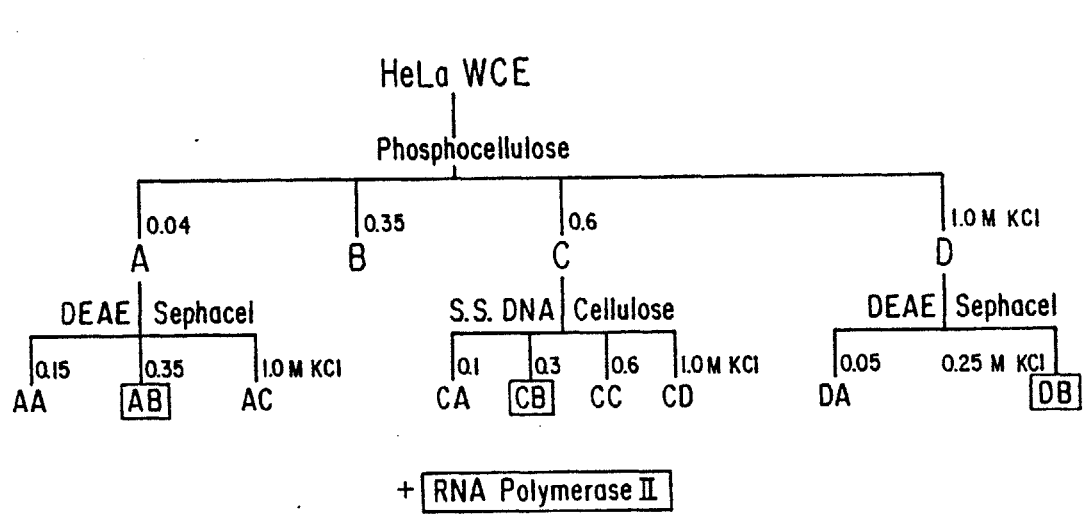
Analogous results were obtained with the reconstituted system using the 3-stage protocol. EIV runoff RNAs were synthesized under conditions of either high ATP or high UTP concentration in the pulse. These RNAs were analyzed for differences at the 5' end by the method of Hansen et al.[4] For this analysis, the long runoff RNAs were hybridized to a single strand M13 recombinant containing a segment of adenovirus complementary to the first 250 bases of the EIV transcripts (see Fig. 2A and Methods). The runoffs were truncated by ribonuclease digestion, leaving a 250-nucleotide promoter proximal RNA, which was selected by filtration through nitrocellulose. This procedure yields RNAs with common 3' termini so that length variations must reflect differences at the 5' end.

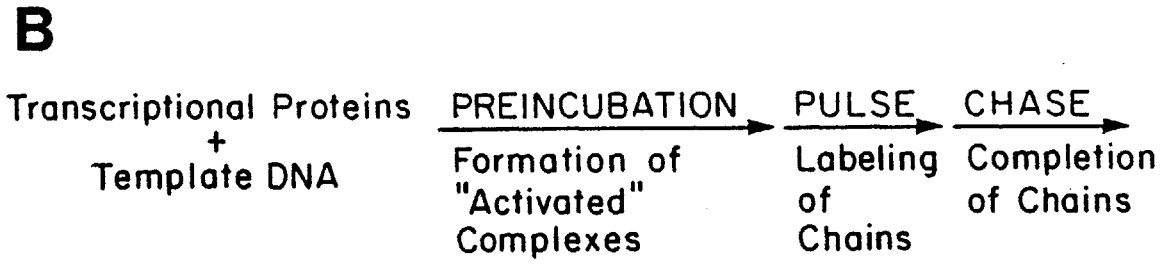
On high resolution gels, the high [UTP] reaction product migrated more slowly than the high [ATP] reaction product (Fig. 2C, lanes 2 and 3), suggesting a shift to upstream initiation sites. Direct analysis

Figure 1

A. The reconstituted system. Fractions [A], [B], [C], and [D] resulted from phosphocellulose chromatography of a whole cell extract. Fractions [A] and [D] were further separated on DEAE-Sephacel (Pharmacia), and [C] was fractionated on single-strand DNA cellulose. RNA polymerase II was separately purified from calf thymus as noted in Methods. Transcriptions used fractions [AB], [CB], [CD], [DB], and polymerase.

B. The 3-stage protocol. RNA polymerase II, [AB], [CB], [CD] and [DB] were preincubated with template DNA for one hour. Pulse nucleotides were added for 3 or 4 min pulse, after which excess unlabeled nucleotides were added to chase nascent chains into full-length runoff RNA. The elongation rate during the chase was 400-600 nucleotides/min. Runoff RNA's ranged from 660-974 bases; a three minute chase was ample for completion of chains. For exact concentrations of all components see Methods.





of short runoff RNAs gave similar results (data not shown). Since all labeled transcripts were subject to this shift, these transcripts must be initiated during the pulse phase.

Initiation with dinucleotides at the EIV promoter

At low nucleotide concentrations, E. coli RNA polymerase will utilize dinucleotides to initiate RNA chains, incorporating the dinucleotides according to base-pairing rules (24-28). To test whether RNA polymerase II behaved similarly, 5' shift experiments were performed using dinucleotides with the EIV promoter. When UpU was added to a transcription pulse, the resulting RNA migrated more slowly than the control RNA initiated at low nucleotide concentrations (Fig. 2C, lanes 1 and 7). This UpU RNA comigrated with RNAs initiated at high [UTP] (compare lanes 3 and 7), indicating that initiation in the presence of the dinucleotide had occurred in the stretch of T residues upstream of +1. A similar upstream shift was observed with CpU (Fig. 2C, lane 6), as expected from the sequence of the EIV promoter. This shift in mobility was specific to UpU and CpU, as neither ApC nor CpA visibly affected the migration of the EIV RNA (Fig. 2C, lanes 4 and 5). A shift of one nucleotide (expected if CpA priming occurred) would probably not be detected with this gel system. However, the fact that addition of either ApC or CpA stimulated labeling of the EIV RNA in a short pulse, indicates that dinucleotides probably did prime transcription at the EIV promoter.

Initiation with dinucleotides at the major late promoter

The major late promoter (MLP) is more efficient in a whole cell extract than is the EIV promoter (44). In further contrast to EIV, initiation at the MLP occurs in vivo and in vitro at a single A residue (nucleotide 6039 of adenovirus type 2). Analysis of initiation at the MLP is facilitated by the presence of a large RNase T1-resistant oligonucleotide containing the 5' terminus (45,46).

This oligonucleotide was analyzed directly using the following procedure. The MLP was transcribed using the 3-stage protocol described above. The short pulse with low concentrations of the radioactive nucleotide allowed incorporation of label only near the 5' terminus. After completion of chains during the chase, RNA was hybridized to a single stranded M13 recombinant spanning the MLP (M13

Figure 2. Analysis of M13-truncated RNA

A. M13 selection. RNA was synthesized using the 3-stage protocol with Hind III-cleaved pHinI[5] as template DNA. The long runoff (in this case 660-nucleotides) was annealed to single stranded M13 XE11 DNA. The 5'-proximal 250 nucleotides of the EIV RNA should hybridize to this M13 probe. After hybridization, the free 3' tail of the runoff was digested with RNase T1, which also hydrolyzed much of the non-specific background RNA. Hybrids were collected by filtration through nitrocellulose membranes, and the 250-nucleotide truncated RNA was resolved on an 8% sequencing-type gel.

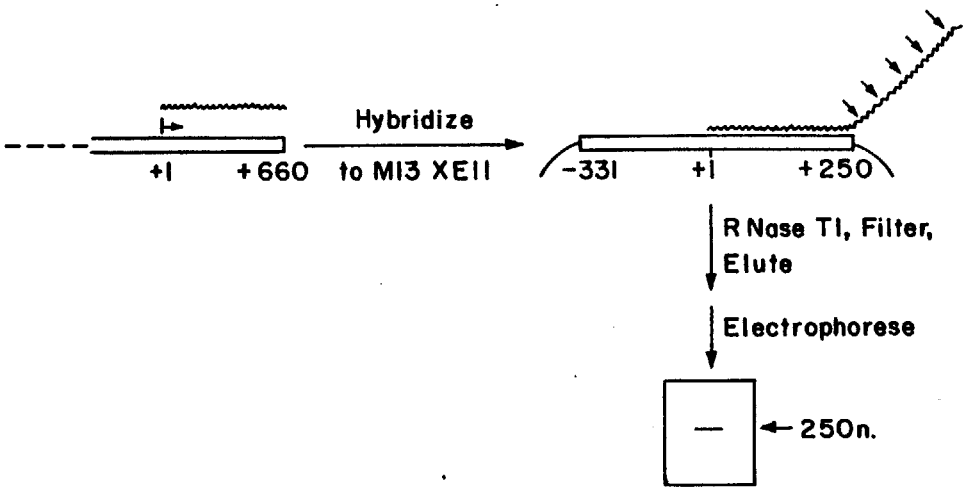
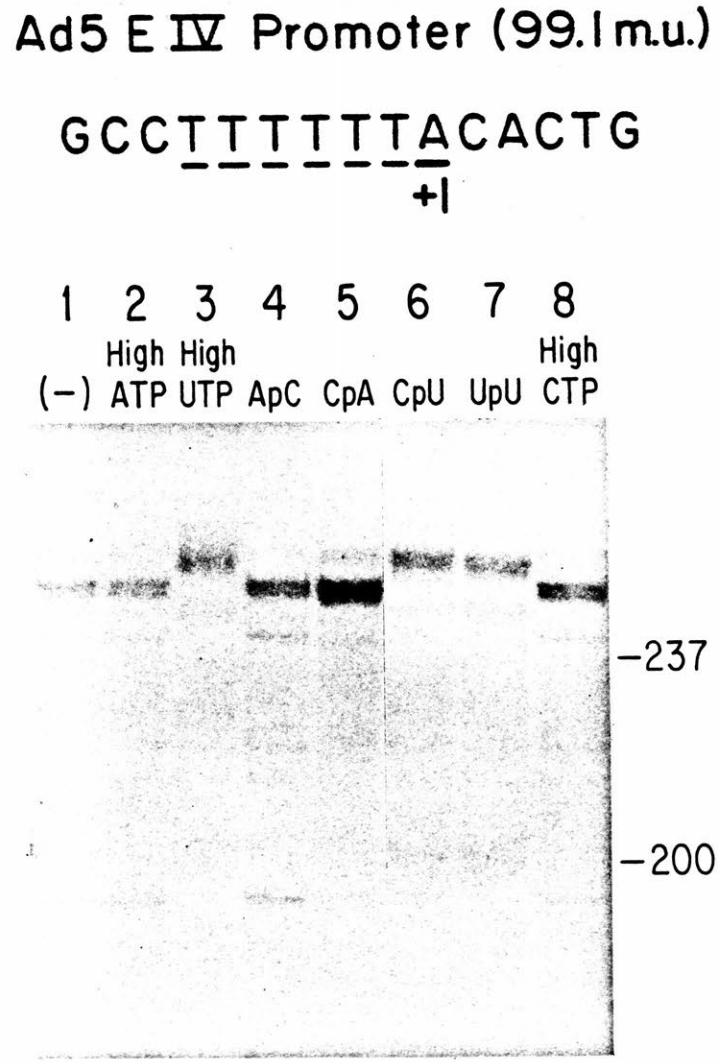


Figure 2 (continued)

B. The sequence around the Ad5 early region IV promoter. The cap sites observed in vivo are underlined, and the A start is defined as position +1 (51).

C. EIV transcription products. Long EIV runoffs were synthesized in the 3-stage protocol, and analyzed as described in panel A. The pulses contained 30 μM ATP, CTP, and UTP, 1 μM [α-^{32}P]GTP with the following additions: lane 1, no additions; lane 2, ATP to 300 μM; lane 3, UTP to 300 μM; lane 4, ApC; lane 5, CpA; lane 6, CpU; lane 7, UpU; lane 8, CTP to 300 μM. Dinucleotides were added to a final concentration of 2 mM. All of the additions increased the intensity of the EIV RNA band reproducibly (in this gel, only half of samples were loaded in lanes 2 and 3). When recombinant phage M13 XE10, containing the noncoding DNA strand, was annealed to transcription products, no labeled material was protected from ribonuclease digestion.

The 237 nucleotide marker was derived from Hinf I-cleaved, end-labeled SV40 DNA. The 200 nucleotide marker was a truncated runoff RNA from the major late promoter.



XH11, containing sequences from -261 to +197 of the MLP). Remaining single stranded RNA was digested with ribonuclease and the RNA/DNA hybrid was selected on nitrocellulose filters. RNA was released from the hybrid and digested to completion with RNase T1. The labeled oligonucleotide products yielded a simple pattern when resolved by electrophoresis in a 15% acrylamide-8M urea gel (Fig. 3A, lane 1). Five major bands, designated A-E, were observed. Bands A, B, C and D migrated in the size range of 1 to 6 bases relative to a size marker, while band E migrated in the range of 9-13 bases. Tentative assignment of these oligonucleotides is shown in Figure 3B. Since the label was [α-^{32}P]-GTP and each oligomer contained only one G residue (at the 3' end), the intensity of each band was proportional to the amount of the corresponding oligomer synthesized during the pulse. The gradient of band intensities ran E-D-B-(A,C). This agreed perfectly with the assignments based on size. Long exposure revealed a faint T1 product running at 16 nucleotides. This longer oligomer corresponded to the 16-mer ending at position +62. Its low intensity suggests that very few polymerase molecules reached that position during the pulse.

In contrast to the results obtained with the EIV promoter, varying nucleoside triphosphate concentrations had no effect on the site of initiation of the MLP. Concentrations of CTP or UTP as high as 1 mM did not result in initiation at C and T residues adjacent to +1 (data not shown).

Examination of the MLP sequence (Fig. 3B) suggested the dinucleotides ApC and CpA as good candidates for priming RNA synthesis. Addition of either dinucleotide to the pulse resulted in a shift of the putative 5' terminal oligonucleotide (Fig. 3A, lanes 2 and 3). The internal oligomers (A-D) migrated identically in each reaction. As predicted from the sequence, the CpA band (E_{12}) migrated more slowly than the ApC band (E_{11}). The ApC band, in turn, migrated more slowly than the control band E, which presumably resulted from initiation with ATP at position +1. A possible reason for the faster migration of band E was the presence of phosphates at the 5' end of this oligomer. The effect of phosphatase treatment on the mobility of these oligonucleotides was therefore investigated.

RNAs primed without dinucleotides (control), or with either ApC or

Figure 3.

A. T1 oligonucleotide analysis of major late promoter RNA. Long runoff RNA was synthesized using the 3-stage protocol with linearized pFLBH as template. The standard pulse nucleotides had the following additions: lane 1, no additions; lane 2, ApC; lane 3, CpA. RNA's were hybridized to single stranded M13 XH11 DNA, containing coding sequences between -261 and +197 of the MLP. The hybrids were RNase treated and filter-selected as in Fig. 2A. After elution from the filter, the RNA was digested to completion with RNase T1. The products were resolved on a 15% sequencing-type gel. Bands A-D refer to T1 oligonucleotides whose sequence is in Fig. 3B. Bands E, E_{11}, and E_{12} correspond to the putative 5' terminal oligonucleotides resulting from the control, ApC, and CpA reactions respectively.

B. Sequence around the adenovirus type 2 major late promoter cap site. Initiation *in vivo* occurs exclusively at the indicated A residue (45,46). Slash marks indicate sites of cleavage of RNA transcripts by RNase T1; the letters A-E designate the 5'-proximal T1 oligonucleotides (see Fig. 3A).

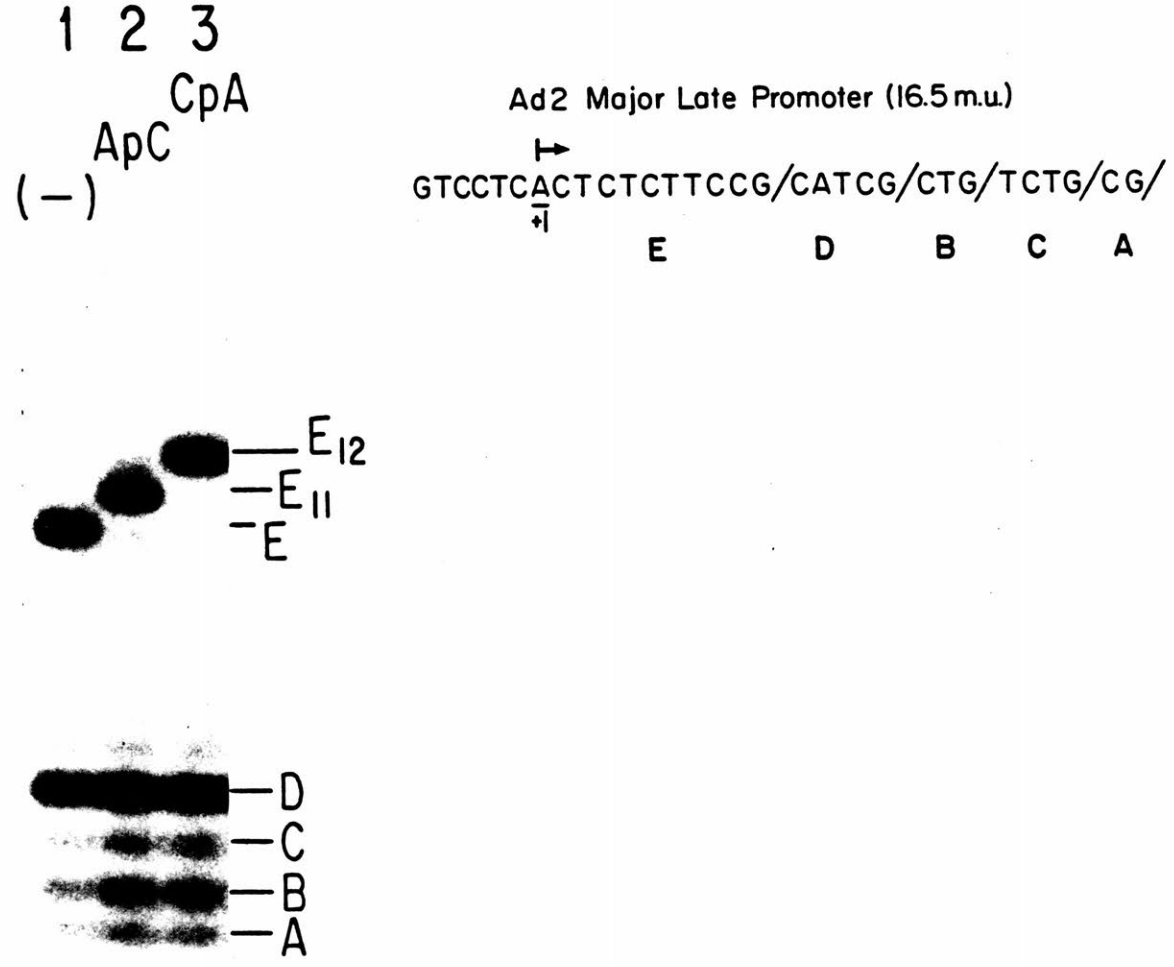
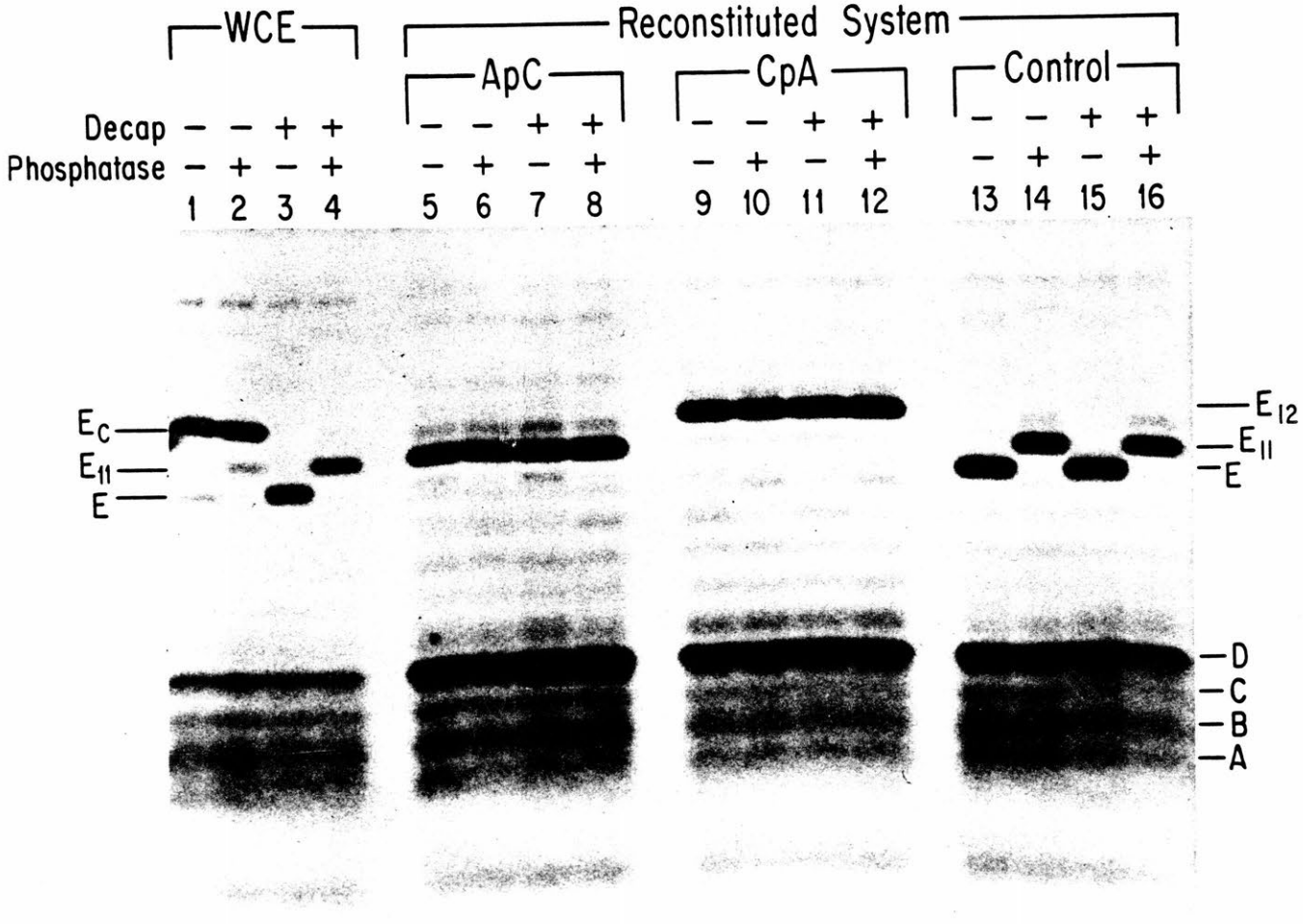


Figure 3 (continued)

C. Sensitivity of late promoter RNA's to decapping and phosphatase. Long runoff RNA's synthesized with the 3-stage protocol as in Fig. 3A were hybridized to M13 XH11, truncated with RNase T1, and filter purified. Aliquots were removed and treated with decapping and/or phosphatase reagents as in Methods. Products in lanes 1-4 were made in a whole cell extract; those in lanes 5-16 were made in the reconstituted system. The standard pulse nucleotides received either no additions (lanes 1-4 and 13-16), ApC (lanes 5-8), or CpA (lanes 9-12). Samples were treated with phosphatase (lanes 2, 6, 10, 14), decapped (lanes 3, 7, 11, 15) or decapped followed by phosphatase treatment (lanes 4, 8, 12, 16). After these treatments, the RNA's were digested to completion with RNase T1 and electrophoresed. Bands A-D refer to oligonucleotides whose sequence is in Fig. 3B. Bands E, E_{11}, E_{12}, and E_c refer to 5'-terminal oligonucleotides synthesized under the different pulse conditions.



CpA, were treated with phosphatase before complete RNase T1 digestion. In this manner, only 5' terminal phosphates should be susceptible to hydrolysis. As expected, phosphatase digestion did not affect the migration of internal T1 oligomers (Fig. 3C, bands A-D). The ApC band (E_{11}) and CpA band (E_{12}) were likewise not affected (lanes 6, 10), whereas band E synthesized in the control reaction was affected by phosphatase (lane 14); after treatment it comigrated with band E_{11}. Thus 5' terminal phosphates were present on ATP primed but not on ApC and CpA primed RNA.

The phosphatase sensitivity demonstrated that the control RNA was not capped. In contrast, previous results have shown that RNA synthesized in the whole cell extract is efficiently capped (2). The latter results were confirmed using RNA prepared in a whole cell extract by the 3-stage protocol. This RNA was analyzed by M13 selection followed by complete RNase T1 digestion. Internal oligomer bands A-D were identical with those of the reconstituted system (Fig. 3C, lanes 1, 13). However, the largest whole cell extract T1 oligomer (E_c) migrated more slowly than the corresponding band E from the reconstituted system. The mobility of E_c was not affected by phosphatase treatment (lane 2). Chemical decapping of the whole cell extract RNA prior to T1 cleavage altered the mobility of band E_c such that it comigrated with band E (lane 3). Decapped E_c was also rendered phosphatase sensitive; the combined decap + phosphatase treatment altered the mobility of band E_c such that it comigrated with ApC band E_{11} (compare lanes 4 and 5). Decapping had no effect on RNAs made in the reconstituted system (lanes 7, 11, 16).

The following structures, with initiating nucleotides underlined, were postulated for the 5' terminal oligonucleotides:

E_c (WCE, ATP-initiated): GpppACUCUCUUCCGp

E (Rec. sys., ATP-initiated): pppACUCUCUUCCGp

E_{11} (Rec. sys., ApC-initiated): ACUCUCUUCCGp

E_{12} (Rec. sys., CpA-initiated): CACUCUCUUCCGp

These T1 oligonucleotides were further examined by digestion with RNase U2, which cleaves after A and G residues. As expected, RNase U2 digestion converted each of the four oligomers, E_c, E, E_{11}, and E_{12} to a single 10-mer (data not shown). The sensitivity of the capped

Table I

Secondary analysis of 5' terminal oligonucleotides. Late promoter runoff RNA's were synthesized using the 3-stage protocol. Each sample was scaled to five standard reaction volumes. The preincubation contained Pst I-cleaved pFLBH at 20 μg/ml. Pulses contained 100 μCi of [α-^{32}P]CTP, UTP, or GTP (1 μM) plus the other three unlabeled nucleoside triphosphates at 30 μM. With each radioactive nucleotide a reaction was done with these standard pulse conditions, or with the standard pulse plus ApC or CpA. From each reaction the 5' terminal RNase T1 oligomer was gel-purified following M13-selection, and was further digested with RNase T2 or A. Products of CTP labeling were separated in two dimensions as described in Methods; products of GTP and UTP labeling were separated only in the first dimension.

Label [α-^{32}P]CTP

RNA Initiator	2° Enzyme	Expected	Observed
ApC	RNase T2	Cp,Up,Gp	Cp,Up,Gp
CpA	RNase T2	Ap,Cp,Up,Gp	Ap,Cp,Up,Gp
ATP	RNase T2	pppAp,Cp,Up,Gp	*Ap,Cp,Up,Gp

Label [α-^{32}]GTP

RNA Initiator	2° Enzyme	Expected	Observed
ApC	RNase A	Cp	Cp
CpA	RNase A	Cp	Cp
ATP	RNase A	Cp	Cp

Label [α-^{32}P]UTP

RNA Initiator	2° Enzyme	Expected	Observed
ApC	RNase T2	Cp,Up	Cp,Up
CpA	RNase T2	Cp,Up	Cp,Up
ATP	RNase T2	Cp,Up	Cp,Up

*Loss of 5' terminal phosphates in this sample apparently resulted from a phosphatase contaminant.

oligomer (E_c) to RNase U2 indicated that none of the capped A was 2'-O-methylated. The extent of methylation at the 7 position of guanine in the cap was not ascertained.

Definitive proof of dinucleotide incorporation was obtained by the following label transfer experiments. RNA was made with control nucleotides, or with added ApC or CpA in the pulse. Each reaction contained a single labeled nucleoside triphosphate ([α-^{32}P] GTP, CTP, or UTP). The 5' terminal oligomers were gel purified and further digested with RNase A (pyrimidine specific) or T2 (nucleotide non-specific). The digestion products were analyzed by thin layer chromatography. Table I shows the products expected and observed in each case.

The [α-^{32}P] CTP reactions provided the critical test. This nucleotide should transfer label to A residues only in the control and CpA reactions. In the ApC reaction, the 3' phosphate of the A residue should be provided by the dinucleotide and thus should not be labeled. These were the observed results. As expected, label was also transferred from [α-^{32}P] CTP to Up, Cp and Gp. [α-^{32}P] GTP and [α-^{32}P] UTP transferred label to the anticipated nucleotides (Table I).

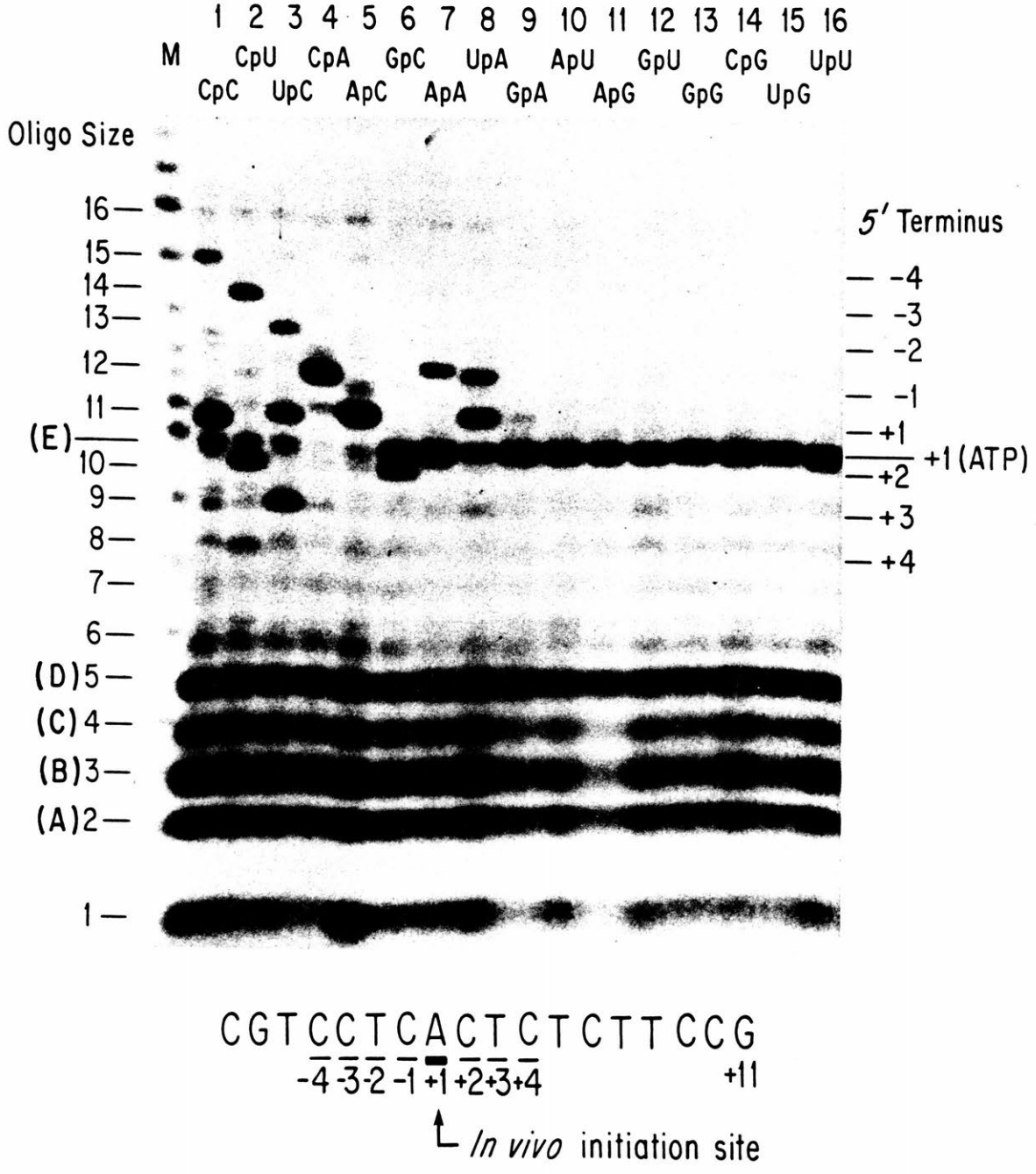
It has been suggested that the requirement for an RNA polymerase I factor can be obviated by the use of a dinucleotide primer (29). This was tested in the RNA polymerase II transcription system with the adenovirus major late promoter. The dinucleotide CpA did not substitute for any of the HeLa cell fractions [AB], [CB], or [DB] in stimulating accurate transcription mediated by RNA polymerase II (data not shown).

A region accessible for dinucleotide priming

The extent of sequences available for priming at the MLP was tested using all 16 dinucleotides. In parallel reactions, each with a different dinucleotide added to the pulse, a variety of patterns was generated. All 16 reactions yielded the same pattern of internal T1 oligomers (Fig. 4, bands A-D). The dinucleotides GpA, ApU, ApG, GpU, GpG, CpG, UpG, and UpU (Fig. 4, lanes 9-16) were not utilized as primers; only the ATP initiated 5' oligomer (band E) was observed. The remaining eight dinucleotides yielded different patterns for the 5' oligomer (lanes 1-8). As previously shown, ApC and CpA gave an 11-mer

Figure 4

Dinucleotide-primed initiations. Major late promoter runoffs were synthesized in the 3-stage protocol, and analyzed as in Fig. 3A. In each reaction a different dinucleotide was added to the pulse mix to give a final concentration of 2 mM. Lane 1, CpC; 2, CpU; 3, UpC; 4, CpA; 5, ApC; 6, GpC; 7, ApA; 8, UpA; 9, GpA; 10, ApU; 11, ApG; 12, GpU; 13, GpG; 14, CpG; 15, UpG; 16, UpU. Numbers on the left side of the gel refer to the lengths of T1 oligonucleotides; those marked (A)-(E) refer to Fig. 3A. Numbers on the right refer to DNA sequence positions of 5' termini (as noted in the figure), and relate only to 5' terminal oligonucleotides. The +1(ATP) line marks the band corresponding to a 5' terminal 11-mer initiated with ATP and presumably carrying a triphosphate 5' end. Nucleotide spacings were compared to an alkaline, formamide hydrolysis ladder of 5' end-labeled tRNA (lane M). Below the gel the sequence around the in vivo site of initiation (the A residue marked +1) is provided.



and a 12-mer, respectively, corresponding to initiation at positions +1 and -1. CpC, CpU, and UpC each yielded novel bands which were interpretable as resulting from dinucleotide incorporation according to the DNA sequence: bands of length 15 (CpC), 14 (CpU), and 13 (UpC) were observed. Also in accord with the sequence, CpU gave bands at 10 and 8 nucleotides, and UpC gave a 9-mer. These oligomers ranging from 8 to 15 bases long corresponded to initiations at each position from -4 (CpC) to +4 (CpU). No initiations further upstream or downstream were detected.

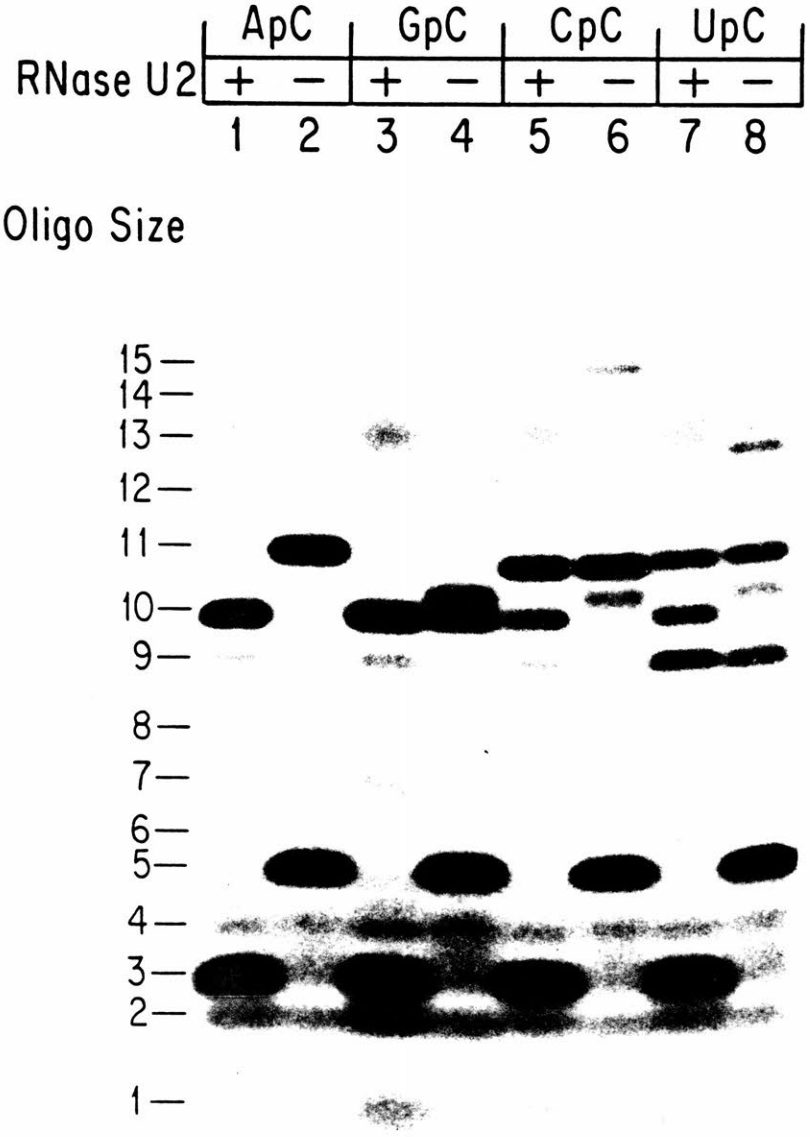
Several bands could not be explained by strict complementarity between dinucleotide and DNA. These included 11-mers generated with CpC (lane 1), UpC (lane 3) and UpA (lane 8), 12-mers generated with ApA (lane 7) and UpA (lane 8), and a 10-mer generated with GpC (lane 6). One possible explanation for these bands was that the dinucleotides were incorporated using complementarity limited to the second residue. For example, CpC and UpC might generate 11-mers by incorporation in the same position as ApC. Further digestions with RNase U2 supported this "misincorporation" hypothesis: the 11-mers made with CpC and UpC were U2-resistant (Fig. 5, lanes 5-8). As expected, U2 treatment converted the ApC-initiated 11-mer to a 10-mer by cleavage at the A residue (lanes 1 and 2). The T1 resistant GpC 10-mer was similarly explained by incorporation of GpC in the same position as ApC. In this case, however, the complete T1 digestion released the 5' terminal Gp leaving a 10-mer. This 10-mer was RNase U2 resistant as expected. The aberrant UpA and ApA oligomers were all converted to 10-mers by RNase U2 treatment (data not shown). This suggests that UpA and ApA generated 12-mers by incorporation in the same position as CpA. The UpA 11-mer is difficult to explain but could be due to contamination in this dinucleotide.

Dinucleotide titrations

Addition of dinucleotides in no case stimulated overall transcription from the major late promoter. To determine the efficiency of dinucleotide utilization relative to standard pulse triphosphates, ApC and CpU were individually titrated in the pulse phase. Complete T1 digestion products were resolved on gels, and the label in various 5' terminal oligomers was measured by densitometry.

Figure 5

RNase U2 digestions of dinucleotide-primed products. Late promoter runoffs were synthesized in the 3-stage protocol with 2 mM ApC, GpC, CpC, or UpC in the pulse. After hybridization to M13 XH11, digestion of hybrids with RNase and filter selection, RNA was eluted and digested to completion with RNase T1 [(lanes 1 (ApC), 3 (GpC), 5 (CpC), 7 (UpC)]. Aliquots from each reaction were removed and further digested with RNase U2 [(lanes 2 (ApC), 4 (GpC), 6 (CpC), and 8 (UpC)]. RNase U2 digestion was complete as monitored by the digestion of all 5-mers (which have the sequence CAUCG) to 3-mers (UCG). U2 also cleaved the CpC 15-mer, the UpC 13-mer, the ApC 11-mer, and all ATP 11-mers. The UpC 9-mer should have no A residue and was U2-resistant. The CpC and UpC 11-mers were likewise U2-resistant, proving that these 11-mers did not contain a purine at position +1. A diffuse band migrating near 13 nucleotides was observed in all U2 digestions.



As the concentration of each dinucleotide was increased, the amount of corresponding dinucleotide-primed transcripts increased at the expense of ATP initiated RNA's. Under standard pulse conditions (30μM ATP, CTP, UTP, 1μM [α-^{32}P]GTP), 50% of the total initiations were dinucleotide-primed at about 60μM ApC (data not shown). The CpU initiations at -3 and +2 increased concomitantly with CpU concentration; at about 200μM CpU the sum of initiations at these two sites was 50% of the total (data not shown).

Rate of initiation

The ability to quantitatively shift 5' termini using high concentrations of dinucleotide in the pulse suggested an experiment to analyze the actual rate of chain initiation. This "dinucleotide challenge" experiment took advantage of the different mobilities of CpA and ATP-initiated RNAs. Transcriptional factors and polymerase were preincubated with template DNA. Pulse nucleotides were added for 4 min, followed by the usual chase. The dinucleotide CpA was added in excess, either with the pulse nucleotides, or at some time during the pulse, or with the chase. The CpA and ATP initiations were resolved by complete T1 digestion of M13 selected RNA (see Fig. 6).

At 2 mM, CpA competed effectively for initiation when added at the beginning of the pulse (lane 1). However, when CpA was added after one minute of pulse, 90% of observed chains had pppA at their 5' terminus; thus, most of the observed chains had already initiated with ATP (lane 2). By three minutes of pulse almost all chains were ATP-initiated (lane 3). In a control reaction, the addition of CpA to the chase had no effect on the 5' oligomer (lane 4). Similar results were obtained when the ATP, CTP, and UTP concentrations were 300 M in the pulse, ten times higher than previously (lanes 5-8). In a separate experiment, it was shown that the incorporation of label into ATP-initiated 11-mers reached a plateau after 2-4 minutes of the pulse (data not shown). This independently confirmed that initiation occurred synchronously during the standard pulse.

The preincubation-pulse-chase protocol thus functions as a one-round assay for activated complexes. Using this assay the number of templates and RNA polymerase molecules productively involved in transcription was readily determined by quantitating the amount of

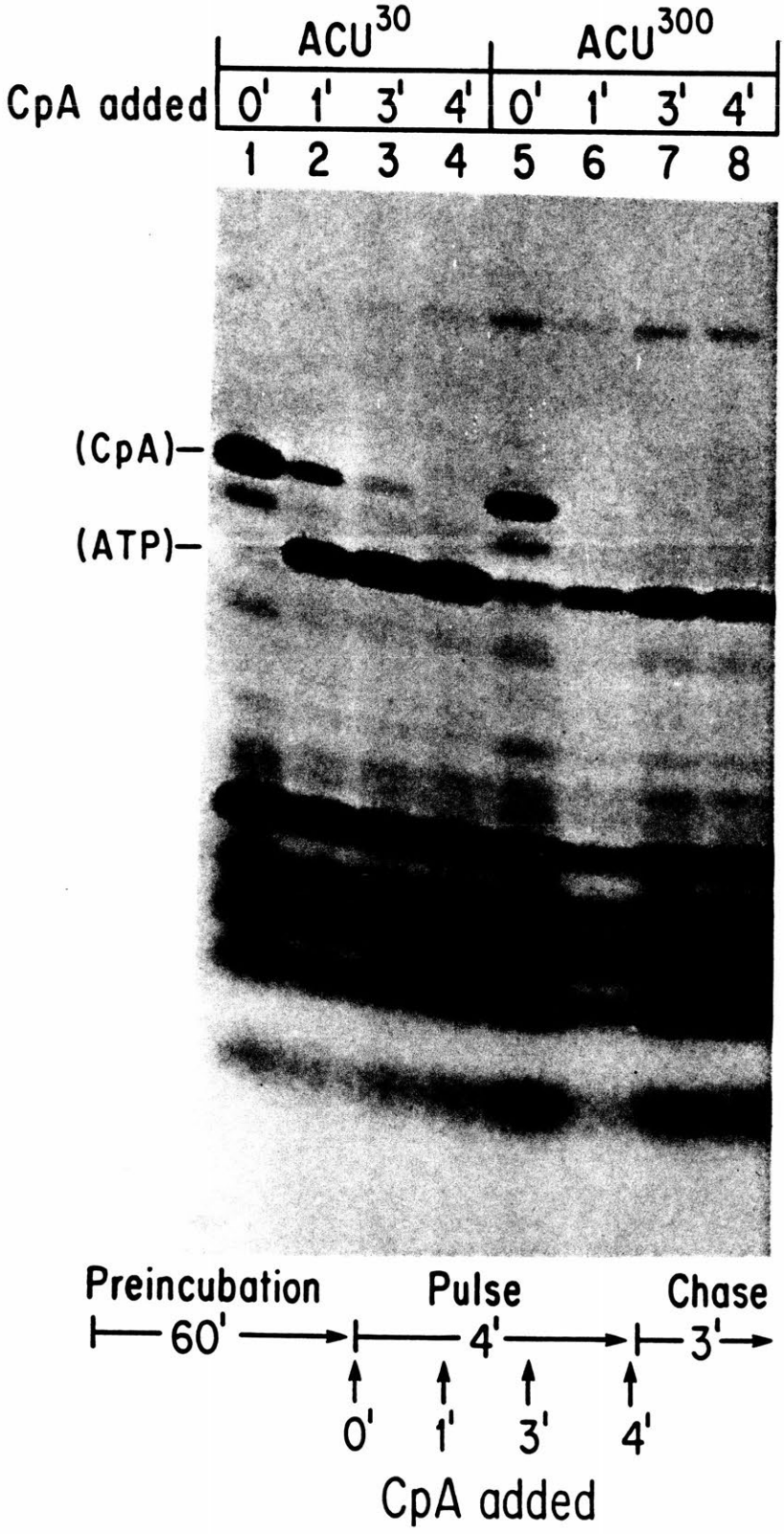


Figure 6. Dinucleotide challenge experiment

After preincubation of transcriptional proteins with a late promoter template, four minute pulses were conducted. ATP, CTP, and UTP were present during the pulse, each at 30 μM (lanes 1-4) or 300 μM (lanes 5-8) with 20 μCi of [α-^{32}P]GTP at 1 μM. Pulses were followed by three minute chases with 1 mM unlabeled nucleotides. The dinucleotide CpA was added together with pulse nucleotides (lanes 1, 5), after one minute of pulse (lanes 2, 6), after three minutes of pulse (lanes 3, 7), or with chase nucleotides (lanes 4, 8). RNA's were analyzed as in Fig. 3A. The different 5'-terminal oligomers are indicated by arrows.

radioactivity in 5' terminal RNase T1-resistant oligonucleotides. 0.1% of templates were transcribed, and 0.1% of the RNA polymerase II molecules functioned in the promoter-specific reaction. The low fraction of participating polymerase molecules did not reflect a poor preparation, as the specific activity of the enzyme was close to the maximum reported values. Kadesch and Chamberlin have directly measured the fraction of active molecules in a similar preparation of RNA polymerase II, using a non-specific phage DNA template (38). From 15-25% of the enzyme molecules were active in this assay. The fraction of templates used in the specific runoff assay accorded with previous estimates for the efficiency of RNA polymerase II systems (1,2). These earlier measurements left open the possibility that some templates were transcribed in multiple rounds; the one-round assay does not suffer from this ambiguity.

DISCUSSION

We have shown that RNA polymerase II can utilize dinucleoside monophosphates to initiate faithful transcription at the early region IV and major late promoters of adenovirus. This polymerase therefore shares the ability of E. coli RNA polymerase and eukaryotic RNA polymerases I and III to incorporate dinucleotides by sequence complementarity at in vivo initiation sites.

The major late and EIV promoters seemed ideal for the study of transcription initiation, as these two promoters represent extremes: initiation in vivo occurs at a unique position at the major late promoter but is distributed over several adjacent nucleotides of EIV (41,46). Initiation in vitro at the major late promoter was examined in detail by testing the effects of all 16 dinucleotides on the positions of 5' termini of transcripts. Dinucleotides complementary to sequences from -4 (CpC) to +4 (CpU) primed RNA synthesis. In contrast, initiation at sites other than +1 was never observed by varying nucleoside triphosphate concentrations (data not shown). This suggests that initiation by dinucleotides and by nucleoside triphosphates may proceed by different mechanisms.

The homogeneity of initiation at the major late promoter could be due to the occurrence of an A residue at a mechanistically preferred position [see Baker and Ziff (41) for a similar discussion]. In vivo, RNA polymerase II initiates most frequently with purines, and with ATP preferentially to GTP. This could reflect preferential binding of ATP to the initiation site of the polymerase. The experiments presented here suggest that RNA polymerase II also has a strong preference for site +1 of the major late promoter, independent of the initiating nucleotide. The efficiency of dinucleotide priming was greatest at positions near +1, as shown by the ability of different dinucleotides to compete with ATP for initiation. Moreover, the novel phenomenon of misincorporation, mediated by sequence complementarity between the second residue of a dinucleotide and the template, was observed only at postions +1 and +2. Dinucleotides might be envisioned to prime the elongation reaction, thereby bypassing initiation events. This would obviate ATP preference, and identify positions on the template available to the elongation site of the polymerase.

A region accessible for dinucleotide priming

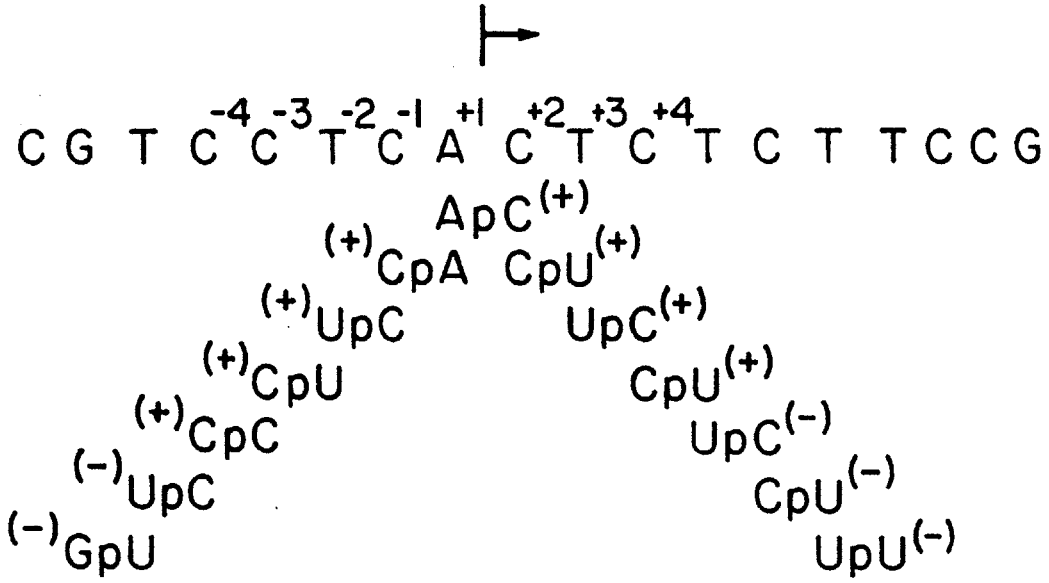
The major late promoter results show that a contiguous region of at least 8-9 bases is accessible for dinucleotide priming, centered around the in vivo cap site (see Fig. 7). This corresponds to a length of about 30 Å of DNA, or almost a full turn of the helix. It is difficult to picture how a static RNA polymerase aligned at a single base could initiate transcripts over this distance. Such a static complex would have to be capable of elongating RNA primers from eight different sites on the enzyme. An alternative and more likely possibility is that the polymerase drifts or oscillates over 8-9 bases around a preferred site of initiation. In this case, it would be possible to align a single site, for instance the elongation site of the enzyme, with any of eight different positions on the template. This plasticity of initiation is apparently distinct to RNA polymerase II as compared to E. coli polymerase and eukaryotic RNA polymerases I and III (26,29,30,47)[1]. A requirement for hydrolysis of the - bond of ATP has been suggested for initiation of transcription by RNA polymerase II (48). This cofactor dependence could be related to oscillation of the enzyme. In any case, the polymerase complex must unwind the DNA to allow base pairing with a complementary primer. E. coli RNA polymerase can initiate from an "open" complex which has unwound approximately one turn of the DNA helix (20-22). Perhaps dinucleotides prime a similar structure formed by RNA polymerase II and factors.

Rate of initiation by the activated complex

In the accompanying paper, we describe the formation of an "activated" complex by preincubation of template DNA, RNA polymerase II and factors. This complex was defined by its ability to rapidly begin transcription when presented with nucleoside triphosphates. The activated complex might therefore initiate directly, or might undergo one or more transitions to form the structure immediately preceding initiation. The time required for activated complexes to initiate RNA chains with ATP at the major late promoter was measured using a dinucleotide challenge experiment. Ninety percent of the observed chains initiated within the first minute of the pulse. It is worth noting that accumulation of activated complexes occurs over an hour

Figure 7. Dinucleotides used to prime RNA synthesis

The sequence of the Ad2 major late promoter cap site is given; *in vivo* initiation occurs at the +1 A residue. RNA polymerase II was shown to initiate with dinucleotides at all positions between −4 and +4, although most efficiently at −1 and +1. Examination of 197-nucleotide M13-truncated runoff RNA's and of complete RNase T1 digestion products, did not detect initiation with dinucleotides upstream of −4 or downstream of +4.



time interval, but that these complexes initiate quite rapidly. The rate limiting steps in transcription of the major late promoter by the reconstituted system can thus occur in the absence of nucleotides.

Initiation with ATP was sufficiently fast that stimulation of label incorporation by dinucleotides was not observed in a four minute pulse. However, stimulation of incorporation by dinucleotides was observed at the EIV promoter. Such stimulation of incorporation was also observed when high conentrations of the initiating nucleoside triphosphate were used. This suggests that initiation at EIV is slower than at the major late promoter.

The role of capping

Late promoter RNA's primed with dinucleotides had no obvious modifications of the 5' hydroxyl terminus, either in the reconstituted system (Fig. 3C) or in the whole cell extract (data not shown). Thus the presence of a cap is not required for, and does not enhance transcription by RNA polymerase II. Moreover, RNA initiated with ATP in the reconstituted system has unprotected phosphates at the 5' terminus. We have already noted that the bulk of guanyltransferase activity (assayed as described in reference 49) is recovered in a fraction not included in the reconstituted system (12). It thus seems highly unlikely that the ATP-initiated RNA's in the reconstituted system were ever capped, although transcripts were efficiently capped in the whole cell extract. Recent experiments with a nuclear wash extract have shown that initiation and some elongation precede capping (50). Together these results indicate that capping plays no direct role in the transcription reaction.

ACKNOWLEDGEMENTS

We are grateful for U. RajBhandary, R. Reilly, H. Drabkin, R. Padgett, S. Hardy, and S. Desiderio for advice on RNA analysis; to H. Drabkin and R. Padgett for marker RNAs; to U. Hansen, R. Kingston, P. Jat and K. Kierkegaard for critical comments; to L. Corboy and M. Esteve for expert technical assistance; to U. Hansen, B. Sollner-Webb, K. Miller, and J. K. Wilkinson for communicating results prior to publication; to the MIT Cell Culture Center for preparation of HeLa cells; to M. Siafaca for preparation of the manuscript; and to members of the Sharp lab for illuminating discussions.

Footnotes

[1]J. K. Wilkinson, K. G. Miller, and B. Sollner-Webb, personal communication

[2]The abbreviations used are: EIV, adenovirus early region four; Hepes, 4-(2-hydroxyethyl)-1-piperazine sulfonic acid; MLP, adenovirus major late promoter

[3]R. Padgett, personal communication

[4]U. Hansen, M. Fromm, and P. A. Sharp, manuscript in preparation

REFERENCES

1. Weil, P. A., Luse, D. S., Segall, J., and Roeder, R. G. (1979). Cell 18, 469-484.
2. Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980). Proc. Natl. Acad. Sci. USA 77, 3855-3859.
3. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980). Science 209, 1406-1415.
4. Grosveld, G. C., Shewmaker, C. K., Jat, P., and Flavell, R. A. (1981). Cell 25, 215-226.
5. Hu, S. L., and Manley, J. L. (1981). Proc. Natl. Acad. Sci. USA 78, 820-824.
6. Wasylyk, B., and Chambon, P. (1981). Nucleic Acids Res. 9, 1813-1824.
7. Grosschedl, R., and Birnstiel, M. L. (1982). Proc. Natl. Acad. Sci. USA 79, 297-301.
8. Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M. P., and Chambon, P. (1982). Proc. Natl. Acad. Sci. USA 79, 7132-7136.
9. Matsui, J., Segall, J., Weil, P. A., and Roeder, R. G. (1980). J. Biol. Chem. 255, 11992-11996.
10. Tsai, S. Y., Tsai, M.-J., Kops, L. E., Minghetti, P. P., and O'Malley, B. W. (1981). J. Biol. Chem. 256, 13055-13059.
11. Dynan, W., and Tjian, R. (1981). In Developmental Biology Using Purified Genes (Brown, D., ed.), pp. 401-414, Academic Press, New York.
12. Samuels, M., Fire, A., and Sharp, P. A. (1982). J. Biol. Chem. 257, 14419-14427.
13. Fire, A., Samuels, M., and Sharp, P. A., manuscript submitted.
14. Wormington, W. M., Bogenhagen, D. F., Jordan, E., and Brown, D. D. (1981). Cell 24, 809-817.
15. Schaak, J., Sharp, S., Dingermann, J., and Söll, D. (1983). J. Biol. Chem. 258, 2447-2453.
16. Davison, B. L., Egly, J.-M., Mulvihill, E. R., and Chambon, P. (1983). Nature (London) 301, 680-686.
17. Chamberlin, M. (1974). Ann. Rev. Biochem. 43, 721-775.
18. Hawley, D. K., and McClure, W. R. (1980). Proc. Natl. Acad. Sci. USA 77, 6381-6385.
19. Rosenberg, S., Kadesch, T. R., and Chamberlin, M. J. (1982). J. Mol. Biol. 155, 31-51.
20. Wang, J. C., Jacobsen, J. H., and Saucier, J.-M. (1977). Nucleic Acids Res. 4, 1225-1241.
21. Siebenlist, U., Simpson, R. B., and Gilbert, W. (1980). Cell 20, 269-281.
22. Kierkegaard, K., Buc, H., Spassky, A., and Wang, J. C. (1983). Proc. Natl. Acad. Sci. USA 80, 2544-2548.
23. Niyogi, S. K., and Stevens, A. (1965). J. Biol. Chem. 240, 2593-2598.
24. Downey, K. M., and So, A. G. (1970). Biochemistry 9, 2520-2525.
25. Hoffman, D., and Niyogi, S. K. (1973). Proc. Natl. Acad. Sci. USA 70, 574-578.
26. Minkley, E. G., and Pribnow, D. (1973). J. Mol. Biol. 77, 255-277.
27. Kupper, H., Contreras, R., Khorana, H. G., and Landy, A. (1976). In RNA Polymerase (R. Losick, and M. Chamberlin, eds.), pp. 473-484, Cold Spring Harbor Press, New York.

28. Glaser, G., and Cashel, M. (1979). Cell 16, 111-121.
29. Wilkinson, J. K., and Sollner-Webb, B. (1982). J. Biol. Chem. 257, 14375-14383.
30. Learned, R. M., and Tjian, R. (1982). J. Mol. Appl. Gen. 1, 575-584.
31. Cowie, A., Jat, P., and Kamen, R. (1982). J. Mol. Biol. 159, 225-255.
32. Shatkin, A. J. (1976). Cell 9, 645-653.
33. Furuichi, Y. (1978). Proc. Natl. Acad. Sci. USA 75, 1086-1090.
34. Babich, A., Nevins, J. R., and Darnell, J. E. (1980). Nature (London) 287, 246-248.
35. Jove, R., and Manley, J. L. (1982). Proc. Natl. Acad. Sci. USA 79, 5842-5846.
36. Randerath, K., and Randerath, E. (1967). Methods in Enzymol. 12, 323-347.
37. Hodo, H. G., and Blatti, S. P. (1977). Biochemistry 16, 2334-2343.
38. Kadesch, T. R., and Chamberlin, M. J. (1982). J. Biol. Chem. 257, 5286-5295.
39. Slattery, E., Dignam, J. D., Matsui, T., and Roeder, R. G. (1983). J. Biol. Chem. 258, 5955-5959.
40. Manley, J. L., Fire, A., Samuels, M., and Sharp, P. A. (1982). Methods in Enzymol., in press.
41. Baker, C. C., and Ziff, E. B. (1981). J. Mol. Biol. 149, 189-221.
42. Silberklang, M., Gillum, A. M., and RajBhandary, U. L. (1979). Methods in Enzymology 59, 58-109.
43. Padgett, R., Hardy, S., and Sharp, P. A., in press.
44. Fire, A., Baker, C. C., Manley, J. L., Ziff, E. B., and Sharp, P. A. (1981). J. Virol. 40, 703-719.
45. Gelinas, R. E., and Roberts, R. J. (1977). Cell 11, 533-544.
46. Ziff, E., and Evans, R. (1978). Cell 15, 1463-1475.
47. Majors, J. (1975). Proc. Natl. Acad. Sci. USA 72, 4394-4398.
48. Bunick, D., Zandomeni, R., Ackerman, S., and Weinmann, R. (1982). Cell 29, 877-886.
49. Venkatesan, S., and Moss, B. (1982). Proc. Natl. Acad. Sci. USA 79, 340-344.
50. Coppola, J. A., Field, A. S., and Luse, D. S. (1983). Proc. Natl. Acad. Sci. USA 80, 1251-1255.
51. Steenbergh, P. H., and Sussenbach, J. L. (1979). Gene 6, 307-318.

Appendix

DNA DEPENDENT TRANSCRIPTION OF ADENOVIRUS GENES IN A SOLUBLE WHOLE CELL EXTRACT

J.L. Manley, A. Fire, A. Cano, P.A. Sharp, and M.L. Gefter.

Proc. Natl. Acad. Sci. USA
Vol. 77, No. 7, pp. 3855–3859, July 1980
Biochemistry

DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract

(RNA polymerase II/transcription initiation/recombinant DNA/mRNA promoter/gene regulation)

JAMES L. MANLEY, ANDREW FIRE, AMPARO CANO, PHILLIP A. SHARP, AND MALCOLM L. GEFTER

Department of Biology and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Communicated by J. D. Watson, April 4, 1980

ABSTRACT **We have developed a cell-free system for studying the synthesis of mRNA in mammalian cells. The system consists of a dialyzed and concentrated whole-cell extract derived from HeLa cells, small molecules and cofactors needed for transcription, and exogenously added DNA. Accurate transcription by RNA polymerase II is entirely dependent upon addition of promoter-containing eukaryotic DNA. At optimal DNA and extract concentrations, transcription initiation from the adenovirus serotype 2 late promoter is readily detectable, and specific transcripts over 4000 nucleotides in length are observed. The RNA synthesized *in vitro* contains the same 5′ capped RNase T1 undecanucleotide as does the *in vivo* transcript. RNA synthesis also initiates accurately at both an early and an intermediate adenovirus promoter site.**

The availability of *in vitro* systems for the study of transcription has been of prime importance in elucidating mechanisms of gene regulation in prokaryotes (1, 2). The lack of similar eukaryotic systems has left many important problems unresolved. For example, virtually nothing is known about the signals and factors required for accurate initiation by RNA polymerase II (3). Control of initiation of transcription plays a major role in gene regulation in mammalian cells. At the moment, it is not clear that prokaryotic models of operators, repressors, and activators can be extended to transcription control in animal cells. We have been working to develop eukaryotic *in vitro* systems, and have recently shown that an *in vitro* system consisting of isolated nuclei is capable of carrying out many of the reactions involved in mRNA synthesis, including accurate initiation of transcription by RNA polymerase II (4, 5). Recently, Weil *et al.* (6) demonstrated that accurate initiation of transcription occurs on exogenously added DNA at a known viral transcription start site when purified RNA polymerase II is mixed with a cytoplasmic extract. Here we describe a simple system consisting primarily of a soluble whole-cell extract that accurately initiates transcription on exogenously added DNA templates. We feel that this system will prove useful for studying the biochemical mechanisms involved in the control of transcription and gene expression in mammalian cells.

MATERIALS AND METHODS

Preparation of DNA. Adenovirus serotype 2 (Ad2) virus and DNA were prepared as described (7). Restriction endonuclease DNA fragments *Bal* I-D [21.5–28.3 map units (m.u.)] and *Bal* I-E (14.7–21.5 m.u.) were inserted into plasmid pBR322 with *Bam*HI linkers and cloned (8). *Sma* I-F fragment (2.9–11.3 m.u.) was similarly inserted into pBR322 but with *Eco*RI linkers (gift of Kathleen Berkner and Frank Laski). Plasmid DNA was prepared by standard protocols (9).

Preparation of HeLa Cell Extracts. Cell lysates were prepared according to the method of Sugden and Keller (10). HeLa cells were grown in suspension culture in Eagle's minimal essential medium supplemented with 5% horse serum to a density of approximately 8×10^5 cells per ml. All further operations were done between 0 and 4°C. Cells were washed in phosphate-buffered saline containing $MgCl_2$ and the cell pellet was resuspended in four packed-cell volumes of 0.01 M Tris·HCl, pH 7.9/0.001 M EDTA/0.005 M dithiothreitol. After 20 min, the cells were lysed by homogenization in a Dounce homogenizer, using eight strokes with a "B" pestle. Four packed-cell volumes of 0.05 M Tris·HCl, pH 7.9/0.01 M $MgCl_2$/0.002 M dithiothreitol/25% sucrose/50% (vol/vol) glycerol were then added and the mixture was stirred gently. To this suspension one packed-cell volume of saturated $(NH_4)_2SO_4$ was added dropwise; the lysate was then gently stirred for an additional 20 min. The extract was then centrifuged at 50,000 rpm for three hours in a Beckman 60 Ti rotor. The supernatant was decanted so as not to disturb the pellet and precipitated by addition of solid $(NH_4)_2SO_4$ (0.33 g/ml of suspension). After the $(NH_4)_2SO_4$ was dissolved, 0.01 ml of 1 M NaOH per 10 g of $(NH_4)_2SO_4$ was added, and the suspension was stirred for an additional 30 min. The precipitate was collected by centrifugation at $15,000 \times g$ for 20 min and resuspended with $^1/_{10}$ the volume of the high-speed supernatant into a buffer containing 50 mM Tris·HCl, pH 7.9, 6 mM $MgCl_2$, 40 mM $(NH_4)_2SO_4$, 0.2 mM EDTA, 1 mM dithiothreitol, and 15% glycerol. This suspension was dialyzed for 4–8 hrs against each of two changes of 100 vol of the same buffer. The dialysate was centrifuged at $10,000 \times g$ for 10 min and the supernatant was quick frozen in small samples in liquid N_2. Samples stored at −80°C retained full activity for at least 2 months. In some later experiments (e.g., the one shown in Fig. 5) the dialysis buffer was 20 mM Hepes, pH 7.9/100 mM KCl/12.5 mM $MgCl_2$/0.1 mM EDTA/2 mM dithiothreitol/17% glycerol.

***In Vitro* Incubations and Purification of RNA.** Standard 50-μl reaction mixtures contained 15 mM Tris·HCl, pH 7.9, 7 mM $MgCl_2$, 32 mM $(NH_4)_2SO_4$ 0.2 mM EDTA, 1.3 mM dithiothreitol, 10% glycerol, 500 μM ATP, CTP, and GTP, 50 μM UTP containing 10 μCi of [α-^{32}P]UTP, 15 μl of extract, and 2.5 μg of DNA (1 Ci = 3.7×10^{10} becquerels). More recently, such as in the experiment shown in Fig. 5, we have altered the reaction mixtures so that the Tris was replaced with 12 mM Hepes, pH 7.9, the $(NH_4)_2SO_4$ was replaced with 60 mM KCl, and the amount of extract was increased to 30 μl. Reaction mixtures were incubated at 30°C for 60 min. RNA was extracted from the reaction mixture as described (4) and precipitated with ethanol. The pellet was redissolved in 0.2% NaDodSO$_4$. After addition of NH_4OAc to 1 M (a total volume of 400 ml), RNA was again precipitated with 2 vol of ethanol, redissolved in 0.2% NaDodSO$_4$/0.3 M NaOAc, pH 5.2, and precipitated with ethanol. This pellet, free of unincorporated

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: Ad2, adenovirus 2; m.u., map unit(s); DBAE-cellulose, dihydroxylborylcellulose.

nucleotide triphosphates, was dissolved in 0.2% Sarkosyl/2 mM EDTA and stored at −20°C. RNA was analyzed by glyoxalation and agarose gel electrophoresis as described by McMaster and Carmichael (11). For cap analysis, RNA was digested with RNase T1, bound to columns of dihydroxylborylcellulose (DBAE-cellulose), eluted, and fingerprinted as described (4, 12, 13).

RESULTS

Extract and DNA Optima for Accurate Initiation. The initiation site for late Ad2 transcription has been positioned at 16.5 m.u. (Fig. 1) (14–16). We have shown that transcription initiates here in nuclei isolated from Ad2-infected HeLa cells (4), and Weil *et al.* (6) recently demonstrated that purified RNA polymerase will initiate at this site on exogenously added viral DNA in the presence of a cytoplasmic extract obtained from uninfected HeLa cells. To test the transcription potential of whole cell extracts (prepared as described in *Materials and Methods*) *Bam*HI-digested pBR322-*Bal* I-E DNA mixture was used as exogenously added template. Initiation at 16.5 m.u. and elongation to the end of *Bal* I-E would produce a 1750-nucleotide RNA (Fig. 1).

The effect of varying the concentration of the whole cell extract on transcription from the late promoter of Ad2 is shown in Fig. 2*A*. Increasing concentrations of extract (protein 1–7.5 mg/ml) were added to reaction mixtures containing either pBR322-*Bal* I-E or pBR322-*Bal* I-D DNA (unseparated, restriction endonuclease-digested DNAs). A quarter of the transcription products were treated with glyoxal and resolved by electrophoresis in an agarose gel. When low concentrations of extract were added to a reaction mixture containing pBR322-*Bal* I-E, a heterogeneous mixture of RNA product was observed (Fig. 2*A*, lane 1). In striking contrast, if the amount of extract in the reaction mixture was doubled, synthesis of RNA was essentially abolished (Fig. 2*A*, lane 2). Addition of still higher concentrations of extract to reaction mixtures containing the pBR322-*Bal* I-E DNA specifically stimulated synthesis of a 1750-nucleotide RNA derived from the late Ad2 promoter (Fig. 2*A*, lanes 3 and 4). At the highest concentration of extract tested, this RNA species represents approximately 40% of the applied sample. (Three examples of analysis of *in vitro* transcription of the *Bal* I-E fragment are shown in Figs. 2*A*, lane 4, 2*B*, lane 4, and 3*A*, lane 3. The specificity of transcription seen in these samples is typical of our total results.) A comparison of lanes 1 and 4 in Fig. 2*A* suggests that, in addition to enhancing specific transcription from the late promoter of Ad2, higher extract concentration also suppresses nonspecific transcription. At either high or low extract concentration, the pBR322-*Bal* I-D fragments are less effective in stimulating RNA synthesis. In particular, a high concentration of extract showed no detectable bands when pBR322-*Bal* I-D DNA was

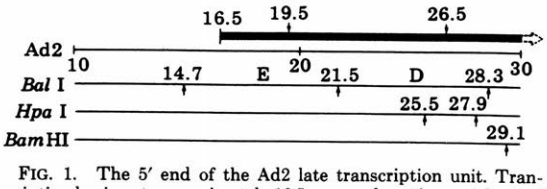


FIG. 1. The 5′ end of the Ad2 late transcription unit. Transcription begins at approximately 16.5 m.u. and continues rightward to almost the end of the Ad2 genome (100 m.u.). One m.u. is roughly 350 base pairs. The coordinates of the three leader segments on late mRNAs are indicated by the numbers above the solid arrow. The sites at which three restriction endonucleases used in these experiments cleave this region of the Ad2 genome are indicated.

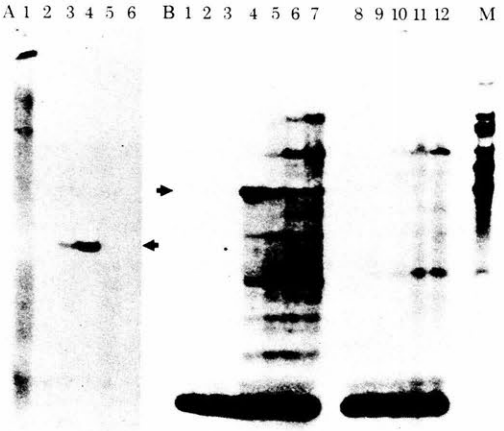


FIG. 2. Extract and DNA concentration optima for accurate transcription. RNA was synthesized in standard reaction mixtures except that either the extract (*A*) or DNA (*B*) concentration was varied, as follows. (*A*) Lanes 1–4, *Bal* I-E-pBR322 DNA template: 2.5, 5, 10, or 15 μl of extract; lanes 5 and 6, *Bal* I-D-pBR322 DNA template: 2.5 or 15 μl of extract. (*B*) Lanes 1–7, 0, 12.5, 25, 50, 75, 100, or 125 μg of *Bal* I-E-pBR322 DNA template per ml; lanes 8–12, 25, 50, 75, 100, or 125 μg of *Bal* I-D-pBR322 DNA template. After extraction and glyoxalation, 25% of each sample was resolved in a 1.4% agarose gel. The number of cpm (Cerenkov radiation) loaded in each slot was as follows: (*A*) 1, 9700; 2, 3400; 3, 2900; 4, 2500; 5, 2600; 6, 1100; (*B*) 1, 2700; 2, 3100; 3, 3400; 4, 4900; 5, 5800; 6, 5900; 7, 6700; 8, 3700; 9, 4000; 10, 4400; 11, 4700; 12, 4900. (The cpm in *B* were approximately 2 times higher than usually obtained. This may have resulted from incomplete removal of nucleotide triphosphates during ethanol precipitation.) The arrows indicate the position of the 1750-nucleotide transcript that originates from the Ad2 late promoter. M is an *Eco*RI digest of Ad2 DNA labeled *in vitro* by nick translation.

used as template (Fig. 2*A*, lane 6). This is not due to inhibition of RNA synthesis by *Bal* I-D DNA because a mixture of pBR322-*Bal* I-E and pBR322-*Bal* I-D DNAs still resulted in the synthesis of the 1750-nucleotide species (data not shown).

The results of varying the concentration of DNA present in standard reaction mixtures are shown in Fig. 2*B*. Concentrations of template DNA less than 50 μg/ml produced virtually no transcription (Fig. 2*B*, lanes 1–3). (We believe the small labeled RNA resulted from labeling of endogenous tRNA present in the lysate.) Reaction mixtures containing DNA at 50 μg/ml (Fig. 2*B*, lane 4) yielded the same high levels of specific transcription shown in Fig. 2*A*, lane 4. At higher DNA concentrations, 75 μg/ml (Fig. 2*B*, lane 5) and 100 μg/ml (Fig. 2*B*, lane 6), the amount of the 1750-base-long transcript synthesized decreased and other discrete transcripts appeared. S1-nuclease mapping suggests they arise from "end-to-end" transcription of the *Bal* I-E DNA fragment and the pBR322 DNA molecule (data not shown). We have not determined which class of polymerase is responsible for this end-to-end transcription. The origin of the lower molecular weight species seen in Fig. 2*B*, lanes 5 and 6, is not clear. Fig. 2, lanes 8–12, shows the effects of increasing concentrations of pBR322-*Bal* I-D DNA as template. The 1750-nucleotide transcript was not detected. At the higher DNA concentrations, we again observed the transcription of lower and higher molecular weight RNAs. The amount of this synthesis was lower with pBR322-*Bal* I-D DNA than with the pBR322-*Bal* I-E DNA.

Additional Properties of *in Vitro* Transcripts Originating from the Ad2 Major Late Promoter. Fig. 3*A* shows that synthesis of the 1750-nucleotide transcript was sensitive to α-

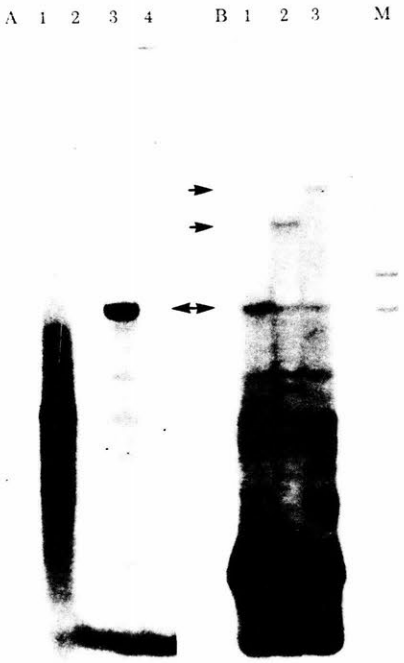


FIG. 3. (*A*) α-Amanitin sensitivity of transcription from the Ad2 late promoter. RNA was synthesized in 100-μl reaction mixtures, each containing 400 μCi of [α-^{32}P]UTP. Lanes 1–3 contained *Bal* I-E-pBR322 DNA as template and lane 4, *Bal* I-D-pBR322 DNA. Lane 1 contained 5 μl of extract; lane 2, 30 μl of extract + α-amanitin at 0.5 μg/ml; lanes 3 and 4 each contained 30 μl of extract. One percent of the RNA extracted from each reaction mixture was glyoxalated and analyzed by gel electrophoresis. The number of cpm loaded in each slot was: 1, 7100; 2, 8100; 3, 3300; 4, 1900. (*B*) *In vitro* transcription of total Ad2 DNA. RNA was extracted from standard reaction mixtures that had contained 50 μCi of [α-^{32}P]UTP. The Ad2 DNA template had been previously digested with *Bal* I (lane 1), *Hpa* I (lane 2), or *Bam*HI (lane 3). Twenty-five percent of each sample was glyoxalated and resolved in a 1.4% agarose gel; 34,000 cpm was loaded in slot 1, 28,600 in slot 2, and 14,500 in slot 3. The arrows indicate the positions of runoff transcripts from the Ad2 late promoter.

amanitin at 0.5 μg/ml (compare lanes 2 and 3 of Fig. 3*A*). This implicates RNA polymerase II as the enzyme responsible for the specific synthesis. Fig. 3*B*, lanes 1–3, shows that virion DNA is also active in these extracts. Here the added viral DNA had been cleaved with *Bal* I, *Hpa* I, or *Bam*HI restriction endonuclease, and these fragments stimulated synthesis of 1750-, 3150-, and 4400-nucleotide RNA chains, respectively (arrows). These are the expected RNA products for initiation occurring at the late promoter site at position 16.5 m.u. and elongation continuing to the end of each fragment (see Fig. 1). Note that the results in Fig. 3*B* show that the whole cell extract system will elongate RNA chains over 4.4 kilobases.

The origins of most of the lower molecular weight species in Fig. 3*B* have not been established. Because several adenovirus promoters other than the major late promoter are active in this system (see below), at least some of these transcripts likely result from initiation at these other sites. Some may also result from processing of one of these primary transcripts. Note that lanes 2 and 3, as well as 1, contain an RNA species of approximately 1750 nucleotides. The presence of these bands in lanes 2 and 3 may have resulted from an occasionally observed actinomycin D-insensitive DNA-independent reaction, which we believe is due to end-labeling of endogenous 18S ribosomal RNA (unpublished results). The heavily overexposed band at the bottom of each track represents one or both of the two small virus-associated (VA) RNAs, which are RNA polymerase III products (17).

Fingerprint Analysis of 5′ "Cap" Structure. The structure of the 5′ end of the Ad2 late mRNAs has been well studied (12, 13). We therefore decided to examine the structure of the 5′ end of the Ad2 RNA synthesized *in vitro*. For this, RNAs extracted from the reaction mixtures analyzed in Fig. 3*A* were digested with RNase T1. About 5% of samples made with *Bal* I-E pBR322 and *Bal* I-D-pBR322 DNA as template were directly analyzed by two-dimensional fingerprint analysis (Fig. 4 *A* and *B*, respectively). The remainders were resolved by selection of "capped" oligonucleotides on columns of DBAE-cellulose and then fingerprinted. When we analyzed the RNA from the *Bal* I-E reaction after selection on DBAE-cellulose, only two large oligonucleotides were detected (Fig. 4*C*, arrows). To identify these oligonucleotides, we eluted them, as well as several of the

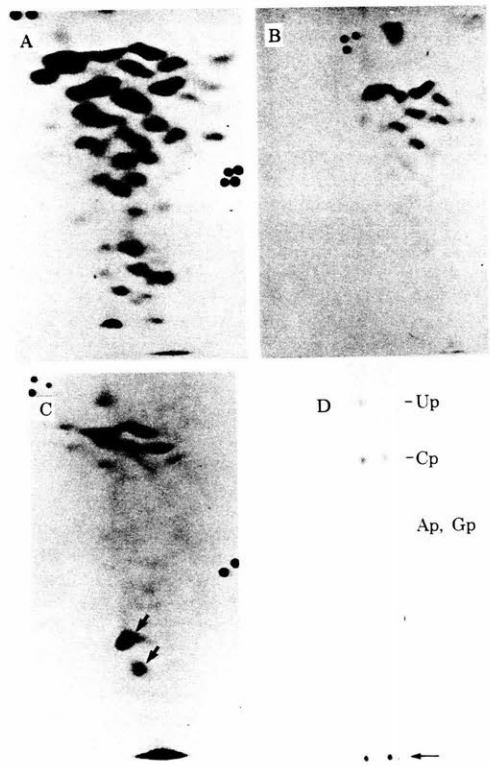


FIG. 4. Fingerprint analysis of RNA synthesized *in vitro*. The remainder of the RNA from each of the samples that had been analyzed by gel electrophoresis in Fig. 3 was digested with RNase T1. Five percent of each sample was analyzed directly by two-dimensional fingerprinting. (*A* and *B*) Fingerprints of RNA synthesized with *Bal* I-E-pBR322 DNA and *Bal* I-D-pBR322 DNA respectively, as templates. Electrophoresis was from left to right, and homochromatography from bottom to top. (*C*) Fingerprint of the oligonucleotides selected by DBAE-cellulose chromatography from the reaction mixture that had contained *Bal* I-E-pBR322 DNA. The arrows indicate the position of the Ad2 capped T1 oligonucleotide (see text). (*D*) Products obtained when aliquots of the major capped oligonucleotide in *C* were digested with RNase A (left) or RNase T2 (right) and analyzed by electrophoresis at pH 3.5 on DEAE-paper. Electrophoresis was from bottom to top. The arrow shows the position of the T2-resistant cap structure.

smaller oligonucleotides at the top of the fingerprint, and digested them with either RNase A or RNase T2. The only oligonucleotides that showed the partial resistance to RNase T2 expected of a 5′ cap structure (Fig. 4*D*) were the two indicated by the arrows in Fig. 4*C*. Furthermore, these two oligonucleotides gave identical nearest-neighbor analyses when digested with either of the two nucleases. The existence of two forms of the late capped oligonucleotide m^7GpppAmpCpUpCpUpCpUpUpCpCpGp(C) (14) has been routinely observed (e.g., ref. 12). Further results (not shown) demonstrated that the cap oligonucleotide was completely resistant to RNase U2 digestion, thereby showing the absence of non-2′-*O*-methylated A residues. Similar analysis of RNA made in the presence of a low concentration of α-amanitin showed that not only does the synthesis of the entire 1750-nucleotide transcript require RNA polymerase II, but also synthesis of the very first 11 nucleotides is dependent on this enzyme. The fingerprint of the RNA synthesized in reaction mixtures containing a low concentration of extract (see Fig. 3*A*, lane 1) was extremely complex. However, when we analyzed the material which bound to DBAE-cellulose we could detect no cap-containing oligonucleotides. Likewise, no capped oligonucleotides were detected in the RNA extracted from reaction mixtures that had contained *Bal* I-D-pBR322 DNA as template (results not shown).

Synthesis from Other Adenovirus Promoters. In order to determine whether any other adenovirus promoters could be specifically recognized *in vitro*, we examined the transcripts obtained from reaction mixtures in which the cloned *Sma* I-F (2.9–11.3 m.u.) fragment of Ad2 had been used as DNA template. Current evidence suggested that two promoters might be located within this fragment. The 5′ ends of early region 1B mRNAs are located at 4.9 m.u. (18), and the 5′ end of the mRNA, which encodes the virion polypeptide IX, a mRNA that is expressed from intermediate to late times after infection, has been mapped at approximately 10.3 m.u. (19). The 3′ ends of all these mRNAs are located at 11.5 m.u., and are thus not contained on the *Sma* I–fragment. We expected transcripts of 2200, 1100, or 340 nucleotides if transcription initiates at the early region 1B promoter and continues to the end of DNA templates generated by cleavage with *Eco*RI, *Hin*dIII, or *Kpn* I, respectively (see Fig. 5). Transcripts of approximately 350 and 380 nucleotides should be produced by transcription initiation at the IX promoter and elongation to the end of the *Eco*RI- or *Hin*dIII-cleaved DNAs, respectively. Fig. 5, lanes 2–4, shows that transcripts migrating at the expected mobilities are synthesized. We therefore conclude that, in addition to the major late promoter, the early region 1B and polypeptide IX promoters are accurately recognized *in vitro*.

One use of the *in vitro* transcription system will be to compare the "strengths" of various eukaryotic promoters. An approach to this question is shown in Fig. 5, lanes 5 and 6. An aliquot of *Eco*RI-cut *Sma* I-F-pBR322 DNA was mixed with a given amount of either *Bal* I-E-pBR322 or *Bal* I-D-pBR322 DNA and transcribed *in vitro*. The results show that both the IB and late promoters functioned when added in the same reaction mixture. The late promoter appears to be the stronger of the two.

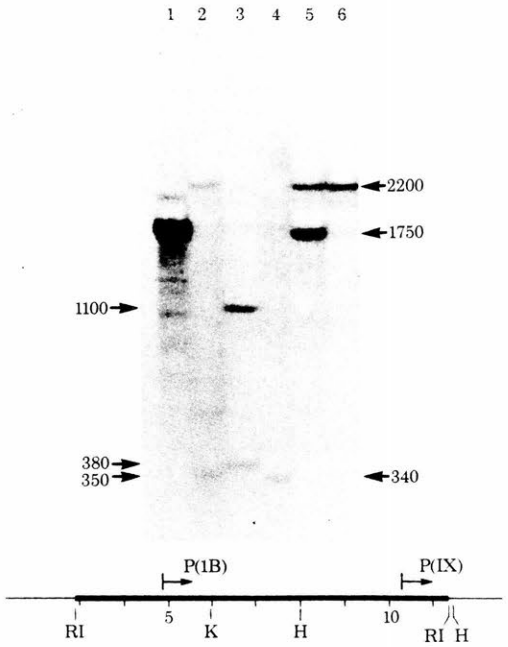


FIG. 5. Transcription from other adenovirus promoters. RNA was synthesized in standard reaction mixtures. DNA templates were: 1, *Bal* I-E-pBR322; 2, adenovirus 5 *Sma* I-F-pBR322; 3, *Sma* I-F-pBR322 digested with *Hin*dIII; 4, *Sma* I-F-pBR322 digested with *Kpn* I; 5, 1 μg of *Bal* I-E-pBR322 + 1 μg of *Sma* I-F-pBR322 + 0.5 μg of *Bal* I-D-pBR322; 6, 1 μg of *Sma* I-F-pBR322 + 1.5 μg of *Bal* I-D-pBR322 DNA. The drawing shows the adenovirus 5 map coordinates and the sites at which the restriction enzymes *Eco*RI (RI), *Kpn* I (K), and *Hin*dIII (H) cleave the adenovirus 5 DNA (thick line) or the pBR322 DNA (thin line). (*Eco*RI sites in the recombinant plasmid correspond to *Sma* I sites in adenovirus 5.) The inferred locations of the two adenovirus promoters are indicated by the arrows above the line. The arrows indicate the position in the gel to which RNA molecules of the indicated sizes (nucleotides) would migrate, as determined by comparison with the mobilities of glyoxalated DNA fragments obtained from a *Sma* I digest of ^{32}P-labeled Ad2 DNA and a *Hph* I digest of ^{32}P-labeled pBR322 DNA (not shown).

DISCUSSION

The development of a DNA-dependent soluble *in vitro* transcription system for mammalian RNA polymerase II has been a long-standing objective in the study of eukaryotic gene regulation. Therefore, it was unexpected when we found that concentrated whole cell extracts are proficient in initiating transcription at suspected promoter sites on exogenously added DNA. We have shown that a whole cell extract initiates transcription at the late promoter site of Ad2, and we have evidence suggesting that two other viral promoters are utilized by the system as well. In collaboration with T. Maniatis and H. Handa, respectively, we have obtained preliminary evidence that the human β-globin gene and the early and late simian virus 40 genes are accurately transcribed *in vitro* (unpublished results).

Two types of evidence suggest that the whole cell extract initiates transcription at the late promoter site of Ad2. First, addition to *in vitro* reaction mixtures of viral DNA fragments cleaved by restriction endonucleases at sites 530 (*Sma* I, data not shown), 1750, 3150, or 4400 base pairs from the late promoter site at 16.5 m.u. stimulates synthesis by RNA polymerase II of RNA chains of these lengths. Second, RNA transcribed *in vitro* from late promoter is modified to form one predominant capped T1-resistant oligonucleotide, which is identical to the *in vivo*-synthesized undecanucleotide analyzed previously (12, 4). This *in vitro*-synthesized cap is completely modified by methylation at the 2′ position of the ribose on the penultimate adenosine. RNA synthesized *in vitro* from exogenously added DNA in the system of Weil *et al.* (6) was also shown to be capped with methylation of the 2′ position of ribose. On the

basis of sizing of runoff products, the whole cell extract also appears to initiate transcription from at least two other adenovirus promoter sites, early region 1B (4.9 m.u.) and the region coding for IX polypeptide (10.3 m.u.). During the lytic cycle of Ad2, the early region 1B is expressed from 4–6 hr after infection (20), whereas synthesis of the mRNA for polypeptide IX is prominent slightly later at an intermediate to late stage of infection (20). It is interesting that three viral promoters that appear to be utilized at three different stages of infection *in vivo* are all recognized *in vitro* by an extract prepared from uninfected cells.

Specific initiation by RNA polymerase II at the late adenovirus promoter site has been previously reported by Weil *et al.* (6). In this case, saturating amounts of purified RNA polymerase II were added to reaction mixtures containing a S-100 cytoplasmic extract prepared from uninfected cells. Under optimal salt, divalent cation, RNA polymerase II, and DNA concentration we calculate that these authors obtained approximately 5.2×10^{-4} pmol of a specific transcript from the late adenovirus promoter site per 50-μl reaction mixture. The whole cell extract system, described here, yields 3.0×10^{-3} pmol from the same promoter site in an identical volume. Thus the concentrated whole cell extract is at least as active as the RNA polymerase II-supplemented system. It is interesting that 100–150 units of purified RNA polymerase II is required to saturate the cytoplasmic extract system, while assays of the amount of RNA polymerase II activity present in the whole cell extract, by the method of Schwartz *et al.* (21), suggest that only 6–10 units of active RNA polymerase II are present in our comparable reaction mixture (results not shown). Perhaps the whole cell extract contains factors lost in the cytoplasmic extract system that enhance the efficiency of initiation by RNA polymerase II. The major advantages of the whole cell extract transcription system are its ease of preparation, flexibility, and possible retention of factors disgarded in other protocols. The ability to follow specific transcription at such an early stage in the fractionation of the extract will permit reconstitution of the system with purified factors.

The availability of a DNA-dependent soluble transcription system will also permit the investigation of regulatory factors that control the rate of initiation at different promoter sites. That these studies will be interesting is suggested by the fact that the late promoter of adenovirus is efficiently recognized by a whole cell extract from uninfected HeLa cells. As mentioned previously, Weil *et al.* (6) have also shown that RNA polymerase II and cytoplasmic extracts from uninfected HeLa cells will recognize the late promoter of adenovirus. Introduction *in vivo* of this same segment of DNA as part of the virus genome into the nuclei of HeLa cells does not result in detectable levels of transcription from this promoter until the onset of viral DNA replication (22). However, early promoters from the same viral genome are efficiently transcribed before viral DNA replication. One of these early promoters is the 1B region promoter, which is transcribed, but less efficiently than the late promoter in the whole cell extract from uninfected HeLa cells (see Fig. 5, lane 5). It is possible that recognition of a promoter site in the *in vitro* reaction by RNA polymerase II might reflect a nonphysiological interaction between the protein complex and the initiation site. Hence, within the cell, additional factors would be required for promoting initiation. Alternatively, early in the lytic cycle the late promoter site of Ad2 could be sequestered by proteins or modified so that RNA polymerase II does not recognize it. Another possibility is that the late promoter actually does function at early times after infection, but RNA has not been detected because it is degraded, or prematurely terminated, *in vivo*. This model would suggest that control of adenovirus gene expression might occur, at least in part, at a step other than initiation.

The availability of *in vitro* systems for transcription of mammalian genes will provide a wealth of information on the biochemistry of RNA processing and modification. Highly labeled substrates can now be synthesized that are similar, if not identical, to intracellular nuclear intermediates and that can be used to assay for specific processing enzymes. It is intriguing that several smaller specific size RNA species are resolved after electrophoresis of RNA synthesized *in vitro* from the late Ad2 promoter (e.g., Fig. 2*B*). Some of these RNAs are of the correct size to be intermediates in the processing of late viral RNA by RNA splicing. The purification and characterization of all those hypothetical activities that cleave, splice, and modify RNA in the nucleus of a mammalian cell, as well as identification of the signals on nucleic acids that bring about these reactions, should be facilitated by this *in vitro* system.

We thank T. Weil, R. Roeder, and C. Parker for communicating their results to us prior to publication, and S. Huang for excellent technical assistance. We thank the Massachusetts Institute of Technology Cell Culture Center for providing HeLa cells for these studies. P.A.S. acknowledges support from the American Cancer Society (Grant MV-37D) and the National Science Foundation (Grant PCM78-23230); M.L.G. acknowledges support from the American Cancer Society (Grant NP-6H) and the National Institutes of Health (Grant AI13357-04). J.L.M. was supported by National Institutes of Health Training Grant CA09255, A.F. by a Graduate Fellowship from the National Science Foundation, and A.C. by a Fellowship from the Spanish Council for Scientific Research.

1. Zubay, G., Chambers, D. & Cheong, L. (1970) in *The Lactose Operon*, eds. Beckwith, J. & Zipser, D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 375–391.
2. Losick, R. & Chamberlin, M., eds. (1976) *RNA Polymerase* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
3. Roeder, R. G. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 285–330.
4. Manley, J. L., Sharp, P. A. & Gefter, M. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 160–164.
5. Manley, J. L., Sharp, P. A. & Gefter, M. L. (1979) *J. Mol. Biol.* **135**, 171–197.
6. Weil, P. A., Luse, D. S., Segall, J. & Roeder, R. G. (1979) *Cell* **18**, 469–484.
7. Petterson, U. & Sambrook, J. (1973) *J. Mol. Biol.* **73**, 125–130.
8. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Ruther, W. J. & Goodman, H. M. (1977) *Science* **196**, 1313–1318.
9. Davis, R., Roth, J. & Botstein, D. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 66–67.
10. Sugden, B. & Keller, W. (1973) *J. Biol. Chem.* **248**, 3777–3788.
11. McMaster, G. K. & Carmichael, G. C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4835–4838.
12. Gelinas, R. E. & Roberts, R. J. (1977) *Cell* **11**, 533–544.
13. Klessig, D. F. (1977) *Cell* **12**, 9–21.
14. Ziff, E. & Evans, R. (1978) *Cell* **15**, 1463–1475.
15. Evans, R. M., Fraser, N. W., Ziff, E., Weber, J., Wilson, M. & Darnell, J. E. (1977) *Cell* **12**, 733–739.
16. Goldberg, S., Weber, J. & Darnell, J. E. (1977) *Cell* **10**, 617–621.
17. Jaehning, J. A., Weinmann, R., Brendler, T. G., Raskas, H. J. & Roeder, R. G. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp 819–834.
18. Berk, A. J. & Sharp, P. A. (1978) *Cell* **14**, 695–711.
19. Alestrom, P., Akusjärvi, G., Perricaudet, M., Mathews, M. B., Klessig, D. F. & Pettersson, U. (1980) *Cell* **19**, 671–682.
20. Spector, D. J., McGrogan, M. & Raskas, H. J. (1978) *J. Mol. Biol.* **126**, 395–414.
21. Schwartz, L. B., Sklar, V. E. F., Jaehning, J. A., Weinmann, R. & Roeder, R. G. (1974) *J. Biol. Chem.* **249**, 5889–5897.
22. Sehgal, P., Frazer, N. & Darnell, J. R. (1979) *Virology* **94**, 185–191.

Biographical Note

Name: Andrew Zachary Fire
Birthdate: April 27, 1959
Birthplace: Palo Alto, California

Education

1972-1975	Fremont High School, Sunnyvale, California
1975-1978	University of California, Berkeley A.B. in Mathematics June 1978
1978-1983	Graduate School, Massachusetts Institute of Technology Ph.D. in Biology June 1983

Teaching

Fall 1976	Reader for Math 51C (Differential Equations), UC Berkeley
Fall 1979	Teaching Assistant for 7.03, (Genetics) MIT
Fall 1981	Teaching Assistant for 7.03, (Genetics) MIT

Research

6/77-9/77	High Energy Physics Group, Lawrence Berkeley Lab; with Dr. R. Ely.
4/78-6/78	Molecular Biology Lab, UC Berkeley; with Dr. R. Calendar.
7/78-9/78	Biology Department, MIT; with Dr. B. Magasanik.
5/79-5/83	Center For Cancer Research, MIT; thesis research in the laboratory of Dr. P. A. Sharp.

Publications

Manley, J. L., Fire, A., Cano, A., Sharp, P. A. and Gefter M. L. (1980). DNA dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77: 3855-3859

Sharp, P. A., Manley, J. L., Fire, A., and Gefter, M. L. (1980). Regulation of adenovirus mRNA synthesis. Ann. N.Y. Acad. Sci. 354: 1-15

Fire, A., Baker, C. C., Manley J. L., Ziff, E. B. and Sharp, P. A. (1981). In vitro transcription of adenovirus. J. Virol. 40: 703-719

Crawford, N., Fire, A., Samuels, M., Sharp, P. A., and Baltimore, D. (1981). Inhibition of transcription factor activity by poliovirus. Cell 27: 555-561

Fire, A., Baker, C. C., Ziff, E.B., and Sharp, P. A. (1981). Transcription of adenovirus DNA in infected cell extracts. In: Developmental Biology Using Purified Genes. ICN-UCLA Symposia 23, D. Brown and C.F. Fox eds. (New York: Academic Press), pp. 387-399

Samuels, M., Fire, A., and Sharp, P. A. (1982). Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. 257, 14419-14427

Manley, J. L., Fire, A., Samuels, M., and Sharp, P. A. (1983). In vitro transcription: whole cell extract. Methods in Enzymology, 101, 568-582

Laski, F., Fire, A., RajBhandary, U., and Sharp, P. A., Characterization of tRNA precursor splicing in mammalian extracts. J. Biol. Chem., in press

Fire, A., Samuels, M., and Sharp, P. A., Interactions between RNA polymerase II, factors, and template leading to accurate transcription, submitted for publication June 1983

Samuels, M., Fire, A., and Sharp, P. A., Dinucleotide priming of transcription mediated by RNA polymerase II, submitted for publication June 1983