Problem Outline

Can we prevent parasites from using the host transcription machinery? Analysis of transcription, replication, recombination and translation, in parasites, are still (1993) virtually virgin fields. Biochemical approaches may be combined with genetics (in vivo tools, for example, the two hybrid system). Structural studies of transcription factors are likely to be a gold mine for basic science and small molecule discovery for therapeutics.

Transcription Factors in Parasites

Transcription in parasites is poorly understood. The establishment of successful transfection procedures for Leishmania [Dyann Wirth, Harvard], Plasmodium [Dyann Wirth, Harvard; Louis Miller, NIH], Toxoplasma [John Boothroyd, Stanford] and Trypanosomes [George Cross, Rockefeller] are likely to aid investigators in analysis of gene expression. Scientists with expertise in transcription may discover fundamental principles of parasite biology and shed light on host-parasite interactions. Identifying regulatory molecules may provide clues to vaccines or develop therapeutic strategies.

Little is known about the subunit structure of RNA polymerases either from Plasmodium or Trypanosome. Characterization of RNA polymerases and associated basal transcription complex from these and/or other parasites are essential. Most eukaryotic RNA polymerase II is present in two phosphorylated forms [Young RA (1991) Ann Rev Biochem 60 689-715]. Phosphorylation is mainly located on the carboxy terminal domain [CTD] of the largest subunit [RPB1]. The CTD of yRPB1 consists of 26 or 27 nearly identical copies of a hepta-peptide sequence [Tyr-Ser-Pro-Thr-Ser-Pro-Ser]. In mammals there are about 52 repeats. CTD in P. falciparum is distinct with a 63 aa tail [Li et al NAR (1989) 17 9621-9636]. Trypanosomes lack this type of CTD [Smith et al (1989) Cell 56 815-827]. We do not know whether CTD-associated factors identified in other organisms [Koleske et al (1992) Cell 69 883-894 and Thompson et al (1993) Cell 73 1361-1375] may interact with similar regions of the Plasmodium CTD. Unique regions of Plasmodium CTD may be used as an affinity tag to identify molecules which interact with the region. In a
parallel genetic approach, yCTD may be modified to resemble *Plasmodium* CTD [or parts of it] and genetic selection in yeast [Nature (1993) 364 121-126] may identify factors, which have homologues in *Plasmodium*.

For regulated transcription by pol II, the binding of TBP to the TATA sequence is essential but structurally TBP is not universal. *P. falciparum* TBP [pfTBP] is bit different [Hyde *et al* (1992) Nature 360 541; McAndrew *et al* (1993) Gene 124 165-171]. The authors point out that such differences "could be exploited chemotherapeutically" in future. It is of interest to note the highly conserved Proline [β sheet 1] in TBP [except *Plasmodium* which is essential for specificity of transcription by pol II as well as pol III [Schultz *et al* (1992) Cell 69 697-702]. Substituting Proline abolishes transcription by pol II and pol III but not by pol I. In pfTBP, Leucine replaces this Proline at position 50. It is exciting to speculate the pol I bias (?) of pfTBP with reported transcription of VSG and PARP by RNA polymerase I [Chung *et al* (1993) MCB 13 3734-43]. Comparison of pfTBP with *Arabidopsis thaliana* TBP-2 [Nikolov *et al* (1992) Nature 360 40-46] reveal differences in the region identified as the yTFIIA interaction site by mutational analysis. Upon closer analysis of H2 and S1' the differences appear to be located mostly in H2. Given the subunit [heterodimeric in yeast and heterotrimeric in humans] structure of TFIIA, perhaps one subunit is more conserved across species while other(s) may be species specific. Two-hybrid systems may be useful to search for parasite homologues. Overall, the TBP core that interacts with DNA is better conserved than the faces, that are likely to interact with other proteins in a complex. Solving the structure of pfTBP complexed with its specific (?) sequence will be valuable. The canonical TATA sequence may not be the binding site for pfTBP.

The differences in *Plasmodium* CTD and TBP may offer exciting clues. If pfTBP-DNA interaction is kinetically equivalent to the mammalian system, we could use the protein-DNA complex to assemble other factors in an *in vitro* transcription system for studying parasite gene expression. The feasibility of this approach is substantiated from my work (Datta, S., unpublished) using fractionated components from the HeLa system [Buratowski S *et al* (1989) Cell 56 549-561]. *Plasmodium* CTD used as an affinity tag may identify interacting target molecules (unique to the infected state) of therapeutic interest.
In *Trypanosomes*, hybridization with TBP probes had limited success [George Cross, personal communication] in view of differences between ‘classical’ and parasite TBP. In *Trypanosomes*, two abundant messages [VSG and PARP] are transcribed by α-amanitin resistant RNA polymerase in parasites containing disruption in 2 of the 4 alleles coding for the largest subunit of RNA polymerase II [Chung *et al* (1993) *MCB* 13 3734-3743]. There is little doubt that TBP interacts with most RNA polymerases [Sharp PA (1992) *Cell* 68 819-821] and *Trypanosome* TBP identification may take the pol I route.


Once characterized, RNA polymerase subunit[s] and transcription factors may be mutated and transfected in parasites containing disruption of the chromosomal gene [Steven Beverly, *Leishmania*; Lex van der Ploeg, *Trypanosome*]. Developing tools to use suppressor genetics in parasites may further aid in dissecting the transcription complex.
Molecular Medicine: Statement of Research & Pre-proposal for Transcription Targets in Molecular Parasitology

PRE-PROPOSAL

Introduction

Parasitic diseases are a major cause of mortality and morbidity worldwide. In the tropics, it causes about 200 million new infections and claims more than 2 million lives, annually. In 1992, there were about 800,000 child deaths due to malaria in Africa [1]. The re-emergence of opportunistic parasitic infections in the affluent industrialized nations are increasing with the spread of AIDS.

Diseases caused by animal parasites afflicted our ancestors. Ancient documents from China, Assyria and India describe high intermittent fevers, typical of malaria. During 5th century B.C., Hippocrates characterized the fevers and symptoms of malaria while Aristotle wrote about [the cysticercus stage of] a certain worm [*Taenia solium*] in the muscles of pigs. Probably, the first scientific report of a parasitic human infection [caused by the intestinal tapeworm *Diphyllobothrium latum*] was documented in 1602 by Plates, a Swiss physician [2].

Malaria still presents the single greatest threat in the tropics. In 1880, Laveran described the asexual stages of *Plasmodium falciparum*, the causative agent of human malaria. In 1898, Sir Ronald Ross, working in Calcutta (India), showed that the parasite developed in the mosquito and was transmitted by the bite of the insect [3]. Almost immediately, malaria control became synonymous with mosquito control. Ross and Laveran received the Nobel Prizes for their seminal contribution.

Treatment of malaria has not changed much since the 17th century, when the Peruvian bark, cinchona, was first used as chemotherapy. Quinine and cinchonine, the alkaloids of the cinchona tree, were isolated in 1820 and synthetic anti-malarials were developed in Germany [Pamaquine in 1924 and Chloroquine in 1934], England [Proquanil in 1944] and USA [Primaquine in 1952]. Prevention of malaria by mosquito control gained impetus with the development of the insecticide DDT. However, DDT has the potential to wreck the ecological balance and promote other disasters.
In 1955, to control a malaria epidemic in Borneo, the World Health Organization [WHO] sprayed the country with DDT which killed mosquitoes and houseflies. The dead houseflies were consumed by geckos, which in turn died. Without the geckos to prey upon them, Borneo's rat population exploded and resulted in a bubonic plague epidemic. To contain the plague epidemic, WHO air-dropped thousands of cats on the island ["the day it rained cats"]. Several thousand additional mortalities resulted from the second [plague] epidemic before the cats finally brought the rat population under control.

Molecular biology and biotechnology may help produce a malaria vaccine but it is still a work in progress. One claim of a "highly effective" immuno-therapy was only effective in one-third of patients tested [4]. However, parasitology is not completely devoid of molecular milestones. In 1976, Trager and Jensen succeeded to grow human malaria parasites in continuous culture [5]. Transfection procedures were established for Trypanosomes [6], Leishmania [7], Plasmodium [8] and Toxoplasma [9]. Successful gene replacement and disruption has been achieved in Leishmania [10] and Trypanosome [11]. Using an oligonucleotide screening strategy, the equivalent of the human and yeast TATA binding protein [TBP] was isolated from Plasmodium falciparum [12]. Recently the isolation of a red blood cell [erythrocyte] receptor used by Plasmodium to invade the cell was reported [13].

The isolation of TBP from malaria parasite and establishment of transfection procedures for Plasmodium, represent steps toward understanding transcription in Plasmodium since TBP is essential to transcription in eukaryotes [14]. Proliferation of the parasite in the host is dependent on its ability to initiate transcription of the essential genes and direct its asexual stages of development.

After a mosquito bites, the salivary fluid containing sporozoites [motile forms of the parasite] are delivered in the blood. Infection initiates following entry of sporozoites in liver parenchymal cells where parasites divide asexually. For P. falciparum, within 5-7 days, each infected liver cell releases about 40,000 progeny [merozoites] which invade red blood cells and initiate the erythrocytic phase of infection. Within the erythrocyte,
the parasite grows to form early trophozoite, followed by asexual division to become a schizont composed of merozoites. The erythrocytic cycle completes with rupture of the erythrocyte within 48 hours and release of merozoites, which proceed to invade other erythrocytes. The cycle continues with tertian periodicity [48 hours] unless intervened.

Antibodies directed against stage-specific antigens were mostly ineffective due to the intra-cellular location of parasite. Chemotherapy aimed at retarding proliferation of parasites adversely affects normal cells. The mechanism by which the parasite attaches itself to the erythrocyte, and subsequently gains entry, may be better understood with the recent identification of the receptor [13]. However, the ability of the parasite to multiply in the liver and erythrocyte presents a key threat.

Universally, almost all biological processes are mediated through the action of proteins. Transcription of DNA is the first step leading to protein synthesis. Initiation of transcription is a complex process. Biochemical fractionation of extracts made from mammalian cells and mutational analysis combined with suppressor genetics in yeast, have yielded a wealth of information. TBP, isolated as a key component of basal transcription by RNA polymerase II, is a common denominator in transcription by all three eukaryotic RNA polymerases [14]. Most subunits of yeast RNAPII have been cloned and sequenced [15]. TBP, along with other basal transcription factors [mostly specific for the type of RNAP] and RNAP are necessary components of the minimal transcription apparatus [16].

Comparison of the largest subunit [RPB1] of Plasmodium RNAPII with yeast and human RPB1, revealed striking homology and diversity in the conserved carboxy terminal domain [CTD] of RPB1 [15]. Biochemical analysis and suppressor genetic studies in yeast have identified transcription factors which interact with the CTD [17, 18]. TBP from Plasmodium, plant, yeast, frogs and humans show homology in several region but Plasmodium TBP diverges in a region that map to a RNAP interaction site and in another region that may interact with TFIIA, a RNAPII basal transcription factor [14].
The structural genes of *Plasmodium* may be transcribed by RNAPII. However, a similar notion held for VSG genes was dispelled when VSG transcription was unaffected in *Trypanosomes* with chromosomal disruptions in two RNAPII alleles [11]. Irrespective of which RNAP is involved, *Plasmodium* TBP may form complexes with other RNAP basal factors, providing a common biochemical theme. The homology and diversity of *Plasmodium* CTD offers a biochemical probe for isolation of *Plasmodium*-specific CTD interacting factors that may regulate transcription and gene expression.

**Objective**

The primary objective is to identify Malaria parasite [*Plasmodium falciparum*] specific transcription factors, which may serve as targets for therapeutic intervention or provide clues for designing novel vaccine strategies for Malaria. However, within the two year period specified in the grant, it may be feasible only to undertake the most direct and defined aspects of this proposal, as follows:

1. Identifying and characterizing protein[s] that interact with the *Plasmodium* CTD.
2. Identifying and characterizing protein[s] that interact with the *Plasmodium* TBP.
3. Cloning the genes of proteins that interact with the *Plasmodium* CTD and TBP.

**Hypothesis**

Transcription factors regulate and maintain host-parasite interactions. Survival of the parasite necessitates such proteins to provide common functions to "bridge" the host-parasite functional chasm. Characterization of proteins involved in such "cross-talk" may reveal novel targets for small molecule drug discovery geared toward therapeutic intervention. Selective disruption of the parasite-specific component may abrogate the host-parasite interaction without adversely affecting the host physiology.
Rationale

[1] The sequence of RNAPII subunits are conserved in evolution. The three largest subunits [RPB1, RPB2, RPB3] are homologues of bacterial RNAP and their sequences are conserved from bacteria to humans [15]. In most eukaryotes, the C-terminal domain (CTD) of RPB1 contain a consensus hepta-peptide repeat sequence [Tyr-Ser-Pro-Thr-Ser-Pro-Ser]. The sequence is conserved amongst species and the number of repeats increase with the complexity of the organism [yeast = 27, mouse = 52]. Yeast cells with truncated CTD containing fewer than 10 repeats are non-viable. The CTD of yeast RPB1, interacts with RNAPII transcription factors [17, 18]. The sequence of Plasmodium CTD shares homology with other organisms for about 15 repeats and then diverges to form a 63 aa tail with no homology to any published sequence [15]. This feature may be the hypothetical "bridge" and/or a template for "cross-talk" in host-parasite interactions.

In other words, the conserved region of the Plasmodium CTD "acts" like the host [human] CTD and hence utilizes the host functions for transcription. The unique region of Plasmodium CTD may be specific for the biology of the parasite. This, therefore, is a target with no cellular counterpart. Extrapolating from functional conservation, this unique region probably interacts with one or more [transcription] factors, perhaps, in a manner similar to the interaction of transcription factors with yCTD [17, 18]. Identifying factors that interact with Plasmodium CTD [factors specific for the unique region] may offer Plasmodium-specific targets for therapeutic intervention.

[2] Kinetic studies reveal that the first, necessary and essential, step in the formation of a pre-initiation complex for transcription initiation by RNAPII, is the sequence-specific binding of TBP [TATA binding protein] to a TATA-containing promoter [19]. But, RNAPII mediated transcription from TATA-less promoters also contain TBP in the pre-initiation complex [14]. In addition, transcription initiation complexes formed at promoters transcribed by RNAPI and RNAPIII contain TBP, attesting to the universal requirement of TBP in eukaryotic transcription. The isolation of Plasmodium TBP makes it possible to study the biochemical assembly of the pre-initiation complex with TBP anchored to TATA sequence, as the first committed step. However, it may be argued,
given the A+T rich genome of *Plasmodium*, that TATA sequences may be arbitrary. The latter raises doubts about the specificity of the TBP-TATA interaction in *Plasmodium*. However, the following two observations may be cited in favour of the TBP interaction:

[A] During the stepwise assembly of the pre-initiation complex on a TATA-containing promoter, the TATA sequence is bound by TBP and several other TAFs or TBP associated factors [14]. The TBP-TAF complex is referred to as TFIID. The TATA-TFIID complex binds TFIIA [RNAPII basal transcription factor] to form the DA complex. In the third step, TFIIB binds to the DA complex to form the DAB complex. A preformed TFIIF-RNAPII complex associates with the DAB complex, followed by association with TFIIE and TFIIH [20]. Therefore, the kinetics of binding of TBP to the TATA sequence may be subject to modification by one or several associated proteins. While some of the protein-protein contacts with TBP have been mapped, the list is still growing. However, the most striking divergence of *Plasmodium* TBP from other TBPs, coincides with TFIIA contact points which maps in helix 2 [H2] of TBP [21]. Hence, is there a *Plasmodium* specific TFIIA that helps modify TBP-TATA specificity? By conserving some features of the host pre-initiation complex but modifying others, the parasite may have evolved a transcriptional machinery that is compatible with yet distinct from the host. Using *Plasmodium* TBP as a “hook” it may be possible to identify associated parasite-specific factors. The latter may be direct targets or provide clues for novel therapeutics.

[B] TATA-TBP DNA-protein interaction is only one of many possible interactions of TBP [14]. In RNAPI and RNAPIII interactions, TBP interacts with other proteins or protein-DNA complex. *Plasmodium* TBP may interact with accessory proteins and the resulting protein-protein complex, then, may bind with sequence specificity to DNA. It is of interest to note that a highly conserved Proline residue is located in β sheet 1 of all TBPs except *Plasmodium* TBP [12, 21]. By mutational analysis in yeast, it has been shown that replacing this Proline with any other amino acid abolishes the ability of the mutant TBP to sustain transcription by RNAPII and RNAPIII *in vivo* [21] but the mutant TBP remains functional for RNAPI transcription *in vivo*. It will not be without precedent if some *Plasmodium* mRNAs may be RNAPI transcripts, similar to transcription of VSG genes in *Trypanosome* [11].
Methods

A. Preparation of Extracts

B. CTD and TBP Affinity Chromatography

C. DNA Binding Specificity of *Plasmodium* TBP

[A] Preparation of Extracts

Large-scale cultivation of the erythrocytic stage of *P. falciparum* is an established procedure and several methods exist in the literature to make protein extracts [22]. The procedures used for making whole cell and nuclear extracts from mammalian cells will be tested [23]. Extract suitability will be tested by reproducible retention of specific proteins by affinity chromatography and analytical gel electrophoresis to estimate integrity of proteins. To help enrich for any protein that may be retained on the affinity columns, each batch of extract will be fractionated over DEAE and CM matrices and eluted with step gradients. Individual fractions will be concentrated and applied to CTD and TBP affinity columns.

[B] CTD and TBP Affinity Chromatography

Several protocols exist to identify interacting proteins [24]. In addition, GST fusions or similar methods exist where the base protein [GST] is attached to the bead while the fusion [CTD] is free to interact [25]. *Plasmodium* CTD will be subcloned independently or as a fusion and CTD or CTD-fusion purified to homogeniety. This will be immobilized on beads to serve as the CTD affinity column. To select proteins that bind to the unique region of the CTD, the repeats will be deleted in the expression clone or the unique peptide will be synthesized. In the latter case, reversible photoactivated cross-linkers may be used to identify proteins that interact with the peptide [26]. Yeast CTD will serve as control. Similar strategy will be followed for *Plasmodium* TBP [gift of John Hyde, UMIST]. Retained proteins will be sequenced and used for cDNA screening.
[C] DNA Binding Specificity of *Plasmodium* TBP

TATA oligonucleotides and several mutant combinations [gift of Kevin Struhl, Harvard] will be used in gel retardation assays to determine the sequence and binding specificity of *Plasmodium* TBP, if any [27]. If specific binding is detected, then partially fractionated *Plasmodium* extracts will be tested to determine if higher order complexes [pre-initiation complexes?] can form on such a template [20].

Conclusion

Knowledge about RNAPII and TBP obtained from other eukaryotes [humans, yeast] is the key to frame important medically relevant questions about the malaria parasite, *Plasmodium falciparum*, and its interaction with humans. Characterization of transcription factors that mediate parasite-specific functions are of medical significance. The simplicity of this approach may unambiguously identify proteins of interest. The quest for therapy is based on understanding, identifying and disrupting these key molecule[s] involved in the biology of the host-parasite interaction.
References


Summary of Education & Positions

1976
School Leaving Certificate [GCE]
University of Cambridge, Cambridge, UK [Local Examinations Syndicate]

1980
BSc [Honours] Biochemistry and Medical Physiology
Minors: Physics and Chemistry
Presidency College, University of Calcutta, Calcutta, India

1988
PhD [Molecular Biology and Genetics]
Rutgers University School of Medicine & collaboration with Princeton University

1989-1991
Research Fellow in Medicine
Harvard University & Harvard Medical School; Massachusetts General Hospital

1991-present
Research Associate
Whitehead Institute, Massachusetts Institute of Technology

1991-present
Instructor in Medicine
Harvard Medical School

Proposed - 1994
Research Scientist
University of California, Berkeley and
University of California, San Francisco
UCSF School of Medicine

CONTACT INFORMATION

Dr Shoumen Datta
Whitehead Institute for Biomedical Research
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142-1479

Tel: [617] 252-1940
Fax: [617] 252-1902
Eml: shoumen@mit.edu
ABSTRACT

Regulation of transcription initiation provides a key mechanism for control of gene expression. Basal transcription initiation factors, best characterized in mammals, and RNA polymerase II, which is best understood in yeast, are two important components of the transcription apparatus. The objective of this research is to identify and characterize the components of the basal transcription apparatus from yeast. This would allow us to employ the powerful combination of yeast genetics and biochemistry to advance our understanding of basal transcription initiation. As a first step, we aim to fractionate yeast nuclear extracts in order to reconstitute an in vitro transcription system and to analyze the assembly of initiation complexes by native gel electrophoresis. Information and reagents available from fractionation of mammalian extracts are likely to aid in identifying the yeast homologues of the basal transcriptional machinery. In the next step, we plan to purify and clone basal transcription initiation factors that are unavailable. The recombinant factors will be used to replace one or more fractions in in vitro transcription. The information obtained from these experiments will improve our understanding of the basal transcription initiation apparatus and may provide clues to the mechanisms by which the basal transcription apparatus is regulated.

SPECIFIC AIMS

To improve our understanding of the mechanisms of eukaryotic transcription initiation, the project aims to:

1) fractionate yeast nuclear extracts to obtain highly purified components of the basal transcription initiation apparatus to reconstitute an in vitro transcription system capable of accurate initiation;

2) identify basal transcription initiation factors in yeast that have mammalian counterparts by using a gel retardation assay;

3) purify transcription factors to homogeneity and clone and express their genes.

RELATIONSHIP TO THE CANCER PROBLEM
Cancer results from aberrant gene expression. Regulation of transcription initiation provides a key mechanism for control of gene expression. Yet the basic process of transcription initiation and the mechanisms that regulate this process are poorly understood. In this project we propose to characterize the components of the basal transcription initiation apparatus in eukaryotes. A well defined transcription initiation process will provide the foundation to investigate the mechanism by which deregulation of this process can cause cancer. Protein products of several oncogenes (FOS, JUN, MYC, etc.) are important regulatory transcription factors which probably act at the level of transcription initiation to deregulate normal cellular processes. Information obtained by comparing normal proteins with their oncogenic counterparts, will aid in improved drug design and effective therapy.

INTRODUCTION

The process of transcription initiation by RNA polymerase II is complex, yet we have some clues [4, 29, 38, 41, 50]. Components of the transcription initiation apparatus are currently (1991) the focus of considerable attention. However, basal transcription initiation factors in yeast are largely unidentified, despite a relatively detailed picture of yeast RNA polymerase II. The availability of yeast
3 transcription factors would allow us to combine the power of yeast genetics and biochemistry to understand the details of the process of transcription initiation.

In the following sections we present an overview of the field and the rationale to pursue the proposed research:

A. The basal transcription apparatus and initiation of transcription

B. The components of the basal transcription apparatus
   1. TFIID
   2. TFIIA
   3. TFIIB
   4. TFIIF, TFIIE
   5. TFIIG, TFIIH
   6. RNA polymerase II

C. Developing a reconstituted yeast transcription system

D. Suitability of the laboratory environment for this project

The basal transcription apparatus and initiation of transcription

Extracts made from mammalian cells [35] or nuclei [12] support in vitro transcription of TATA-containing templates. To understand the nature of the basal transcriptional machinery, it was necessary to isolate the individual components. Chromatographic fractionation of HeLa cell nuclear extracts and reconstitution of basal transcription in vitro indicated that multiple components are contained within the basal transcription apparatus [37, 40, 47]. Kinetic studies showed that the first committed event in transcription initiation is the sequence-specific binding of TFIID (Transcription Factor, RNA polymerase II, fraction D) to a TATA-containing promoter sequence [11, 15, 46].
Our understanding of the stepwise assembly of basic transcription apparatus in mammalian cells is based on Buratowski et al. [1]. Using partially purified transcription factors and a gel-retardation assay, the authors demonstrated the sequential assembly of the transcription apparatus on a minimal promoter [1]. In summary, the key event is the binding of TFIID to the TATA element. TFIID, which is unable to bind to DNA by itself, binds to the TATA element. Addition of TFIIB creates the DAB complex which associates with a preformed TFIIF-RNA polymerase II macromolecule followed by association with TFIIE [9, 18]. TFIIG [44] and TFIIH [30] have been identified in pre-initiation complex. TFIIG associates with the complex prior to the DABF-polII step [44] and TFIIH is likely to enter the complex either subsequent to or with TFIIE [30].

The components of the basal transcription apparatus

**TFIID**

From fraction TFIID, a single protein was identified which binds to the TATA sequence. Yeast TFIID is a 27 kD protein. The TFIID gene is essential for yeast viability [14] and it has been cloned from humans [32], yeast [24, 27, 28, 42], drosophila [26] and plant [19]. The protein is strikingly conserved in its C-terminal 180 aa (between various species) while containing a highly variable N-terminal domain. Other proteins have been reported to associate with TFIID and may influence transcription [13].

**TFIIA**

The second step in the formation of the transcription complex requires TFIIA [1]. A 35-38 kD protein is the essential component of wheat and human TFIIA [5]. TFIIA from yeast or wheat can replace human TFIIA in a functional assay [5, 23]. Yeast TFIIA is a heterodimeric protein, both subunits of which have been cloned (S. Hahn, personal communication). TFIIA can be purified by
TFIID affinity chromatography [45] and it may associate with TFIID independent of DNA in vivo [1, 23].

**TFIIB**

In step three, DAB complex formation requires fraction TFIIB. A 30 kD protein is the essential component of TFIIB and the human TFIIB gene has been cloned [22].

**TFIIF, TFIIE**

The DAB apparatus associates with a preformed TFIIF-RNA polymerase II complex [9, 18]. Subsequently TFIIE enters the complex [30]. Slower migrating complexes containing these factors can be resolved by native gel electrophoresis [1]. Human TFIIF consists of 30 kD and 74 kD subunits, originally identified as RNA polymerase associated proteins (RAP30, RAP74) [6, 17, 43]. Human TFIIE is a hetero-dimeric protein with 34 kD and 56 kD subunits [30, 39].

**TFIIG, TFIIH**

Mammalian transcription complexes have been reported to contain TFIIG [44] and TFIIH [30]. TFIIG is required for preinitiation complex formation prior to association of RNA polymerase II. It shares functional similarities with, but is chromatographically distinct from, TFIIA [44]. TFIIH enters the complex either with or subsequent to TFIIE [30].

**RNA Polymerase II**

An essential component of the basal transcription apparatus is RNA polymerase II [41, 50]. The enzyme alone is incapable of accurate transcription initiation. Yeast RNA polymerase II is a complex enzyme containing 11 subunits [50 and references therein]. The yeast subunit genes have all been cloned. The three largest subunits are homologues of the prokaryotic RNA polymerase core subunits. Some of the human RNA polymerase II subunit genes have been cloned and sequence analysis reveals that they are closely related to their yeast counterparts. A unique feature of the largest subunit of eukaryotic RNA polymerase II (RPB1) is the presence of a repeated hepta-peptide at its carboxy-terminal domain (CTD). The hepta-peptide repeat sequence is highly conserved amongst species and the number of repeats increases with the complexity of the organism. The CTD appears to have a role in transcription initiation [10, 50].
Developing a reconstituted yeast transcription system

Fractionation of mammalian nuclear extracts and isolation of some of the human transcription factor genes have improved our understanding of transcription initiation [4, 41]. Detailed characterization of transcription factors is limited by the lack of an efficient genetic approach in mammals. For this purpose we have chosen to develop a yeast system which would provide the power to combine biochemical studies and genetic characterization. The high degree of conservation of the transcriptional process [5, 7, 8, 20, 21, 23, 31, 36] suggests that models derived from studies on yeast will be applicable to higher eukaryotes.

Transcription competent yeast nuclear extracts have been reported [33] and a preliminary fractionation scheme has been published [16]. This fractionation scheme does not allow investigators to relate the yeast fractions to transcription initiation factors already described in the mammalian system. Establishing a comparable system in yeast is still necessary.

We plan to use reconstitution of in vitro transcription activity and the gel retardation assay [1] to monitor separation of individual components and assign yeast functional homologues of the mammalian basal transcription initiation factors. A few components of the basal transcription apparatus are at hand from the yeast system: the cloned yeast TFIID gene and purified recombinant yeast TFIID and various mutant versions of the protein (S. Buratowski), cloned yeast TFIIA genes (S. Hahn, personal communication); cloned yeast RNA polymerase II subunit genes, purified yeast RNA polymerase II holoenzyme (S.-M. Liao, S. Buratowski and R. Young). In addition, work in our laboratory is underway to clone yeast TFII B and TFII F genes using mammalian gene probes (S.-M. Liao, J. Zhang, S. Buratowski and R. Young).
In contrast to the mammalian \textit{in vitro} transcription system, which was developed more than a decade ago [35], the yeast \textit{in vitro} system is relatively new [33, 48] and still undergoing modifications [49]. Information obtained from extensive fractionation of HeLa cell nuclear extracts by a number of laboratories over several years made it possible for Buratowski \textit{et al} [1] to demonstrate the assembly of intermediate complexes of a transcription initiation apparatus. Fractionation of yeast nuclear extracts is still in its preliminary stages [16]. Without a reproducible fractionation scheme and assays to monitor \textit{in vitro} transcription initiation and assembly of initiation complexes, it will be difficult to dissect the individual components which make up the basal transcription apparatus.

The objective of this research is to identify individual components of the yeast basal transcription apparatus and to develop reconstituted \textit{in vitro} transcription system. It is expected that this system will be used in combination with a genetic analysis of RNA polymerase II and transcription factor genes. Mutant proteins can then be tested functionally both \textit{in vivo} and \textit{in vitro}.

EXPERIMENTAL APPROACH
The power of resolution of column chromatography makes this method key to the identification and purification of basal transcription initiation factors. Yeast nuclear extracts will be fractionated and assayed for transcriptional activity in vitro by reconstituting various fractions. Active fractions will be used for assembly of initiation complexes and analyzed by the gel retardation assay [1]. Genetic methods and gene cloning subsequent to protein purification will be employed. The experimental approach is discussed in subsections as follows:

A. Yeast nuclear extracts
B. Fractionation and reconstitution of in vitro transcription
C. Gel retardation assay for assembly of initiation complex
D. Preparative purification of transcription factors
E. Transcription factors identified by genetic methods

Yeast nuclear extracts

Preparation of highly active nuclear extracts capable of supporting transcription in vitro is essential to any fractionation scheme. We have made several important modifications to the original yeast nuclear extract protocol [33]. To allow cells to recover adequately from cell wall digestion, the regeneration time has been increased. Nuclei were extracted with ammonium sulfate and contaminating nucleic acids were removed by DEAE-Sephacel treatment [34]. The supernatant was precipitated with solid ammonium sulfate and the recovered protein pellet was resuspended and dialyzed. These modifications resulted in substantial improvement in the quality of the in vitro transcription signal.

Fractionation and reconstitution of in vitro transcription

Yeast nuclear extracts have been chromatographed over DEAE Sepharose and step-eluted with increasing salt concentration. The fractions will be assayed in various combinations to reconstitute transcription activity in vitro. The active fractions from this step will be subjected to further rounds of ion exchange and molecular sieve chromatography to enrich for transcription factors. Reconstitution assays will be performed with each fraction and subfraction to continually
optimize for transcriptional activity. We have purified yeast RNA polymerase II and purified recombinant yeast TFIID to supplement \textit{in vitro} reconstitution assays, as appropriate. These experiments will allow us to identify the yeast homologues of the mammalian basal transcription initiation factors. With increased fractionation, the yeast system may identify additional basal transcription initiation factors. The ultimate goal is to use these purified factors to establish a standardized reconstituted \textit{in vitro} transcription system capable of accurate initiation.

\textbf{Gel retardation assay for assembly of initiation complex}

The reconstitution assay described above identifies active fractions that are sufficient and necessary for basal transcription but lacks in its ability to assign yeast homologues of the mammalian system. To identify which yeast fractions contain the functional equivalent of the previously identified mammalian basal transcription factors, the gel retardation assay will be used [1]. In brief, purified recombinant yeast TFIID and the same promoter used for \textit{in vitro} reconstitution assays will be used to generate the first complex. To this we will add transcriptionally active fractions and monitor the assembly of initiation complexes. It is likely that the yeast pattern of the transcription intermediate complexes will resemble previously identified mammalian complexes and allow us to tentatively identify the yeast homologues of the mammalian factors. To confirm the identity of these homologues we will pursue further experiments.

\textbf{Preparative purification of transcription factors}

Having established a functional fractionation scheme we will isolate preparative amounts of initiation factors. Information from protein micro-sequencing will help in the design of oligonucleotide probes for PCR and hybridization to cDNA libraries to clone the transcription factor genes.

Yeast cells may contain limited copies of some basal transcription factors, and we may find that it is difficult to obtain sufficient amounts of protein from fractionation of nuclear extracts. As an alternative approach, we could use yeast whole cell extracts (yWCE) [49]. The latter extract is easier to prepare and can be scaled up to a greater degree than preparation of nuclei. Though initial yWCE
preparations were less active than nuclear preparations, recent improvements generate yWCE that are highly active [49] and it was successfully used as a source to purify yTFIIA [23].

**Transcription factors identified by genetic methods**

Genetic suppressor analysis has led to the identification of genes whose products may interact with TFIID (S. Buratowski and H. Zhou, personal communication) and/or the largest subunit of RNA polymerase II (A. Koleske, C. Thompson and R. Young, personal communication). The protein products of these suppressor genes may interact with the basal transcription machinery in vivo. We have produced recombinant proteins using the cloned suppressor genes and will determine whether these proteins are active in the in vitro reconstitution assay and the initiation complex assembly assay.

**CONCLUDING REMARKS**

Establish a reconstituted in vitro transcription system using purified yeast basal transcription factors, in a manner that will allow us to relate the yeast components to the mammalian basal transcription initiation factors. The completion of the proposed project will advance our understanding of the process of transcription initiation. It will provide a foundation for identifying biochemical targets within the transcription apparatus which respond to activators and repressors of transcriptional activity.
REFERENCES
