Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

1994

A Molecular Characterization of Human General Transcription Factor IID

Alexander Hoffmann

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

Part of the Life Sciences Commons

Recommended Citation

Hoffmann, Alexander, "A Molecular Characterization of Human General Transcription Factor IID" (1994). *Student Theses and Dissertations*. 353.
http://digitalcommons.rockefeller.edu/student_theses_and_dissertations/353

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.

LD4711.6
H711
c.1
RES

THE LIBRARY

LD 4711.6 H711 1994 c.1 RES
Hoffmann, Alexander.
A molecular characterization
of human general

Rockefeller University Library
1230 York Avenue
New York, NY 10021-6399

A Molecular Characterization Of Human General Transcription Factor IID

Alexander Hoffmann

Thesis submitted to the Faculty of The Rockefeller University in partial fulfillment of the requirements for the degree of Ph.D.

c Copyright by Alexander Hoffmann, 1994

Table of contents

List of Figures.....vi

Acknowledgment.....x

Epigraph.....xii

Abstract.....1

Preface 3

Introduction

(i) An Introduction to Transcription

Definition of transcription.....5
Importance of transcription in development.....5
Approaches to study transcription.....7
Transcription in *Escherichia coli*.....10

(ii) The Biochemistry of Eukaryotic Transcription

Prelude to accurate transcription *in vitro*.....14
The first accurate *in vitro* transcription systems.....15
In vitro transcription in 1988.....19
In vitro transcription in 1994.....24

(iii) An Introduction to TFIID

Identification, chromatographic behaviour.....33
Promoter interactions.....36
Role in template commitment.....38
Implicated in trans-activation mechanisms.....39
Purification of yeast TFIID.....43
Cloning of yeast TFIID.....46

Chapter I: cDNA cloning of TFIIDs

Introduction.....49

(i) *Schizosaccharomyces pombe* TFIID

Isolation of genomic and cDNA clones.....53

Functional analysis.....63

The conserved TFIID 'core' domain.....66

(ii) TFIID from higher eukaryotes

Isolation of human cDNAs.....75

TFIID cloning from mouse, insect, and plant species...86

Homologies in TFIIDs across the eukaryotic spectrum...88

(iii) Comparing recombinant vs. native human TFIID

Functional analysis of the human cDNA-encoded protein.98

TFIID gel shift: interacting factors.....104

Chapter II: Identifying other subunits of TFIID

Introduction.....121

(i) Attempts at TBP-affinity chromatography and reconstitution of TFIID *in vitro*

Strategies.....126

Discussion.....130

(ii) Immunopurification of TBP-containing complexes

Immunopurification of TFIID.....133

Immunopurification of $TFIIIB_2$.....136

Identification of direct TBP-TAF interactions.....139

(iii) Purification of vaccinia expressed His-TBP

Rationale and purification strategies.....143

Characterization of His-TBP-containing eluates.....146

(iv) Purification of epitope-tagged TFIID

Strategy and methodology.....155

Identification of *bona fide* TFIID subunits.....167

Attempts at mapping the TFIID-complex architecture...172

Chapter III : Studies with recombinant TBP

(i) Bacterial His-tag expression system

Construction of His-pET expression vectors.....179

Non-denaturing purification methodology.....183

(ii) Structural studies of TBP

Introduction.....188

NMR analysis attempts.....189

X-ray crystallography.....193

(iii) Designing a convenient mutagenesis of TBP

Introduction and rationale.....200

A cassette encoding the human TBP core domain.....206

Chapter IV : Characterization of TBP-associated TFIID subunits: TAF20/15

Introduction.....212

(i) Primary structure analysis of human TAF20/15

Cloning of human cDNAs encoding TAF20 and TAF15.....213

Sequence analysis and homologies.....220

(ii) Mapping protein-protein interactions

Protein-protein interactions within TFIID.....226

Construction and interaction potential of TAF20 mutants

Discussion.....240

Summaries, Conclusions and Perspectives

Introduction.....244

(i) The TFIID complex

Characterization TFIID subunits.....250

Promoter interactions by TFIID.....256

Function in mediating *trans*-activation.....262

(ii) Perspectives on transcription intiation

Class II initiation pathways.....267

Universal characteristics of initation pathways.....268

List of Figures

Introduction

1. Fractionation of HeLa nuclear extract on P11 18
2. Transcription assay with TFIIA-containing DEAE Sepahacel fractions 21
3. Identification and purification scheme of the general transcription factors 26
4. Summary of the class II general transcription factors 28

5. Characteristics of the TFII activity in the phosphocellulose (P11) D fraction 35
6. Model for TFIID's role in promoter activation by ATF 42
7. Schematic of the sequence motifs found in *Saccharomyces cerevisiae* TFIID 47

Chapter I

8. Strategy for cloning TFIIDs from higher eukaryotes 51

9. Detection of *Schizosaccharomyces pombe* TFIID gene on a genomic Southern blot 55
10. Cloning TFIID-encoding genomic fragments from *Schizosaccharomyces pombe* : Miniprep analysis 57
11. Nucleotide and predicted amino acid sequences of the *Schizosaccharomyces pombe* TFIID gene 59
12. Mapping of *Schizosaccharomyces pombe* TFIID mRNA 62
13. Expression and functional analysis of cloned *Schizosaccharomyces pombe* TFIID 65
14. Sequence comparison of TFIID genes from *Saccharomyces cerevisae* and *Schizosaccharomyces pombe* 70
15. Schematic of the sequence motifs found in the newly identified TFIID core domain 73

chapter I cont'd.

16. Cloning human TFIID by polymerase chain reaction (PCR) 77
17. Nucleotide and predicted amino acid sequence of human TFIID cDNA 79
18. Matrix comparison of amino acid sequences of TFIID from *Saccharomyces cerevisae* and human 82
19. The human TFIID gene and mRNA 85
20. Sequence comparison of the highly conserved TFIID core domain 90
21. Sequence motifs in the N-terminal structure of the human TATA factor 94
22. Evolutionary conservation of the TATA factor 97

23. *In vitro* expression and function of human cDNA-encoded TFIID in TATA box binding and basal transcription 100
24. USF-stimulated transcription with TFIID 103
25. Gel mobility shift assays with *in vitro* translated yeast and human TFIIDs 106
26. Gel mobility shift assay with bacterially expressed human TFIID 108
27. Rabbit reticulocyte lysate and $P11_{0.1}$ fraction contain a factor that stimulates bacterially expressed human TFIID binding to the TATA box 111
28. Bacterially expressed human TFIID combined with purified human TFIIA form complexes with TATA box-containing DNA fragments of the same mobility as partially purified human TFIID 113
29. Comparison of oligonucleotide competition behaviour of native TFIID with *in vitro* expressed TFIID 116
30. The $P11_{0.5}$ fraction contains another TFIID-interacting factor 119

Chapter II

31. Contrasting native TFIID with TBP (TFIIDτ) 123

32. Schemes for *in vitro* reconstitution of the TFIID complex and subsequent affinity purification 128

33. Immunoprecipitation of TFIID with α-hTBP antibodies 135
34. Immunoprecipitation of $TFIIIB_2$ with α-hTBP antibodies 138
35. Far Western of HeLa-derived fractions probed with TBP 141

36. Overexpression of His-hTBP in HeLa cells using the recombinant vaccinia virus system 145
37. Gel shift assay with His-$hTBP^{vac}$-containing fractions 148
38. Testing for activation by USF with His-$hTBP^{vac}$-containing fractions 150
39. Analysis of polypeptides present in His-$hTBP^{vac}$-containing fractions 153

40. Strategy for the native purification of epitope-tagged TBP-containing complexes 157
41. Identification of cell lines expressing HA-hTBP 160
42. Fractionation of HA-TBP across phosphocellulose 162
43. Assessing efficiency of immuno-affinity purification 164
44. Analysis of polypeptides associated with HA-TBP 169
45. Is TopoI a subunit of the TFIID complex ? 171
46. Attempts at partially disrupting the TFIID complex 175
47. Summary of partial disruption experiments 177

Chapter III

48. Construction of His-pET plasmids 181
49. Expression and affinity purification of His-yTBP 185

chapter III cont'd.

50. Limited trypsin digestion of yTBP 192
51. Ribbon diagram of At2-TBP 195
52. Space filling model of At2-TBP complexed to promoter 198

53. Proteins capable of interacting directly with TBP 202
54. Summary of TBP sequence data 205
55. Designing a cassette construct encoding the human TBP core domain for convenient mutagenesis 208
56. Sequence of the resulting syn hTBP cDNA 210

Chapter IV

57. cDNA and amino acid sequence of human TAF20 215
58. Western blot analysis of TAF20/15 in human TFIID 217
59. Immunoprecipitation with α-TAF20 antibodies 219
60 Matrix between human TAF20 and *Dosophila* TAF28 222
61. TAF20 shows sequence homology to histone H2B 224

62. Identifying TBP-independent protein interactions within native TFIID 228
63. Mapping interaction domains of TAF20 232
64. Summary of interaction studies 234
65. Human TAF20-TBP interaction is not species-specific 236
66. Proposed model of TAF20 tertiary structure and protein interaction surfaces 239

Conclusions and Perspectives

67. Comparison of TFIID subunits in yeast, *Drosophila*, and human 247
68. Human TAFs: protein interactions 249
69. Model of TFIID on a class II promoter 260
70. Comparison of initiation pathways 271

Acknowledgment

Biochemical studies are best suited to exploit the power of the reductionist approach that emphasizes individual constituents rather than larger aggregate systems to understand complex biological regulatory mechanisms in detail. The work described in this thesis contains illustrations of many aspects of reductivism put to experimental work: the separation of activities, the identification and purification of the active component within a complex mixture, the complementation of an isolated component with less well characterized fractions to regain lost functional characteristics, the characterization of structure-function relationships of an individual component by direct structural means as well as functional analysis of mutant variants, etc. For my appreciation of the biochemical approach and for nurturing my fascination in modern experimental biology I would like to thank Drs. Ron Laskey and Bill Earnshaw whose early influence on me shaped my decisions concerning advisor and project for my graduate work.

Most importantly, I would like to thank my advisor, Bob Roeder, for providing the kind of work environment that encourages, strength, an independent critical mind, and pride, not only as a member of a laboratory with a tradition of excellent research and the rigorous application of the scientific method. Thanks go to members of my thesis committee, Drs. Jim Darnell, Claude Desplan, and Jim Manley who took time to lend advice and support when needed.

I am especially grateful to Masami Horikoshi who taught me more than anybody not just a repertoire of experimental techniques, but more importantly experimental design, strategy and focus. More particularly, I owe him my introduction to TFIID, decisive supervision in my early projects, and ultimately, much of the confidence that fuelled my subsequent independent work.

Finally, I would like to thank past and current members of the Roeder laboratory, collaborators and particularly attendants of the general factor meeting (Feb 89 to Mar 93) and the TFIID meeting (Sept 93 to ?) for countless discussions and reagents, as well as support, especially Carmen-Gloria Balmaceda, Richard Bernstein, Stephen Burley, Cheng-Ming Chiang, Hope Cohen-Webb, Alex Gasch, Hui Ge, Thomi Gerster, Polly Gregor, Elaine Halay, Satoshi Hasegawa, Koji Hisatake, Jeff deJong, Hildegard Kaulen, Jo Kim, TK Kim, Markus Kretzschmar, Dong-Kun Lee, Yan Luo, Sohail Malik, Earnest Martinez, Michael Meisterernst, Shona Murphy, Dimitar Nikolov, Thomas Oelgeschläger, Yoshi Ohkuma, Camilo Parada, Maria Popov, Eric Sinn, Henk Stunnenberg, Zhengxin Wang, Tony Weil, Tohru Yamamoto, Jong-Bok Yoon

I am much indebted to the Rockefeller University, the Beckman foundation and the Boehringer Ingelheim Fonds for graduate fellowships and financial support.

"It is not impossible that progress in biochemistry will allow one to bypass in some way genetic analysis by conventional means."

Jacob and Monod (1963)

Genetic repression, allosteric inhibition and cellular differentiation. In *Cytodifferentiation and macromolecular synthesis* (ed. M. Locke), pp.30-64. Academic Press, New York.

Abstract

The general transcription initiation factor IID plays a central role in transcriptional control as a direct target for a diverse array of gene-specific regulatory factors and as the only template-bound class II initiation factor.

The work described in this thesis concerns itself with the molecular characterization of this transcription initiation factor, starting with the cloning of cDNAs from a variety of organisms, including human, encoding a protein that may substitute native TFIID fractions in DNA binding and *in vitro* basal transcription assays. Sequence comparisons identify important structural motifs in the protein. Further functional analyses lead to the realization that this protein is the TATA box binding subunit (TBP) of a multi-protein TFIID complex whose other constituents are required for activator-responsive *in vitro* transcription and are responsible for TFIID's characteristic DNA interactions around the initiation region. A variety of biochemical approaches are discussed leading to the identification of more than a dozen class II TBP-associated factors (TAFs) that as a whole make up the TFIID complex and its characteristic functions.

A bacterial expression system allowing for the convenient purification of large amounts of recombinant protein in non-denaturing conditions is presented. This has allowed structural studies on TBP that have culminated in detailed X-ray crystallographic structures of TBP by itself or bound to the TATA box revealing unprecedented DNA distortions. The design of a convenient mutagenesis approach of TBP is presented that may lead to a more refined understanding of TBP and TAF functions.

Cloning of a small subunit (TAF20) of the TFIID complex is taken as a starting point to chart protein-protein interactions within the complex. Furthermore, sequence

homologies suggest possible functions for some TAFs and lead to a proposed revision of TFIID's functional role in eukaryotic chromatin and in mechanisms of transcriptional initiation and regulation.

Preface

Since its identification the activity of the phosphocellulose D fraction has become the most intensively studied eukaryotic transcription factor. As its characterization is intimately linked to the development of an *in vitro* transcription assay system and the study of transcriptional regulation I have attempted to begin the introduction with a brief historical account to emphasize the conceptual evolution surrounding transcription and to recall some of the fundamental questions and early studies underlying my own.

TFIID has been seen to play the central role not only in initiation complex formation and template commitment as the only DNA-bound general transcription factor, but also in the regulation of transcription as the apparantly primary mediator of *trans*-activation effects. With the availability of cloned cDNAs encoding general as well as specific transcription factors, a plethora of studies have addressed such issues in innumerable model systems. I cannot possibly do justice to all of these studies in my attempts to review the context in which my own results need to be discussed, but will have to refer only to the most relevant of these.

Note also that TATA-binding protein (TBP), identified as a subunit of the multi-protein complex TFIID during the course of this thesis work, is now also known to be a central component of multi-subunit complexes SL1/TIF-1B in class I and $TFIIIB_2$ in class III transcription. While future generations of students will undoubtedly approach the study of transcription and TBP in a more integrated fashion - being more aware of the parallels in transcription by three discrete RNA polymerases (with respect to mechanism and functionally homologous protein factors) - my own studies and conceptual education was still rooted in the old world of 'class division' in transcription. Therefore my own research work was limited to the class II transcription factor TFIID,

and thus I will also restrict my introduction to considerations of transcription initiation by RNA polymerase II from classic class II model promoters.

Related to this conceptual development is the change in nomenclature regarding TFIID/TBP. I have decided to mirror this change in the present thesis (with some exceptions for clarity's sake), not only to demonstrate the conceptual evolution but also to avoid confusion when referring to published works. Thus, in the first sections of chapter I, "TFIID", more often than not, really means "TBP". In the last section of chapter I the two are distinguished as "rec hIID" and "native TFIID", in the introduction to chapter II the nomenclature is discussed and clarified, and in the following chapters the current nomenclature is used throughout.

Introduction

(i) An Introduction to Transcription

Definition of transcription

In two seminal articles in 1962, reviewing the recent surge of genetic data on the regulation of gene expression, Jacob and Monod discuss the need for an informational intermediate, which they proposed to be "messenger RNA", between gene and protein (Jacob and Monod, 1962; Monod and Jacob, 1962). The existence of mRNA was elegantly demonstrated soon thereafter (Brenner et al., 1961; Gros et al., 1961). Transcription refers to the process surrounding the synthesis of such a ribo-nucleic acid copy from a deoxy-ribonucleic acid template catalyzed by a DNA-dependent RNA polymerase. RNA polymerization proceeds in the 5′ to 3′ direction in a linear fashion just like translation and thus preserves the collinearity between gene and polypeptide sequence.

Accurate (initiation of) transcription is an operative term referring to the precise physiological start sites used in an experimental system and is therefore an important criterion in evaluating studies on transcription for their biological relevance.

Importance of transcription in development

The study of transcription is motivated by a desire to understand how cells and organisms control the amounts of various gene products (e.g. proteins) which they synthesize and accumulate. Of particular interest are the mechanisms by which the protein composition of cells is changed, in response either to stimuli from the environment or to internal signals accompanying ontogeny.

In bacteria, rates of messenger RNA synthesis from specific genes were shown to vary directly in response to environmental signals (Leive and Kollin, 1967). In addition,

mutations affecting observable steady state RNA or protein levels are found in transcriptional regulatory signals or in effector molecules such as repressors (for examples, see (Miller and Reznikoff, 1978)).

Variations in RNA composition had early on been noted in eukaryotes as well as in prokaryotes. For example, cells at various stages of sea urchin development were shown to contain many similar, but also many different mRNA sequences (Galau et al., 1976). Likewise, differences in mRNA populations were found to exist between chick liver and oviduct (Axel et al., 1976), and between mouse liver, kidney and brain (Hastie and Bishop, 1976). To clarify that these differences in mRNA composition are mirrored in the pattern of transcription and are not due to mRNA processing events such as splicing, polyadenylation and transport, studies involving pulse labelling of nascent RNA chains in isolated nuclei were undertaken. Thus, sequences from liver-specific mRNAs were shown to be absent from nascent transcribed RNA in brain cell nuclei (Derman et al., 1981). Estrogen induction of vitellogenin in *Xenopus liver* cells (Brock and Shapiro, 1983a) and of ovalbumin transcription in hen oviduct cells (McKnight and Palmiter, 1979) also involve increased rates of transcription, although, significantly, the hormone has the incidental effect of increasing the half-life of vitellogenin mRNA (Brock and Shapiro, 1983b). By the late 1980's though, it had become firmly established that the primary mechanism of the control of gene expression in the vast majority of cases studied is in fact transcriptional regulation (Mitchell and Tjian, 1989). Testimony to the pre-eminence of transcriptional regulation in gene expression today are a vast variety of model systems in development (e.g. *Drosophila* embryogenesis), cell differentiation (e.g. yeast mating type, muscle, liver), growth control (e.g. Rb, p53, abl), inducible responses (e.g. HSF, NFkB or hormone-dependent gene expression), as well as the viral life cycle (e.g. E1A, IE, VP16, Zta, HIV Tat).

Approaches to study transcription

Classical genetic techniques have been used to identify many genes whose products govern important cellular and developmental processes in both prokaryotes and eukaryotes through the generation of phenotypic variants. However, some processes have failed to yield to genetic analysis either because they are intrinsically buffered such that several mutations must occur to create a sufficient perturbation to produce an observable phenotype (Gans et al., 1975) or in the other extreme the process might be so fundamental to the basic functioning of the organisim that mutations in its components cause a broad range of phenotypes (e.g. screens genes involved in cell cycle control in yeast) or are simply lethal. While the former possibility might be the case for important regulatory sequences (Bensimhon et al., 1983), the latter is most likely true with respect to the components that make up the fundamental transcription machinery. However, with the advent of recombinant DNA technology and yeast reverse genetics, it has become possible to design more subtle screens in eukaryotes. The most productive have usually involved the introduction of a mutant form of a known component for which a 'suppressor mutation' can be identified, and/or the use of a particular reporter (e.g. a particular promoter construction linked to a selectable marker, such as his, leu, β-lac, or neo, or colour indicator, such as lacZ) rather than having to rely on overall viability and aberrant phenotype of the mutant organism. Both techniques not only allow the screen to be focussed to identify components in a particular process or pathway but also increase the likelihood of identifying otherwise lethal mutations. Recent examples with respect to the general transcription factors are yeast TBP (spt15, (Eisenmann et al., 1989), TFIIB (sua7, (Pinto et al., 1992)), and BRF a likely component of $TFIIIB_2$ (Buratowski and Zhou, 1992; Colbert and Hahn, 1992; Lopez-de-Leon et al., 1992).

A second, arguably more successful approach to the study of transcriptional regulation that was made possible with recombinant DNA technology involves the transfection of plasmid DNAs into higher eukaryotic cells. Countless early studies using a single transfectable plasmid containing variants of the promoter under investigation linked to a reporter (such as CAT or lacZ) has led to the functional identification of an ever increasing number of promoter elements. Experiments with multiple transiently transfected as well as stably integrated plasmids also permit the functional dissection of effector molecules such as tissue-specific transcription factors.

Biochemical analysis has proven to be the most appropriate approach for studying the protein factors that carry out the basic processes of transcription initiation. Such a strategy is fundamentally reductionist, aiming at isolating the minimum number of cellular components that are responsible for a particular physiological process. One of the most important criteria for success with such an approach is the suitability of the biochemical *in vitro* assay. As our understanding of the factors and mechanisms of the biochemical process in question advances, the assay and the questions it may address grow in complexity. Biochemical studies on transcription beautifully illustrate this point: the *in vitro* assay developed for the purification of RNA polymerase is what we now call a non-specific transcription assay; an assay scoring accurate initiation requires a set of general transcription factors; and finally a transcriptional regulator-responsive assay requiring a variety of co-activators. A next step in pursuit of faithfully reconstituting the physiological process of transcription in the test tube will be to include other chromosomal components, as well as addressing questions concerning the efficiency of initiation, multiple rounds of initiation, termination and processing. As the molecular characterization of the transcription factor IID has in many

ways paralleled this development over the last decade (some might say has been central to it), I will briefly summarize the early biochemical foundations in the following section.

Recently, factors relevant to transcriptional regulation have been identified through a variety of ingenious genetic screens in the yeast *Saccharomyces cerevisae*; e.g. suppressors of the cold sensitive phenotype caused by CTD truncation (SRB, (Koleske et al., 1992)), of the VP16 squelching phenotype (ADA2, (Berger et al., 1992)), of the GAL4 activation domain deletion phenotype (SUG1, (Swaffield et al., 1992)) and of Ty element insertions (SPT3, SPT7, SPT8). Other such intermediary/regulatory proteins identified in yeast are GCN5 which is required for full activation by GCN4 for example (Georgakopoulos and Thireos, 1992) and the multi-protein complex containing SNF2/SWI2, SNF5, SNF6, SWI1 and SWI3 which has been implicated in both activation (Laurent et al., 1991; Peterson and Herskowitz, 1992) and chromatin structure (Hirschhorn et al., 1992).

However, the few emerging reports on the possible functional mechanism of these factors in transcription have relied entirely on a biochemical analysis (Koleske et al., 1992; Thompson et al., 1993; Koleske and Young, 1994). It is thus to be expected that the most productive strategy for scientific study of transcription (as well as many other research topics) is a fruitful marriage of complementary approaches and techniques, namely molecular (and often reverse) genetics and biochemistry. Despite the lag in analyzing transcriptional mechanisms in yeast due to biochemical difficulties in preparing transcription initiation-competent extracts, it is to be expected that the fundamental mechanisims surrounding the issue of transcriptional regulation will be understood in greatest detail through studies in the yeast *Saccaromyces cerevisiae*. Investigations aimed at the molecular pathways leading to differentiation events or developmental decisions may continue to be most fruitful and relevant when studied in

higher eukaryotes, such as *C. elegans*, *Drosophila* or mammalian model systems.

Transcription in *Escherichia coli*

Prokaryotic transcription, as studied primarily in *Escherichia coli*, can be thought of as a process in which the specificity and enzymatic activity of RNA polymerase is modulated by at least two distinct RNA polymerase binding proteins, sigma and NusA. While sigma is involved in the initiation of transcription, NusA functions during the elongation/ termination phase of transcription. As this thesis concerns itself exclusively with the initiation of eukaryotic transcription, I will only mention a few interesting features of sigma and its interplay with RNA polymerase.

Although sigma is often referred to as a subunit of the RNA polymerase holoenzyme, it can not only be purified away from the holoenzyme such that the core component of the RNA polymerase retains its enzymatic activity, but it is also present in substantial amounts in the unbound form (Burgess et al., 1969). Thus, sigma does not behave as a subunit in the strictest sense of the word, though it is required for promoter recognition by the polymerase.

Comparison of a variety of bacterial promoter sequences led to the definition of two consensus sequence elements positioned at -10 (Pribnow box) and -35 (Hogness box) relative to the start site of transcription. The -10 consensus sequence is TATAAA and the -35 consensus is TTGACA (Hawley and McClure, 1980), the two being separated by a 17bp spacer whose length is also conserved. Mutational analyses have corroborated this assignment of promoter elements and indicate that the degree of similarity to the consensus determines the relative promoter strength (Youderian et al., 1982; Szoke et al., 1987). However, the -35 element is less strictly required for high activity (Reznikoff et al., 1985).

Interestingly, the lack of such a consensus element in the promoter seems to be correlated with a dependency on positive activators for their function. Recent mutagenesis of σ^{70} resulted in the identification of two DNA-binding domains (Dombroski et al., 1992) shedding light on the earlier observation that contacts between holoenzyme and promoters of highly different sequence are quite homologous in space, orientation, and the extent of melting (Siebenlist et al., 1980).

Much of what is known about the function of sigma comes from comparing the properties of the holoenzyme with those of the core enzyme. Although these differences are generally ascribed to sigma function, they have not classically been directly demonstrated with the purified sigma factor. This was presumed to be due to significant conformational changes that sigma undergoes upon binding to the enzyme. Indeed, a recent mutagenesis study successfully demonstrated that the N-terminal region (region 1.1) of σ^{70} has an inhibitory effect on DNA binding by the C-terminal DNA-binding domains free sigma factor (Dombroski et al., 1993). Furthermore, this inhibition of binding by the N-terminal domain can be observed in *trans*, lending support to the proposal that initiation of transcription in prokaryotes is in part regulated by the modulation of protein conformation, and that the sigma family of proteins belong to a class of intramolecularly regulated transcriptional effectors.

Sigma function has long been known to be specific to initiation: once, initiation by RNA polymerase is completed, sigma is released to begin another cycle of transcription initiation (Travers and Burgess, 1969). UV cross-linking experiments established that sigma binds the promoter in the region of the -10 element (Simpson, 1979), leading to promoter melting and open complex formation.

Multiple sigma factors have been isolated and their respective genes cloned from *E. coli*, *B. subtilis* and several bacteriophages (reviewed in (Helmann and Chamberlin, 1988)).

The existence of multiple sigma factors constitutes a significant regulatory mechanism whereby gene expression is controlled during the sporulation of gram-positive bacteria (Losick and Pero, 1981) or in response to particular physiological stimuli, such as heat shock (reviewed in (Reznikoff et al., 1985; Arnosti and Chamberlin, 1989)). Comparison of the amino acid sequences of the various sigma factors has shown that, although their functional characteristics are extremely well conserved, their sequences are not (Stragier et al., 1985; Gribskov and Burgess, 1986; Lonetto et al., 1992).

Four weakly conserved regions have been identified on the basis of computer analyzed comparisons of multiple bacterial and phage sigma sequences (Gribskov and Burgess, 1986). Homology region 2 is the only one of the four which appears to be conserved in all of the sigma factors compared thus far. For this reason it has been postulated to contain the essential functions associated with sigma: binding to the core component of RNA polymerase, promoter recognition, and DNA template melting. Further subdivision of this domain into 4 subregions, 2.1-2.4 (Helmann and Chamberlin, 1988), has led to the identification of subregion 2.1 as the interaction site with RNA polymerase (Lesley and Burgess, 1989). Mutational studies have specifically implicated subregion 2.4 as the part of sigma which recognizes the -10 T/A-rich promoter element (Hu and Gross, 1988; Zuber et al., 1989). The only indication that subregion 2.3 may be involved in template melting is the similarity of this sequence to a presumptive nucleic acid binding motif present in RNA or ssDNA binding proteins.

Region 4 contains a predicted helix-turn-helix motif of the type that is associated with specific binding to DNA. Mutational studies indicate that it does play a role in recognition of the -35 element (Gardella et al., 1989; Siegele et al., 1989), and it is absent from a sigma factor (e.g. phage T4 gp55) that is functional on promoters that do

not appear to have a -35 element as part of their architecture (Elliot and Geiduschek, 1984; Gribskov and Burgess, 1986). The functional significance of region 3 has, however, remained obscure.

An interesting exception is that σ^{54}, the sigma factor required for expression of nitrogen fixation genes (Magasanik and Neidhart, 1987; Kustu et al., 1989). In this case the gene encoding σ^{54} bears no resemblance to the family of sigma factors compared thus far (Helmann and Chamberlin, 1988), but does contain glutamine-rich regions, resembing the activation region of the mammalian transcription factor Sp1 (Courey and Tjian, 1988). Holoenzyme containing σ^{54} can form a closed complex but cannot proceed further. The activator protein NifA and hydrolyzable ATP are required for the formation of an open complex and the initiation of transcription (Sasse-Dwight and Gralla, 1988; Popham et al., 1989). Thus the functioning of σ^{54} may seem most reminiscent of the initiation pathway of eukaryotic polymerase II and its accessory proteins.

(ii) The Biochemistry of Eukaryotic Transcription

Prelude to accurate eukaryotic transcription *in vitro*

Nuclear DNA-dependent RNA polymerases were first efficiently solubilized from sea urchin and rat liver (Roeder and Rutter, 1969), but shortly thereafter were obtained from a wide variety of sources. Three types of RNA polymerase were identified and resolved by conventional chromatography. The potential for great confusion was moderated by the discovery that the various polymerases exhibited different responses to the mushroom toxin α-amanitin. Specifically, RNA polymerase II (pol B in European nomenclature) was inhibited by α-amanitin at concentrations well below those required to inhibit RNA pol I and III (Chambon et al., 1970; Lindell et al., 1970). This observation was extremely significant as α-amanitin was also shown to inhibit mRNA biosynthesis *in vivo* (Zylber and Penman, 1971). Thus RNA polymerase II activity appeared to be responsible for mRNA transcription.

Unfortunately, initial attempts to demonstrate sequence specificity of *in vitro* transcription by purified RNA polymerase II were unsuccessful or at best ambiguous (reviewed in (Roeder, 1976)). It was therefore assumed that additional components were required to provide specificity for the polymerase, and a number of factors which stimulate RNA polymerase activity were identified and purified to varying degrees. In some cases the stimulatory effect could be ascribed to an increased rate of initiation, rather than elongation (Stein and Hausen, 1970; Sugden and Keller, 1973; Chuang and Chuang, 1975; Sekimizu et al., 1976; Spindler, 1979), but the specificity of the effects, in terms of increasing the rate of initiation at true promoter sites, could not be demonstrated. The fact that RNA polymerase II has difficulty initiating at internal sequences suggests that these factors may provide nicking- or unwinding-type activities which can improve the rate of internal initiation without providing sequence specificity. With one possible

exception (Sekimizu et al., 1982) the physiological significance of these factors for transcription *in vivo* is doubtful.

Even with more sophisticated assays in the late 70's, purified RNA polymerase II failed to show accurate initiation *in vitro*. Several instances of specific initiation at sites near authentic initiation sites have been noted in yeast (Lescure et al., 1981a; Lescure et al., 1981b; Lescure and Arcangioli, 1984), wheat germ and mammalian polymerases (Tsuda and Suzuki, 1982; Mishoe et al., 1984), though it is has always been difficult to rule out that initiation factors contaminate the polymerase and furthermore, the physiological relevance of these phenomena has remained unclear. In some of these reports, initiation ensued from positions with unusual structural features, either a potential site of secondary structure (Tsuda and Suzuki, 1982) or the actual TATA box itself (Lescure and Arcangioli, 1984; Mishoe et al., 1984). Theoretical and experimental results indicate that the TATA sequence embedded in a GC-rich region may undergo denaturation transitions under physiological conditions (Bensimhon et al., 1983; Patel et al., 1983) possibly leading to initiation *in vitro*. It is unlikely, but possible, that these specific yet inaccurate transcription events observed with pure polymerase have biological relevance.

The First Accurate *In Vitro* Transcription Systems

Specific, or more appropriately, accurate transcription could not truly be monitored until the advent of recombinant DNA technology. The reasons for this are twofold: first, sites of transcription initiation *in vivo* had to be accurately mapped (for an early compilation of viral and cellular promoters see (Breathnach and Chambon, 1981)). Second, appropriate DNA templates and/or probes had to be constructed. Though some DNA viruses have sufficiently small

genomes to be used as templates in transcription reactions, plasmids containing fragments with strong viral promoters such as the adenovirus major late promoter (pSmaF, i.e. the F fragment of a SmaI restricted adenovirus genome which contains the Ad2 MLP), or more recently the HIV-1 promoter, soon became favourite model systems in a number of configurations. Linked to a transcript not requiring the nucleotide guanidine ("G-less cassette"), such promoters provide a quick and convenient activity assay for chromatographic fractions (Sawadogo and Roeder, 1985a).

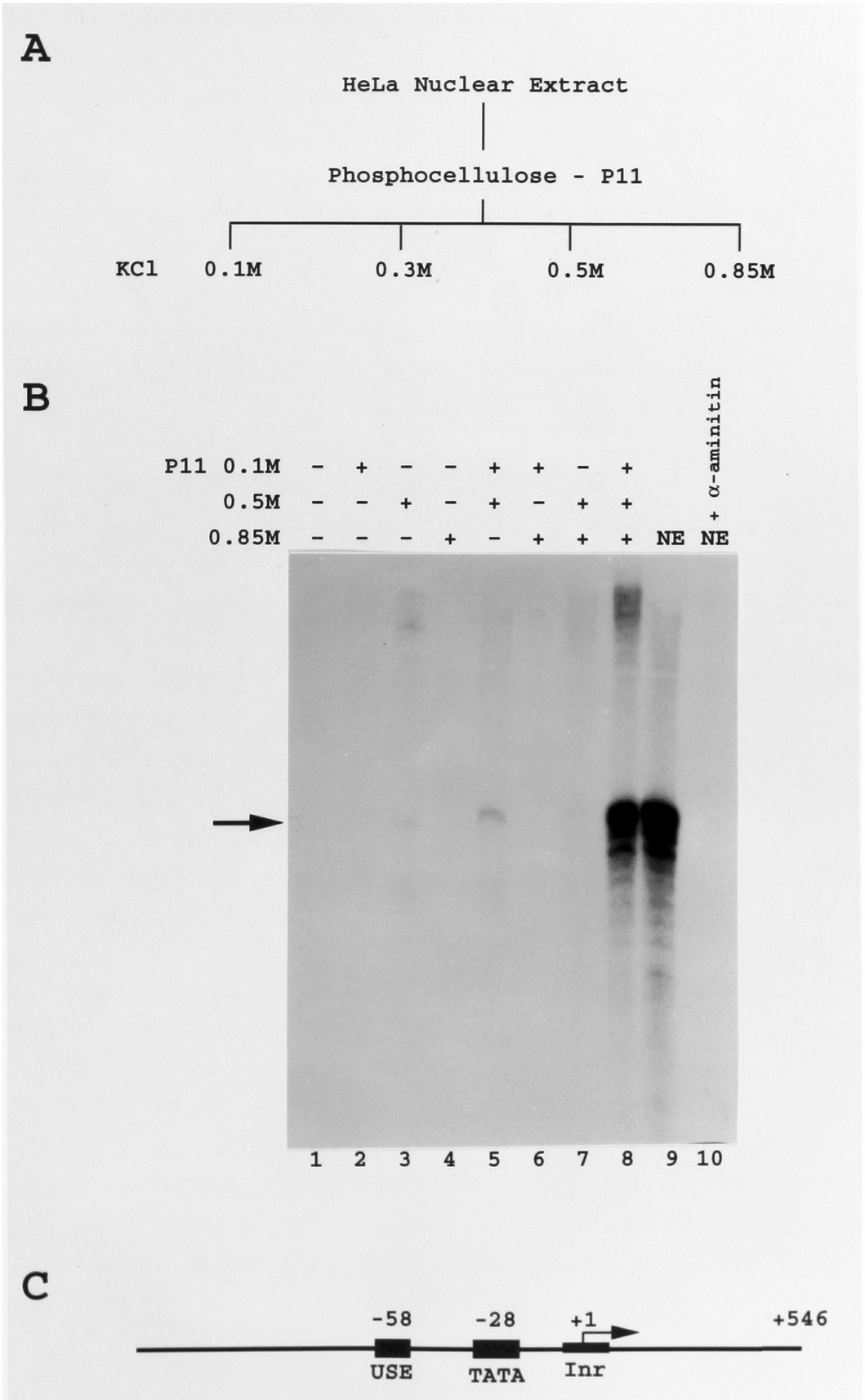
Accurate *in vitro* transcription by purified eukaryotic RNA polymerase preparations was first observed when these were added to isolated nuclei (Sklar and Roeder, 1977) or when purified RNA polymerase III was incubated with isolated chromatin templates (Parker et al., 1976; Parker and Roeder, 1977). Similar experiments with nuclei of adenovirus infected HeLa cells furthermore proved that *de novo* initiation had occurred by demonstrating the incorporation of a radioactive precursor at the second position of the transcript (Manley et al., 1979). These observations provided the first indication that the accessory factors required by eukaryotic polymerases for promoter recognition are template-associated, unlike what biochemical studies on prokaryotic transcription had revealed (Helmann and Chamberlin, 1988). Shortly thereafter, with the advent of cloned genes, accurate transcription from the adenovirus major late promoter was observed in soluble protein extracts supplemented with RNA polymerase II (Weil et al., 1979), and in soluble whole cell extracts containing endogenous polymerase (Manley et al., 1980). In both cases the sequence at the 5' terminus of the *in vitro* transcript was determined, and shown to be identical to that observed *in vivo* (Ziff and Evans, 1978). Given the requirement of accessory transcription components (Weil et al., 1979), in the absence of a suitable genetic approach, a fractionation/reconstitution scheme was necessary for their isolation, as exemplified by the biochemical analysis of

Figure 1. Fractionation of HeLa nuclear extract on phosphocellulose.

A. Nuclear extract from HeLa cells was loaded at 0.1M KCl onto a phosphocellulose P11 column. Bound proteins were eluted with buffers containing indicated KCl concentrations and peak protein fractions within each step elution were pooled.

B. *In vitro* transcription ("run-off) assay with step-eluted P11 fractions in indicated combinations incubated with a linearized template containing the SmaF fragment from adenovirus type 2 encoding the major late promoter (MLP). Accurate transcripts of 536 nucleotides are indicated by an arrow.

C. Schematic of the Ad2 MLP structure showing the TATA box, the initiator element and the upstream sequence element (USE, or "E-box").


A
HeLa Nuclear Extract
Phosphocellulose - P11
KCl
0.1M
0.3M
0.5M
0.85M
B
P11 0.1M
0.5M
0.85M
+ α-aminitin
NE
NE
1 2 3 4 5 6 7 8 9 10
C
-58
-28
+1
+546
USE
TATA
Inr

bacterial replication (Schekman et al., 1974). The validity of this strategy was confirmed by the first successful reconstitution, which demonstrated that several, and possibly many accessory factors are required to supplement RNA polymerase II *in vitro* (Matsui et al., 1980). An example of such a fractionation of nuclear extract on phosphocellulose and transcription assay with reconstituted components is presented in Figure 1. Considerable effort has subsequently been devoted in several laboratories to the purification and characterization of the general transcription factors.

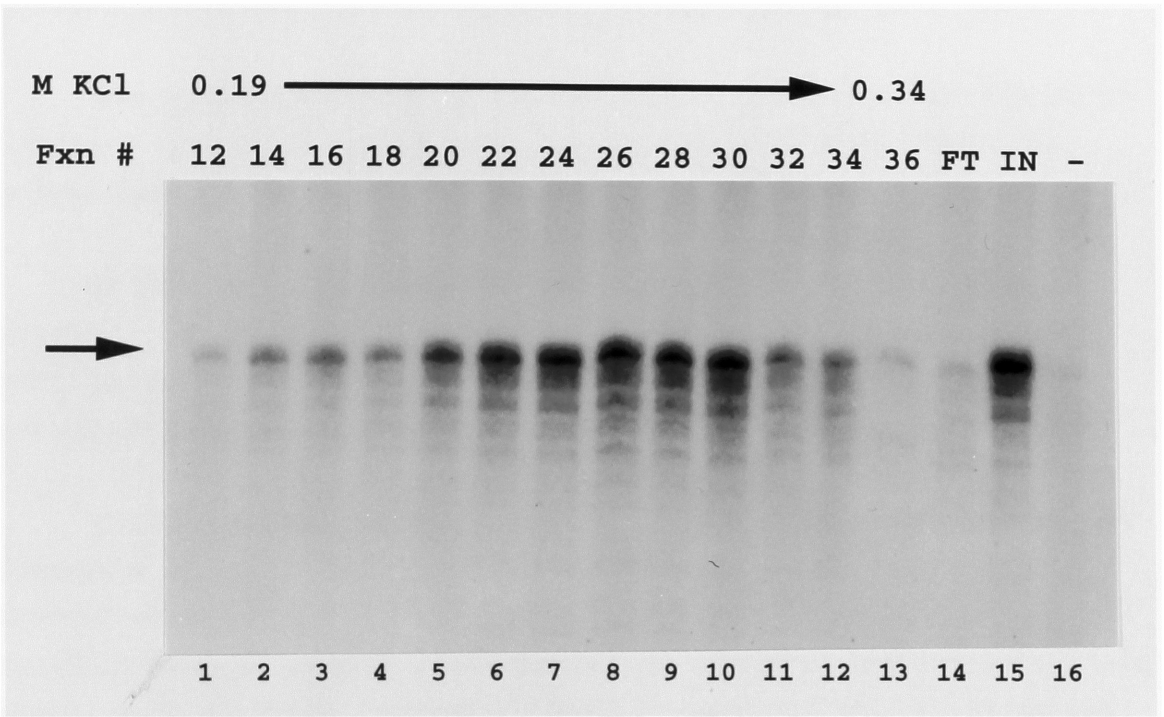
The *in vitro* transcription system in 1988

By 1988, I was introduced to a reconstituted system capable of accurately initiating class II transcription consisting of five partially purified factors in addition to polymerase:

TFIIA (Matsui et al., 1980; Reinberg et al., 1987),also referred to as AB (Samuels et al., 1982; Samuels et al., 1984) or STF (Davison et al., 1983; Egly et al., 1984) The absolute requirement for this factor remained unclear and seemed to depend on the purification stage of the complementing factors. Figure 2, for example, shows second column fractions reconstituting transcription by complementing with first (P11) column fractions; however, no requirement was seen using third column complementing fractions (data not shown, (Sawadogo and Roeder, 1985a)). One curious report suggested that its active component is actin (Egly et al., 1984) but this was inconsistent with data resulting from the use of other chromatographic schemes (Samuels and Sharp, 1986) and has since been disproved (cf. recent purification and cloning of TFIIA subunits (DeJong and Roeder, 1993; Ma et al., 1993; Yokomori et al., 1993a)).

Figure 2. Transcription assay with TFIIA-containing DEAE Sepahacel fractions.

Phosphocellulose P11 flow-through was loaded onto a DEAE Sephacel column. Fractions spanning a linear gradient salt elution as indicated, were incubated with phosphocellulose $P11_{0.5}$ and $P11_{0.85}$ and the pSmaF template. Accurate transcripts of 536 nucleotides are indicated with an arrow.


M KCl
0.19
0.34
Fxn #
12 14 16 18 20 22 24 26 28 30 32 34 36 FT IN -
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Several laboratories had purified transcription factors with molecular weights in the 27,000-35,000 range such as **TFIIB** (Reinberg and Roeder, 1987). Its functional homologue in rat, α, was the first class II general factor to be purified to homogeneity (from rat liver nuclear extracts) and yielded a polypeptide of 35kd (Conaway et al., 1987). BTF3, a fraction from HeLa whole cell extract (Zheng et al., 1987) seemed functionally similar to TFIIB but was subsequently shown to be encoded by a mitochondrial gene and not to be required in nuclear extract transcriptions (S. Malik and R.G. Roeder, personal communication; (Moncollin et al., 1992)).

The active component in the **TFIIE** fraction was found to interact in solution as well as on the DNA with both RNA polymerase II and TFIIB (Reinberg and Roeder, 1987).

TFIIF, an additional component separated from the original TFIIE activity (Flores et al., 1989) was shown to contain as one active component the RNA polymerase II-associated proteins RAP30 (Flores et al., 1988) and RAP74.

TFIID (Matsui et al., 1980), or otherwise designated DB (Samuels et al., 1982), BTF1 (Davison et al., 1983) and D (resolvable in τ and ε; Conaway and Conaway, 1990). Purification attempts from mammalian nuclear extracts had not shown much success beyond a third column fraction (Nakajima et al., 1988) but much work had been done to characterize this factor, which I will discuss in detail below.

In addition, attempts had been made to characterize the process of initiation in mechanistic terms. Early on an energy requirement was described for accurate initiation (Bunick et al., 1982a; Sawadogo and Roeder, 1984). The chemistry of RNA polymerization itself requires that an NTP be hydrolyzed at the α–β phosphate bond. Thus for non-specific initiation and elongation by RNA pol II, NTP analogs

containing a non-hydrolyzable β–γ bond are active and are incorporated normally into the transcript. Interestingly, these β–γ nonhydrolyzable analogs were shown not to support accurate transcription initiation *in vitro*, unless they are supplemented with either ATP or dATP (Bunick et al., 1982b; Sawadogo and Roeder, 1984; Conaway and Conaway, 1988). Which step, however, required this energy, remained unclear.

Other studies addressed the question of possible pathways in the ordered assembly of the general transcription factors into a pre-initiation complex. This functional approach utilized intricate order of addition protocols combined with template challenge, or sarkosyl and heparin (Hawley and Roeder, 1985; Hawley and Roeder, 1987) as assembly blocking or destabilizing agents. Footprinting and gelshift studies with native human TFIID and a corresponding yeast polypeptide (van Dyke et al., 1988; Buratowski et al., 1989) further visualized the ordered assembly pathway, by demonstrating discrete footprints/gel shift complexes with each additional factor incorporated. After binding of TFIIDto the promoter, alone or in conjunction with TFIIA, TFIIB may associate before polymerase can bind in conjunction with TFIIF and TFIIE. Because TFIIB gives rise to some DNase I protection between the TATA box and cap site, it is commonly believed that this molecule 'measures' the distance and determines polymerase's initiation site (van Dyke et al., 1988).

Physical *in vitro* analyses (Buratowski et al., 1989) not correlated with functional assays (e.g. template commitment, (Reinberg et al., 1987; van Dyke et al., 1988)) are more prone to experimental artefact than when coupled to functional assays, and need to be carefully interpreted. In two cases though, a different set of experimental tools (glycerol gradient and immunoprecipitation vs. fractionation) has cast doubt on the physiological relevance of chromatographically separable transcription factor activities. Complex formation as studied by gel shift assays

(Buratowski et al., 1989) seems to follow a particular pathway when these components are partially purified and thereby separated on chromatographic resins. In yeast however, a holoenzyme containing all general factors except TFIID and TFIIE can be isolated by immuno-purification and shown to initiate (partially) activator-responsive transcription when supplied with TBP and TFIIE (Koleske and Young, 1994). This is consistent with previous reports, employing mammalian fractions, demonstrating discrete complexes, in the absence of DNA, of RNA polymerase II with TFIIB and with TFIIF for example (Sopta et al., 1985; Reinberg and Roeder, 1987; Price et al., 1989; Flores et al., 1990; Kitajima et al., 1990; Gerard et al., 1991; Henry et al., 1992; Tschochner et al., 1992).

Similarly, in class III transcription, a multi-protein complex TFIIIC, required for initiation, splits on conventional chromatographic resins into two sub-components $TFIIIC_1$ and $TFIIIC_2$; however, immunoprecipitation experiments have indicated the existence of a single complex (Zhengxin Wang, Eric Sinn, personal communication), as well as an association between TFIIIC and TFIIIA (Eric Sinn, Beth Moorefield, personal communication).

In vitro reconstituted transcription systems in 1994

Several additional general transcription factors have been identified and purified to near homogeneity as is evident from the recently published purification chart reproduced in Figure 3 (Flores et al., 1992). Figure 4 gives a brief overview of the molecular make-up of each factor and a short description of its functional role in the initiation complex. A total of about 30 polypeptides (not counting the >10 constituents of RNA polymerase itself), amounting to well over 10^6 Daltons, make up the class II initiation complex. The complexity might seem surprizing but only reflects the cellular requirement for intricate transcriptional control

Figure 3. Identification and purification scheme of the general transcription factors.

First comprehensive purification scheme for all class II general transcription factors (except TFIID) as published by Flores et al. 1992.

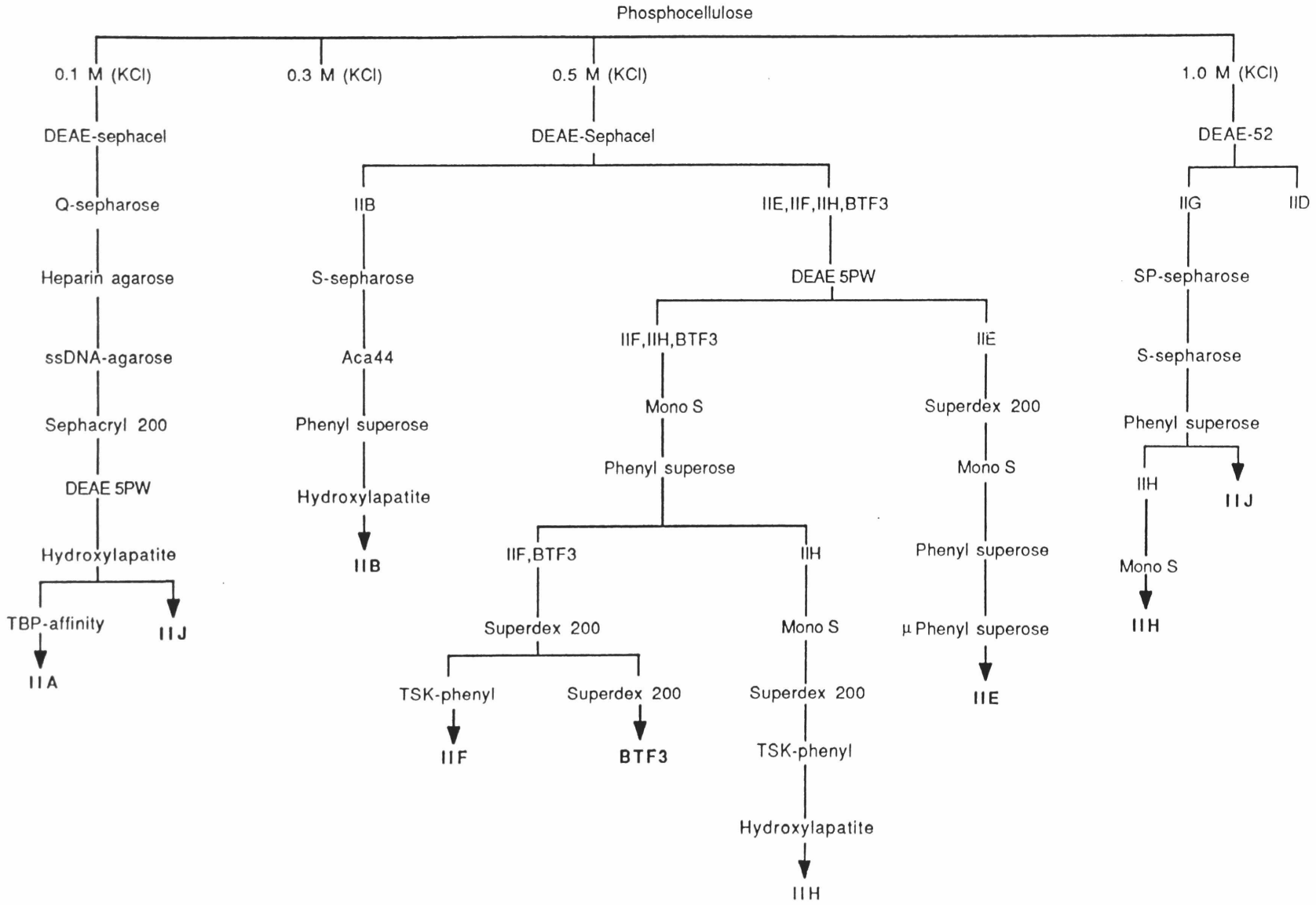

Phosphocellulose
0.1 M (KCl)
0.3 M (KCl)
0.5 M (KCl)
1.0 M (KCl)
DEAE-sephacel
Q-sepharose
Heparin agarose
ssDNA-agarose
Sephacryl 200
DEAE 5PW
Hydroxylapatite
TBP-affinity
IIJ
IIA
DEAE-Sephacel
IIB
S-sepharose
Aca44
Phenyl superose
Hydroxylapatite
IIB
IIE,IIF,IIH,BTF3
DEAE 5PW
IIF,IIH,BTF3
Mono S
Phenyl superose
IIF,BTF3
IIH
Superdex 200
Mono S
TSK-phenyl
Superdex 200
Superdex 200
TSK-phenyl
Hydroxylapatite
IIF
BTF3
IIH
IIE
Superdex 200
Mono S
Phenyl superose
μ Phenyl superose
IIE
DEAE-52
IIG
IID
SP-sepharose
S-sepharose
Phenyl superose
IIH
IIJ
Mono S
IIH

Figure 4. Summary of the class II general transcription factors.

Currently identified general transcription factors with indicated approximate molecular weight (if known) and potential functional role in accurate transcription.

General Class II Transcription Initiation Factors

Factor	Subunits (kD)	Function
TFIID	38 (IID_τ or TBP)	binds to TATA box initiates complex formation direct activator target
	others (TAFs)	required for activator function and initiation at TATA-less promoters
TFIIA	12, 19, 35	stimulates TFIID binding to TATA
TFII-I	120	binds to INR, promotes TFIID binding
TFIIB	35	promotes Pol II-TFIIF binding direct activator target
RNA Polymerase II	>10 (> 600 kD)	catalytic function
TFIIE	34, 57 (tetramer)	stimulates processivity of IIH kinase
TFIIF (RAP 30/74)	30, 74 (tetramer)	targets pol II to promoter stimulates pol II processivity
TFIIH	35, 41, 43, 62, 90	CTD kinase activity, helicase, repairosome
TFIIG/J	?	stimulates pol II processivity ?

during cell cycle, differentiation, and development. Furthermore, it has become apparent that not only the mechanistic principles but also the components of this apparatus are functionally and structurally highly conserved across the eukaryotic spectrum, an aspect that has benefited the progress in our understanding enormously. In the following I will not attempt to give a comprehensive review of past studies but rather to sketch out some of the pertinent questions that are being addressed today.

The recently purified multi-protein complex TFIIH has the exciting capability to phosphorylate the C-terminal domain (CTD) of the large subunit of RNA polymerase II and thus is the most likely candidate for the long-sought CTD kinase (Feaver et al., 1991; Lu et al., 1992; Serizawa et al., 1993). Given an interaction of unphosphorylated CTD with TBP (Usheva et al., 1992), as well as a reported helicase activity of TFIIH (Schaeffer et al., 1993), current models suggest that the addition of TFIIH to the pre-initiation complex allows RNA polymerase to initiate and/or clear the promoter region. The finding that TFIIE is capable of regulating the TFIIH kinase activity (Lu et al., 1992; Ohkuma and Roeder, 1994), suggests that recruitment of this heterotetramer might be regulated and in turn could be a target for activators in a novel mechanism of *trans*-activation.

Furthermore, TFIIH is also under intensive scrutiny in both human and yeast systems following the finding that some of its subunits or transiently associated proteins are coded for by previously identified genes involved in DNA repair. For example, the human p89 subunit corresponds to helicase ERCC3 (Schaeffer et al., 1993), and the yeast p85 and p50 subunits are the previously characterized RAD3 and SSL1 (Feaver et al., 1993). Recently published work demonstrating that transcription is coupled to repair mechanisms (Drapkin et al., 1994; Guzder et al., 1994) is consistent with previous observations indicating that actively transcribed

genes in mammalian cells (Bohr et al., 1985) and in *E. coli* (Mellon and Hanawalt, 1989) are repaired more rapidly than non-transcribed genes or regions in the genome.

The role of the weak initiator (INR) consensus (Smale and Baltimore, 1989) remains unclear. While early studies suggested that components of the TFIID factor interact with the initiator region (Nakajima et al., 1988), this model, while still correct, has failed to explain the sequence preference in this region. More recently a number of initiator binding proteins have been isolated, such as TFII-I (Roy et al., 1991) and YY1 (Seto et al., 1991), which do show sequence specificity, as well as the capacity to interact with previously described components of the general transcription machinery. Studies with TFII-I, in particular, led to the interesting suggestion that on a TATA-less promoter this factor might bind first to the DNA, nucleating the pre-initiation complex assembly by recruiting TFIID and obviating the need for TFIIA (Roy et al., 1993b), the factor which dramatically stimulates TFIID binding to DNA on TATA-containing promoters. Similar interactions are being analyzed for YY1 (Usheva and Shenk, 1994a), that may lend some support to the intriguing possibility of alternative complex assembly pathways in response to different activators: while TFII-I has been reported to interact with USF (Roy et al., 1991) and the oncoprotein Myc (Roy et al., 1993a), YY1 may be recruited by glutamine rich activators, like Sp1 (Seto et al., 1993).

The topic that commands the greatest interest is undoubtedly the mechanism of action of *trans*-activators. As will be detailed below, TFIID was implicated in 1988 in transducing the activation signal (Hai et al., 1988; Horikoshi et al., 1988a; Horikoshi et al., 1988b), and this concern has been a recurring theme throughout my thesis work. However, a number of recent studies have also suggested that certain activators play a role in the recruitment of TFIIB to the committed complex (Choy and Green, 1993). Direct

interactions between the VP16 activation domain and human TFIIB can be detected *in vitro* (Lin et al., 1991), and *in vivo* (Colgan et al., 1993), and point mutations in VP16 that abolish activation correlate with a loss of TFIIB binding (Lin et al., 1991; Lin and Green, 1991). Conversely, critical residues in TFIIB that crucially affect the interaction with acidic activators (Gal4-AH, Gal4-VP16) abolish activated transcription, but only slightly decrease basal levels (Roberts et al., 1993). Utilizing immobilized templates, TFIIB can be seen to be recruited to the promoter by an acidic activation domain (Choy and Green, 1993). However, this phenomenon does not translate into a functional effect; increased levels of transcription still require co-activator components, and thus the functional relevance of the TFIIB interaction remains to be delineated. One candidate for such co-activator might be the yeast protein ADA2 (Berger et al., 1992), identified by its ability to suppress the squelching phenotype of overexpressed VP16 in yeast (Berger et al., 1990). Interestingly, suppressors of ADA2 mutants have in turn been isolated in the SUA7 gene, the yeast TFIIB homolog (R. Knaus, L. Guarente, unpublished results), thus implicating TFIIB in the mechanism of activation.

Mechanistically however, there is still little conceptual framework at this point for how an activator can effect large increases in transcriptional activity other than by affecting recruitment of one or more of the basal factors through direct protein-protein interactions. Given that *trans*-activation effects can be observed with saturating levels of general factors, recruitment as currently interpreted cannot be the mechanism responsible for activator-dependent stimulation in such a case, and thus is clearly not the only physiologically relevant mechanism. The precise role of each component within the initiation complex still remains to be elucidated and assays for conformational changes or similar allosteric effects (as seen with sigma in

prokaryotic transcription) leading to initiation and promoter clearance will have to be developed. Furthermore, initiation factors alone will not give the full picture but soluble co-factor fractions (USA) as well as recently identified polymerase-associated co-activators in yeast such as SWI2/SNF2-containung complexes (Cairns et al., 1994) that are likely to have homologues in *Drosophila* (e.g. brahma (Kennison and Tamkun, 1988; Tamkun et al., 1992)) and human (e.g. PC2 (Kretzschmar et al., 1994)) suggest that mechanisms for transcriptional regulation have evolved to impinge on every conceivable aspect of transcription (for review see (Roeder, 1991)).

(iii) An Introduction to TFIID

Identification, chromatographic behaviour

Following the initial finding that the phosphocellulose D fraction contains an activity required for accurate transcription initiation *in vitro* (Matsui et al., 1980), much effort was devoted to the purification of this mammalian factor. Following cation and anion chromatography, little purification resulted from any of the conventional or affinity media (specific DNA oligo) employed due to spreading of the activity and its tendency to co-elute with the protein peak (M. Horikoshi, personal communication). Best results were obtained with ω -amino octyl agarose and in combination with a number of other steps a 1000-fold purification was obtained (about 1% purity; M. Horikoshi, personal communication). The goal of obtaining a purified functional TFIID preparation is also one of the objectives in parts of this thesis and will be elaborated on especially in chapter 2, sections (iii) and (iv).

A rapid and convenient complementation assay was developed (Nakajima et al., 1988) utilizing the observation that this activity can be selectively heat-inactivated (47°C for 15min). Possible heat-stable components of TFIID that are subsequently chromatographically separable from the heat-labile component would thus be lost, but have not been detected until very recently.

Native molecular mass estimates for this factor have varied widely from an estimate of 130kd for the HeLa derived factor (Samuels et al., 1982; Reinberg et al., 1987) to about 750kd for the homologous rat liver-derived activity (designated τ) found to be a highly elongated macromolecule with an axial ratio >10 (Conaway and Conaway, 1990; Conaway et al., 1991).

Figure 5. Characteristics of the TFII activity in the phosphocellulose (P11) D fraction.

Functional characterization of TFIID was largely accomplished by 1988. (See text for further details.)

Characteristics of the TFII activity in the phosphocellulose D fraction

1. functionally required for all class II promoters
2. acts via the TATA element, hence 'TATA-factor'
3. template commitment factor
4. remains associated with the promoter for several rounds
5. large footprint on a strong promoter (MLP -47 to +35)
6. 3' extension of footprint on a weak promoter (E4) is correlated with presence of activator
7. requires thermal energy for DNA binding to form a slowly migrating gel shift complex
8. best estimates indicate a native mass of about 750kd

Promoter interactions

The so-called TATA-motif, which in higher eukaryotes is generally located around positions -25 to -30 relative to the initiation site, was identified (Gannon et al., 1979)) when eukaryotic promoter sequences were compared in the late 70's in a search for possible eukaryotic counterparts to the bacterial control sequences such as the "Pribnow" box at -10 or the "recognition" region at -35 (Goldberg, 1979, Ph.D. thesis Stanford University). Subsequently, results from *in vitro* studies (Corden et al., 1980; Grosveld et al., 1981; Hu and Manley, 1981; Mathis and Chambon, 1981; Sassoni-Corsi et al., 1981; Tsujimoto et al., 1981; Wasylyk and Chambon, 1981; Grosschedl and Birnstiel, 1982; Zarucki-Sculz et al., 1982; Concino et al., 1984) and *in vivo* (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Grosveld et al., 1982; McKnight and Kingsbury, 1982; Dierks et al., 1983; Charnay et al., 1985) indicated that this is a critical element both for promoter activity and for determining the exact point of initiation. Moreover, for some promoters the TATA box was shown to be the only DNA sequence element that is absolutely required for a low level of initiation *in vitro* (Corden et al., 1980; Grosveld et al., 1981; Hu and Manley, 1981; Sassoni-Corsi et al., 1981; Tsujimoto et al., 1981) and for mediating transcription responses to viral activator proteins *in vivo* (Green et al., 1983; Coen et al., 1986; Wu et al., 1987; Simon et al., 1988) and *in vitro* (Abmayr et al., 1985; Abmayr et al., 1988). The first identification of the factor that interacts with this element came from *Drosophila* embryo extracts (Parker and Topol, 1984b) which exhibit particularly high transcription activity. A preliminary study with HeLa derived TFIID fraction showed that this factor interacts specifically - and in the absence of the other factors - with the TATA region (Sawadogo and Roeder, 1985b). This was consistent with an earlier report that the formation of stable preinitiation complexes in crude fractions is dependent on TATA-containing sequences (Davison

et al., 1983). Thus, since the TATA element had been identified as the major component of the core promoter, a key role for TFIID in initiating the assembly of a functional preinitiation complex was indicated.

Promoter interactions by TFIID were characterized in more detail using a 300 fold purified mammalian preparation which distinguished two general patterns of TFIID interactions by partial digestion with DNase I and exonuclease III (Nakajima et al., 1988). In the first, exemplified by the Ad2 major late promoter (MLP), interactions were observed over a broad region extending from positions -47 to +35 (Nakajima et al., 1988). Interestingly, comparisons between wild-type and mutant ML promoters (Sawadogo and Roeder, 1985b; Nakajima et al., 1988), and histone H4 promoters (M. Horikoshi, personel communication) indicated that downstream interactions are sequence independent. Furthermore, they could be selectively eliminated by competition with nonspecific DNAs (M. Horikoshi, M. van Dyke, personel communication). That the TATA element is indeed the key recognition (tight binding) site for TFIID on the ML promoter had been demonstrated in experiments using methidiumpropyl-EDTA•Fe(II) (MPE) cleavage of the template (Sawadogo and Roeder, 1985b). Similar experiments subsequently showed that downstream protection from this reagent was dependent on the presence of the other general factors (van Dyke et al., 1988). However, a later report demonstrated that at least on one promoter (gfa) interactions with a weak TATA consensus sequence by natural TFIID are dependent on the presence of downstream sequences (+10 to +40) indicating their potential importance in weak TATA containing or TATA-less promoters (Nakatani et al., 1990).

In contrast, on other promoters, exemplified by the hsp-70 promoter, the TFIID fraction exhibited a second mode of binding, giving rise to a much more restricted DNase I footprint of 16 bp over the TATA box region. Competition

studies (H4 vs. H2B) seemed to indicate that the the same factor is responsible for both types of interactions (M. Horikoshi, unpublished data) These differences were seen to correlate with promoter strength.

As had first been seen with *trans*-activating sequence specific binding proteins, the TFIID-DNA complex was also shown to withstand native gel electrophoresis (M. Horikoshi, unpubished data). However, this phenomenon required particular reagents and conditions, including at least a 100-fold purified TFIID preparations, a relatively large promoter fragment, special gel and buffer conditions (see Materials and Methods) and importantly a 30°C incubation for the formation of the complex prior to electrophoresis. Subsequent analysis confirmed that the gel shifted promoter DNA exhibited the same DNAse I protection pattern as observed in previous studies (M. Horikoshi and H. Fujii, unpublished data).

Role in template commitment

Soon after initial fractionation of HeLa nuclear extract, it was demonstrated using a template competition assay that the activity residing in the D fraction (DB) acts at a very early step in the reaction (Fire et al., 1984; Reinberg et al., 1987; Reinberg and Roeder, 1987) and in combination with TFIIA (AB) makes the DNA-protein complex refractory to challenge with competing templates (Fire et al., 1984; Reinberg et al., 1987). Furthermore, while re-initiation occurs rapidly when TFIID stays bound on the promoter, the other general factors are able to cycle in a manner similar to polymerase (Hawley and Roeder, 1985). An alternative characterization of functional steps with the detergent sarkosyl led to similar conclusions. This experimental approach was based on the observation that pre-initiation complex formation, (re-) initiation, and elongation each inhibited by increasing concentrations of sarkosyl (Hawley

and Roeder, 1985; Hawley and Roeder, 1987). Such findings were substantiated by complementary physical analyses with more highly purified preparations of TFIID using footprinting, and template exclusion assays (van Dyke et al., 1988) or gel shift assays (Buratowski et al., 1989) that demonstrated that TFIID bound to the TATA region is sufficient for template commitment.

What 'template commitment' may translate into physiologically was shown in elegant work combining *in vitro* transcription with an *in vitro* chromatin assembly system. Chromatin templates cannot be transcribed with a reconstituted system unless TFIID has been allowed to pre-bind and therefore commit the template to be active (Workman and Roeder, 1987; Workman et al., 1988). Thus TFIID may provide us with a molecular and functional link between DNA packaging into chromatin and the control of RNA synthesis from a DNA template.

Implication in activation

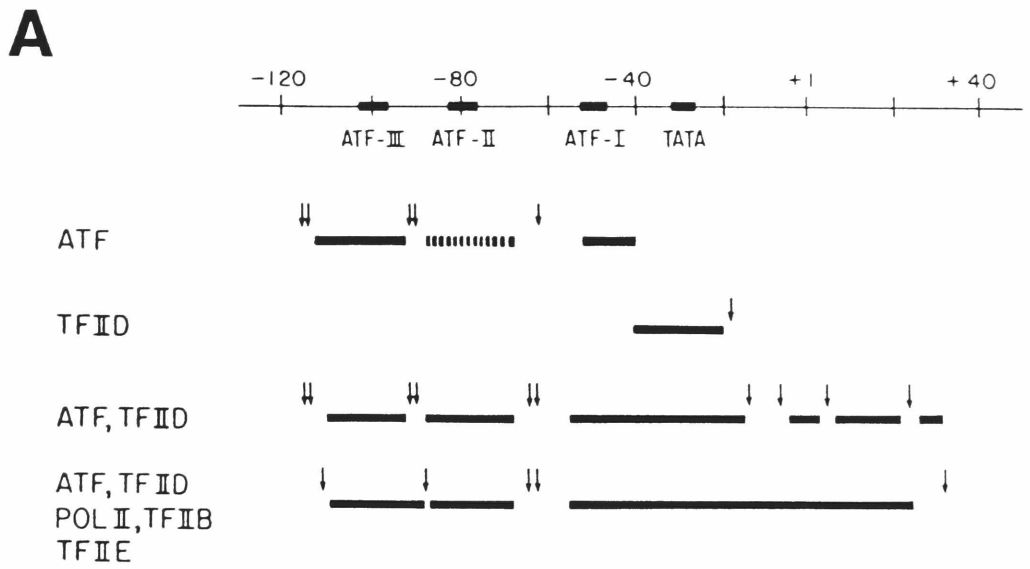
Initial attempts to reproduce *in vitro* the regulation of RNA polymerase II promoter function observed *in vivo* were only in part successful : while repression effects with SV40 large T (Rio et al., 1980; Hansen et al., 1981), Ad2 E2 protein (Handa et al., 1983), poliovirus (Crawford et al., 1981), and vesicular stomatitis virus (McGowan et al., 1982) were observed early on, experiments with E1A for example did not show the expected *trans*-activation effect. Soon, however, enhanced transcription of specific genes was seen in response to viral immediate-early proteins (Abmayr et al., 1985), to factors present in the S phase of the cell cycle (Heintz and Roeder, 1984), and a heat shock-induced factor (Parker and Topol, 1984a). In 1986 the first report demonstrating the *in vitro* function of a genetic element that modulates developmental regulation described a B-lymphocyte-derived, cell-free system that mediated both accurate and a high level

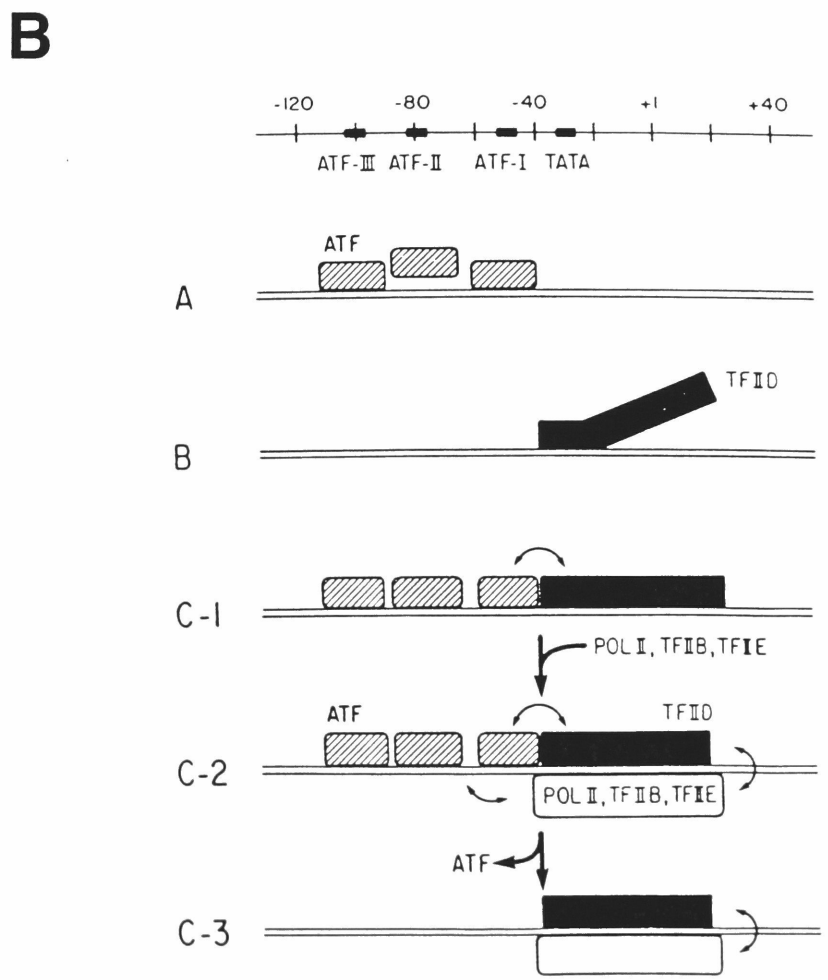
of B-cell-specific transcription of a κlight chain gene (Mizushima-Sugano and Roeder, 1986). This phenomenon was dependent upon DNA sequences previously implicated in tissue-specific promoter function *in vivo* (Bergman et al., 1984).

In this study the proximity of the upstream regulatory sequence element to the TATA box was noted. Another study also presented evidence for a co-operative interaction between a regulatory transcription factor and TFIID on the Ad2 ML promoter, in the sense that the off-rate of the specifically bound activator USF was markedly slowed when TFIID was bound as well (Sawadogo and Roeder, 1985b). It was only in 1988, however, that TFIID was more directly implicated in mediating the *trans*-activation effect. On one hand, mechanistic studies involving TFIID preincubation protocols showed that the pseudorabies immediate early (IE) protein activates transcription by facilitating TFIID promoter interactions (Abmayr et al., 1988); pre-bound TFIID would not allow the *trans*-activation effect. On the other hand, footprinting studies with the cellular activator ATF (Hai et al., 1988; Horikoshi et al., 1988b), as well as a GAL4 derivative (Horikoshi et al., 1988a) showed a correlation between activation and extension of a normally restricted TFIID footprint on the E4 promoter : when multiply bound these activators effected similar qualitative changes in TFIID-promoter interactions, which further correlated with more efficient binding of the other general factors (Fig. 6); when only the proximal binding site was occupied, similar footprint changes were observed but with little increase in preinitiation complex formation (M. Horikoshi, personal communication), suggesting that either TFIID might not be the only target for activators or that conformational changes not detectable by footprinting in TFIID are required for increased pre-initiation complex formation.

Figure 6. Model for TFIID's role in promoter activation by ATF.

A. Summary of the Transcription factor interactions on the E4 promoter as revealed by DNase 1 footprinting and presented in Horikoshi et al., 1988b. The DNase 1-protected regions (thick horizontal lines) and hypersensitive sites (thin vertical arrows) are indicated below the template diagram.
B. Model proposed for promoter activation by ATF by Horikoshi et al., 1988b. Upstream bound ATF causes a conformational change in TFIID that facilitates the assembly of the general transcription factors and RNA polymerase, which in turn form a stable pre-initiation complex even when upstream bound ATFs dislocate.


A
-120
-80
-40
+1
+40
ATF-III
ATF-II
ATF-I
TATA
ATF
TFIID
ATF, TFIID
ATF, TFIID
POL II, TFIIB
TFIIE


B
-120
-80
-40
+1
+40
ATF-III
ATF-II
ATF-I
TATA
ATF
A
TFIID
B
C-1
POL II, TFIIB, TFIIE
ATF
TFIID
C-2
POL II, TFIIB, TFIIE
ATF
C-3

Purification of yeast TFIID

At about the same time when TFIID was implicated in mediating the *trans*-activation effect and thus attracted increasing interest from a wider audience, the research field opened up dramatically when the major hurdle to further progress - the difficulty in purifying native TFIID - was removed. Just as human cdc2 was originally identified by *in vivo* complementation of a mutant yeast (Lee and Nurse, 1987), an experiment was attempted to achieve the reverse *in vitro* for TFIID. Two groups reported that a chromatographic fraction derived from a *Saccharomyces cerevisiae* whole cell extract was capable of functioning in a basal transcription assay in place of TFIID when complemented with partially purified human factors (Buratowski et al., 1988; Cavallini et al., 1988). Furthermore, in template challenge (Buratowski et al., 1988) and oligo competition (Cavallini et al., 1988) experiments this yeast factor behaved analogously to human TFIID.

In contrast to the native human factor, though, the activity resided in a protein of the surprizingly small native mass of 23-27,000 MW (Buratowski et al., 1988; Horikoshi et al., 1989b) and gave a DNase I footprint restricted to the TATA element on the ML promoter (Buratowski et al., 1988; Horikoshi et al., 1989b). However, the more favourable chromatographic behaviour allowed this activity to be purified to homogeneity in a series of four conventional chromatographic steps (Buratowski et al., 1988; Cavallini et al., 1989; Horikoshi et al., 1989b; Schmidt et al., 1989), which was followed by peptide sequencing and gene cloning.

These significant contributions were made in the context of accumulating evidence for the remarkable evolutionary conservation of class II transcription in eukaryotes (see (Guarente, 1988) for review). The largest subunit of RNA polymerase II had been known for some time to be highly conserved over eukaryotic evolution (Allison et al., 1985; Corden et al., 1985). Sequence and functional similarity

between mammalian and yeast transcriptional activators was also demonstrated with yeast activator GCN4, chicken oncogene jun (Struhl, 1987; Vogt et al., 1987), and its mammalian counterpart AP-1 (Bohmann and al., 1987), as well as CCAAT box-binding proteins (Olesen et al., 1987; Chodosh et al., 1988a; Hahn and Guarente, 1988), which were also shown to function as yeast/mammalian hybrid proteins (Chodosh et al., 1988b). These findings lay the conceptual foundations for the first cross-species gene fusion experiments, following the observation that the yeast protein GAL4 stimulates transcription from a promoter containing the appropriate binding site in mammalian cells (Kakidani and Ptashne, 1988; Webster et al., 1988) as the fos oncogene does in yeast cells (Lech et al., 1988).

Recently, the extent of molecular homology of general transcription factors was analyzed in detail, by exchanging *Saccharomyces cerevisae* factors with those purified from *Schizosaccaromyces pombe* (Li et al., 1994), an organism whose genes exhibit a promoter architecture (distance between TATA and initiation site) more similar to those of higher eukaryotes. All general transcription factors, except for TFIIB, TFIIE, and RNA polymerase, could be swapped individually between the two reconstituted systems. The exceptions were, however, functional when swapped in the pairs TFIIE-TFIIH and TFIIB-RNA polymerase II, and this latter pair was sufficient to shift the transcription start site according to its species derivation. It may be concluded that the need for a partner factor in the swap is indicative of functional interactions between them.

Cloning of yeast TFIID

Purification of yeast TFIID as a 27kd polypeptide allowed for its rapid cloning in a number of laboratories following peptide sequencing by PCR-mediated techniques (Hahn et al., 1989a; Horikoshi et al., 1989a) and/or direct screening

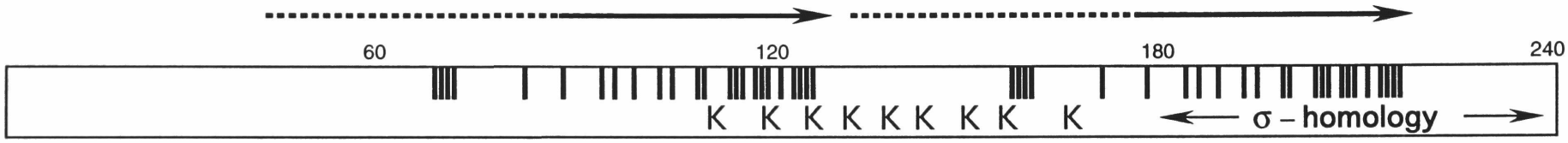
methods (Cavallini et al., 1989; Schmidt et al., 1989). Sequencing of positive clones revealed in each case a 720bp open reading frame encoding a protein of 240 amino acids and calculated molecular mass of 26986 Daltons. Thus TFIID became the first general transcription factor to be cloned and to be available in purified recombinant form (except for RAP30 subunit of TFIIF). Functional analysis of the recombinant protein expressed in bacteria (Schmidt et al., 1989), rabbit reticulocyte lysates (Hahn et al., 1989b; Horikoshi et al., 1989a), and Hela cells (Cavallini et al., 1989) confirmed its expected activity in TATA box binding and basal transcription assays when complemented with HeLa derived fractions containing the other basal transcription factors.

Interestingly, physical mapping of clones, hybridization, and sequencing established the identity of this gene to the previously identified SPT15 allele (Eisenmann et al., 1989). Though mutations in SPT15 had been isolated as suppressors of δinsertion mutations that alter the transcription of adjacent genes, their effects are highly pleiotropic, including defects in sporulation, mating and viability. Other complementation groups obtained from this particular screen with similar phenotype were identified as loss of function mutations in histone genes (Eisenmann et al., 1989), suggesting a possible physiological link between TFIID binding and nucleosome formation around the transcription initiation sites (cf. competition between nucleosome assembly and TFIID binding *in vitro* (Workman and Roeder, 1987)).

Detailed sequence analysis of the protein revealed a number of primary structure features (Fig.6): the most prominent is a directly repeating unit that accounts for almost two thirds of the protein. The 67 amino acid long segmens have 25 identical residues (30%) and if conservative amino-acid replacements are considered, the sequence

Figure 7. Schematic of the sequence motifs found in *Saccharomyces cerevisiae* TFIID.

The protein sequence is presented as a box with amino acid positions for indicating the scale. The approximate position and extend of the direct repeat is marked by horizontal arrows; amino acid residues identical in each repeat are indicated by a short vertical line. Lysines within a predicted amphipathic helix (basic repeat) are shown with the letter "K". The location and extend of a putative weak homology to the 2.3 and 2.4 regions of bacterial sigma factors is indicated. See text for further details and references.


60
120
180
240
K K K K K K K K K
σ – homology

similarity of these two domains is 60% (Cavallini et al., 1989; Hoeijmakers, 1990). The two repeats are separated by a putative amphipathic helix that displays basic residues, especially lysines, on one surface (Horikoshi et al., 1989a). The N-terminal region is highly hydrophilic (65%) with a large number of charged residues (Hoeijmakers, 1990). Finally, at the C-terminal end a very weak similarity to the 2.4 region of bacterial sigma factors can be found (Horikoshi et al., 1989a). This is of interest as there is genetic evidence for direct contacts between the σ-factor 2.4 region and the -10 promoter region which is AT-rich (Helmann and Chamberlin, 1988).

The availability of recombinant yeast TFIID protein allowed a great variety of questions surrounding transcriptional regulation in general and or focussing on TFIID in particular to be addressed by a much larger number of laboratories, probing its structure and function mapping interactions or using it as a starting point for genetic or biochemical screens. This thesis is not the place for a comprehensive review of these studies, but I will mention and refer to them when relevant to my own studies, while attempting to preserve, as much as possible, the overall chronology of the ongoing characterization of TFIID.

Chapter 1 : cDNA cloning of TFIIDs

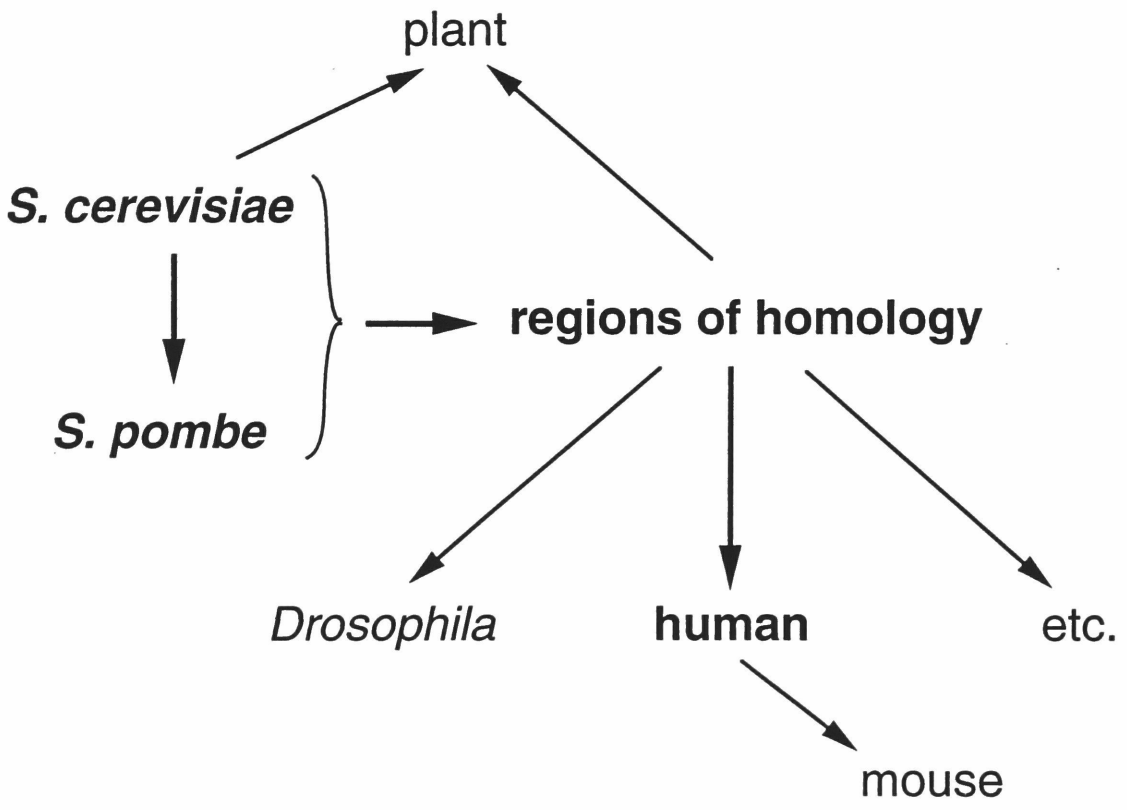
Introduction

Despite the ability of the yeast 27kd protein to substitute for the human TFIID fraction in all of the important assays tried at the time, it had become apparent that the two preparations were different in two significant characteristics: native size, as shown by gel filtration and gel shift analyses, and footprinting pattern on the Ad2 major late promoter, as I described in the previous section. Native human TFIID preparations had been shown to be a target for some regulatory factors, as evidenced both by physical (Sawadogo and Roeder, 1985b; Horikoshi et al., 1988a; Horikoshi et al., 1988b) and functional (Workman and Roeder, 1987; Abmayr et al., 1988; Workman et al., 1988) studies of interactions between these components. A detailed *in vitro* analysis of thus inferred molecular mechanisms underlying gene control necessitated the development of a homologous system.

The apparent evolutionary functional conservation of TFIID between the yeast and human factors suggested a similarly conserved primary structure in at least portions of its amino acid sequence. Initial attempts to obtain cross-hybridizing signals with the TFIID gene from *Saccharomyces cerevisiae* or pieces thereof in nucleic acid pools from human sources by either genomic Southern blotting (data not shown), cDNA library screening (S. Buratowski, personal communication), or PCR mediated cloning (data not shown) remained unsuccessful. An alternative strategy depicted in Figure 8 involves the cloning of a TFIID gene from another species of small genomic complexity, facilitating the use of low stringency hybridization techniques. The fission yeast *Schizosaccaromyces pombe* with an estimated genome size of 10^7bp (cf. 3.3×10^9bp for human genome) not only fullfills this criterion but is evolutionarily quite separate from the budding yeast, *Saccharomyces cerevisiae*. Fundamental

Figure 8. Strategy for cloning TFIIDs from higher eukaryotes.

Direct cross-hybridization with *Saccharomyces cerevisiae* TFIID may be used to identify the TFIID gene from *Schizosaccharomyces pombe*. Regions of homology between these two TFIID genes can be localized and used to screen for TFIID genes in higher eukaryotes, including human, by hybridization methods or PCR-aided cloning.


plant
S. cerevisiae
S. pombe
regions of homology
Drosophila
human
etc.
mouse

molecular and functional differences between the two yeasts that include mitosis and cytokinesis (Hiraoka et al., 1984), the control of cell cycle and growth (Russell and Nurse, 1986) (highlighted especially by the cloning of human cdc2 (Lee and Nurse, 1987)), as well as the molecular mechanism of RNA splicing (Käufer et al., 1985; Padgett et al., 1986) show a greater similarity of *S.pombe* than *S. cerevisiae* to higher eukaryotes. Finally, initiation of transcription also appears to be somewhat different : the position of the TATA box relative to the transcription start site varies in S. cerevisiae from -40 to -120bp, whereas it is rather more constant in the fission yeast (-25 to -40bp), conforming more to the characteristic TATA-box position (-25bp) in higher eukaryotes (Losson et al., 1985; Russell, 1985). Furthermore, genomic DNA sequence comparisons indicate that the two yeasts have diverged from each other as much as they have from humans (Russell and Nurse, 1986).

The cloning of a TFIID gene from two such divergent species may provide evidence for a structural homology that lies at the basis of the observed functional conservation, as well as lead to the mapping of particularly homologous regions within the primary structure of the protein. Such information might be used not only as a 'stepping stone' in the effort to cloning homologous cDNAs from higher eukaryotic species, but can also provide the basis for, or can complement an *in vitro* mutagenesis approach to provide a detailed structure-function analysis of the protein.

(i) *Schizosaccharomyces pombe* TFIID

The bulk of the work presented in this section was previously published in Hoffmann et al. (1990).

Isolation of genomic and cDNA clones

Although Southern blotting with genomic DNA from *Schizosaccaromyces pombe* probed with radio-labelled TFIID gene fragments at high stringency had not yielded any cross-hybridizing signals (data not shown; (Eisenmann et al., 1989)), similar experiments at lower stringency demonstrated a single cross-hybridizing band in genomic DNA digested with a number of different restriction enzymes (Fig. 9A, lanes 2, 4, 6, 7). This suggested that a single gene in *S. pombe* has a high level of structural homology to the *S. cerevisae* IID gene, which is shown to hybridize to its own genomic *Eco*R I fragment at high stringency (Fig. 9B). A 1.37kb *Hin*d III fragment that reproducibly cross-hybridized with the *S. cerevisae* probe at optimized conditions and apparently contained no recognition sites for restriction enzymes *Xho* I, *Sph* I and *Eco*R I was gel-isolated following a preparative *Hin*D III digest and ligated to form a subgenomic plasmid library. Minicultures pooled from the resulting colonies (12 in one) were similarly screened by Southern blotting for the presence of the cross-hybrizing fragment following the additional selection based on the lack of *Xho* I, *Sph* I and *Eco*R I restriction sites (Fig. 10A). A single culture proved to be positive (Fig. 10B, lane 11) and the corresponding colonies were subsequently analyzed individually in the same way yielding a single positive clone. Its insert was sequenced and found to contain sequence that showed a high degree of similarity at the amino acid level in one reading frame to that of *S. cerevisiae* TFIID. However, stretches of similarity were interupted by intervening sequences that in two out of three cases cause a frameshift in the reading frame and in all three cases contain stop codons in all three

Figure 9. Detection of *Schizosaccharomyces pombe* TFIID gene on a genomic Southern blot.

A. *Schizosaccharomyces pombe* genomic DNA probed with *Saccharomyces cerevisiae* TFIID gene (1-672 of the open reading frame) under low stringency hybridization conditions after digestion with *Xho*I, *Sph*I, *Eco*RI, and *Hin*dIII alone (lane 2, 4, 6,7) show a single cross-hybridizing fragment in each lane. When cut in combination with *Hin*dIII (1, 3, 5) the genomic HindIII fragment is shown to contain no internal XhoI, SphI, or EcoRI sites, a fact which alows for an additional level of selection in further cloning steps.
B. *S. cerevisiae* genomic DNA digested with *Eco*RI and probed with *S. cerevisae* TFIID DNA under high stringency conditions.

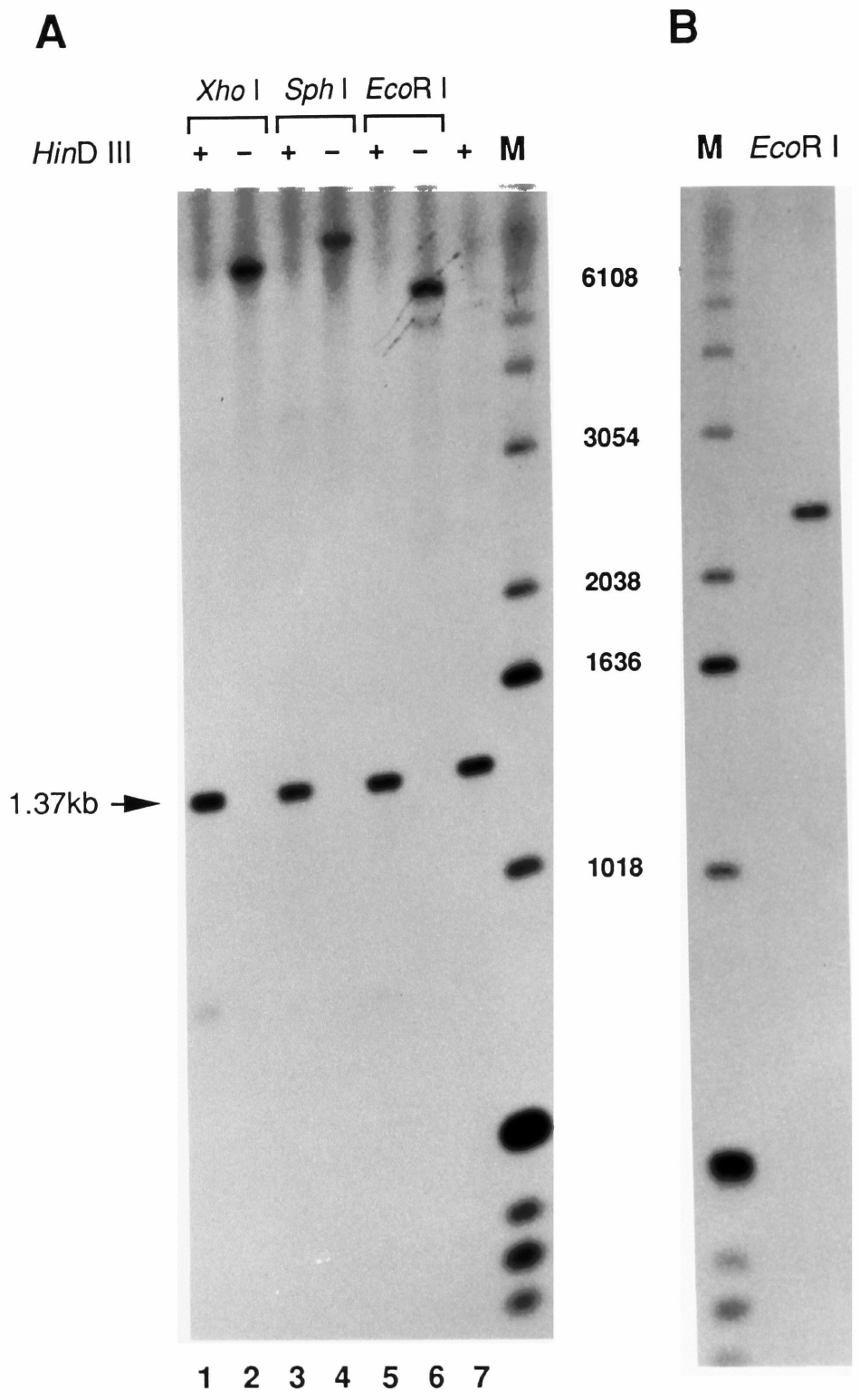

A
B
Xho I
Sph I
EcoR I
HinD III
+ − + − + − + M
M EcoR I
6108
3054
2038
1636
1018
1.37kb
1 2 3 4 5 6 7

Figure 10. Cloning TFIID-encoding genomic fragments from *Schizosaccharomyces pombe* : Miniprep analysis.

A. A plasmid library consisting of *S. pombe* genomic HindIII fragments of 1.35 to 1.4 kb in length was used to transform supercompetent bacteria and resulting colonies were pooled into 26 cultures. Resulting MIniprep DNA was cut with *Hin*dIII, *Xho*I, *Sph*I and *Eco*RI and size fractionated on an agarose gel along with size markers (lanes 1, 2) and genomic DNA from *S. cerevisiae* (cut with *Eco*RI, lane 3) and *S. pombe* (cut with *Hin*dIII, lane 4).

B. Agarose gel from Fig. 10A was transferred to a nitrocellulose membrane and probed with *S. cerevisae* TFIID DNA at low stringency hybridization conditions (as in Fig. 9A). A cross-hybridizing *S. pombe* genomic fragment of 1.37kb (lanes 4 and 11) is indicated by an arrow.

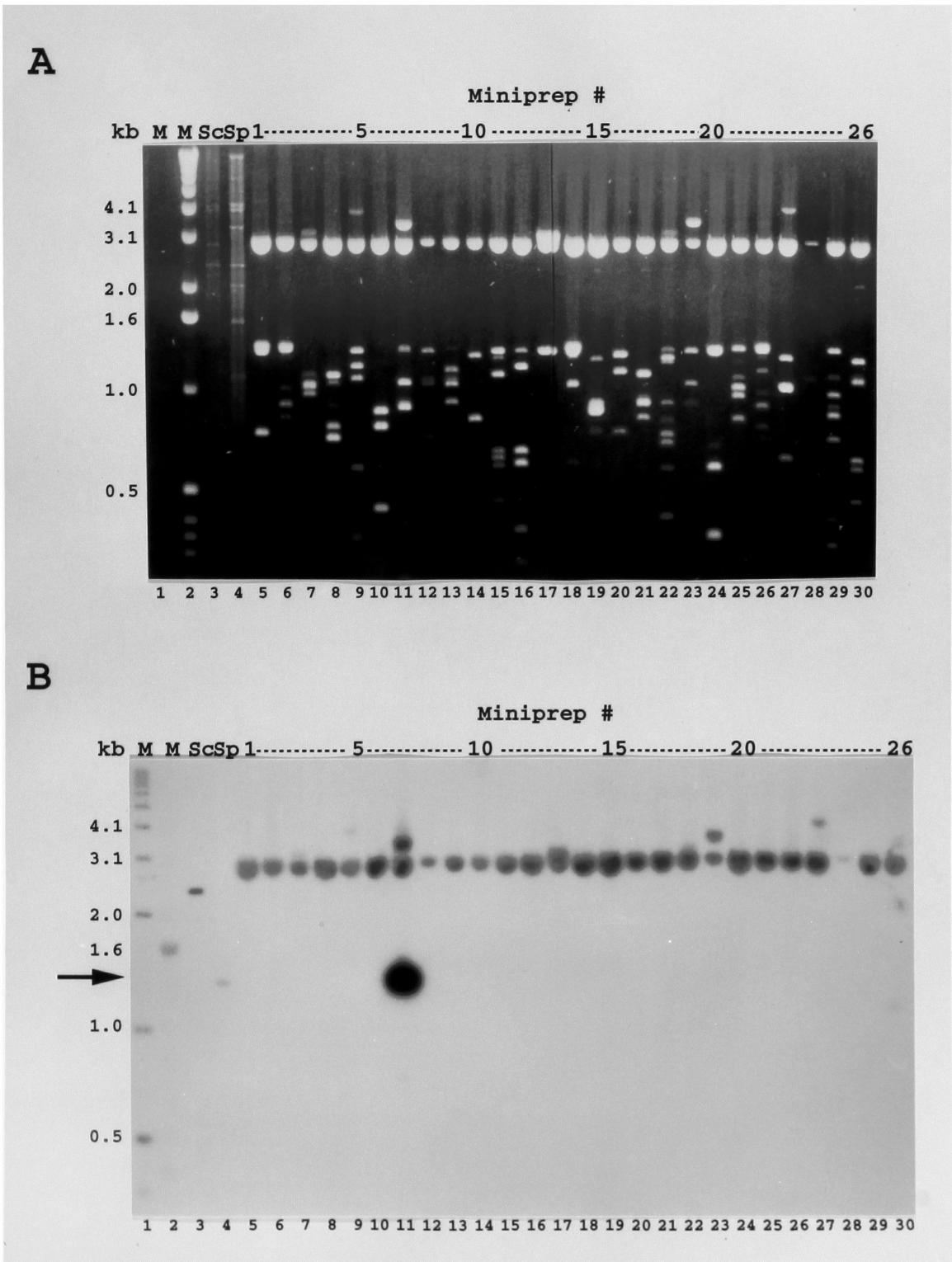
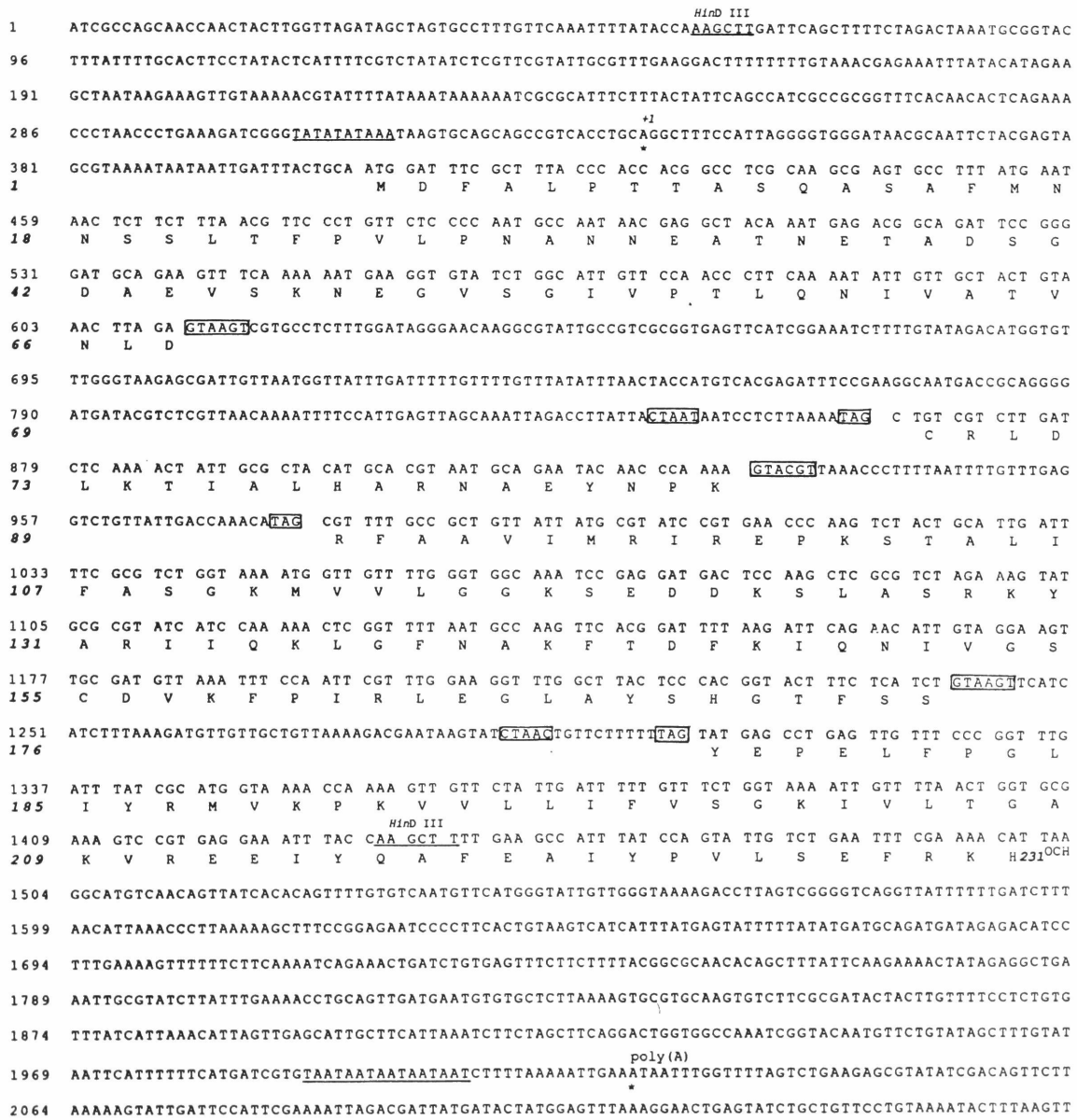

A
Miniprep #
kb M M ScSp 1 5 10 15 20 26
4.1
3.1
2.0
1.6
1.0
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
B
Miniprep #
kb M M ScSp 1 5 10 15 20 26
4.1
3.1
2.0
1.6
1.0
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Figure 11. Nucleotide and predicted amino acid sequences of the *Schizosaccharomyces pombe* TFIID gene.

The exon/intron boundaries are based on sequencing of genomic and cDNA clones and are indicated by boxes around splice consensus sequences. The *Hin*dII sites for the genomic fragment identified in Fig. 1A are shown. Characterization of the mature message of this gene has allowed the mapping of the transcription start site (+1), a putative polyadenylation signal (underlined), and the location of the poly(A) tail. The predicted translation of the TFIID open reading frame is shown.


HinD III
+1
HinD III
poly(A)

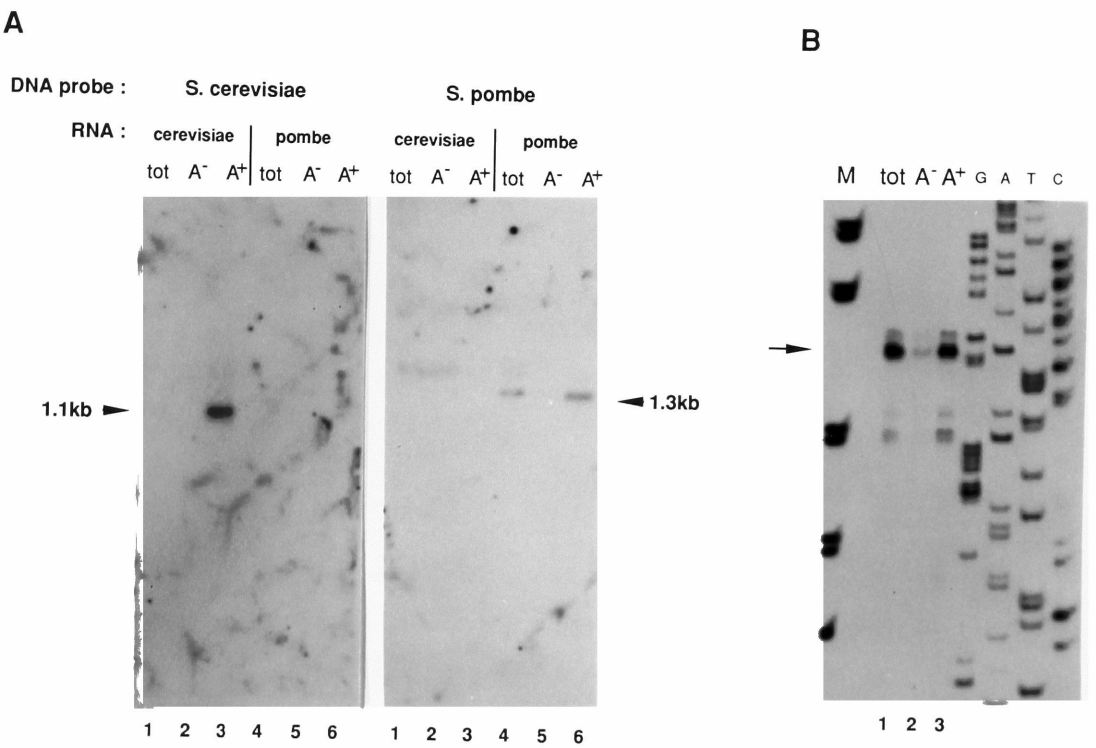
frames. The cloned genomic 1.37 kb *Hin*D III fragment clearly did not contain the complete C-terminal portion of the putative gene, however, it could be used to screen a genomic *S. pombe* library under high stringency conditions to yield several clones falling into two categories.

The two fully independent clones isolated were restriction mapped and partially sequenced. Both contained the previously identified *Hin*D III fragment with an additional ~10kb of 5′ and 3′ flanking sequences. On the basis of amino acid homology to *S. cerevisiae* TFIID the complete gene was identified, including three putative introns (Fig. 11, 255bp, 52bp, and 70bp in length), which contain the consensus splice donor and acceptor sites (boxed in Fig. 11, GTANGT and TAG, (Käufer et al., 1985)). The larger two putative introns also contain splice recognition boxes conforming to the consensus previously established for *S. pombe* (C T A/G A T/C in the 3′ part of the intron, (Langford et al., 1984; Käufer et al., 1985)), whereas the third may be too short to require such a recognition sequence, but is spliced out by a different mechanism. In order to verify the putative exon/intron structure and obtain a functional cDNA an oligo dT-primed cDNA library was constructed from *S. pombe* polyA$^+$ selected RNA and exhaustively screened. PCR was performed also on total RNA following a primer extension reaction using several specific oligonucleotides to increase the specificity of the products. The resulting clones were sequenced and shown to contain the previously predicted open reading frame, which is preceded by several stop codons.

Northern Blot analysis (Fig. 12A) was used to compare the messenger RNAs of TFIID genes from *S. pombe* and *S. cerevisiae*. At high stringency conditions the probes failed to crosshybridize but in each case revealed a discrete message of 1.3kb to 1.4kb for S. pombe and 1.1kb for *S. cerevisiae* (as previously shown, (Horikoshi et al., 1989a)).

Figure 12. Mapping of *Schizosaccharomyces pombe* TFIID mRNA.
A. Northern blots of 20μg of total RNA, poly(A)$^-$ and poly(A)$^+$ RNA from *S. pombe* and *S. cerevisae* probed under high stringency conditions with both *S. pombe* and *S. cerevisae* TFIID fragments. The positions of the 1.3kb (*S. pombe*) and 1.1kb (*S. cerevisiae*) hybridizing species were determined from size markers (not shown).
B. Primer extension analysis of total, poly(A)$^-$, and poly(A)$^+$ RNA from *S. pombe*. Primer extension products were rum along size markers and a DNA sequence ladder generated from the cloned TFIID gene with the same primer used for RNA analysis.

(These experiments were carried out in collaboration with the laboratory of P.A. Weil.)


A
DNA probe :
S. cerevisiae
S. pombe
RNA :
cerevisiae
pombe
cerevisiae
pombe
tot A- A+ tot A- A+ tot A- A+ tot A- A+
1.1kb
1.3kb
1 2 3 4 5 6 1 2 3 4 5 6
B
M tot A- A+ G A T C
1 2 3

The 5' transcription initiation site of the *S. pombe* IID message was mapped using primer extension (Fig. 12B) ~25bp from a putative strong TATA box found in the genomic sequence (Fig. 11) yielding a ~68bp 5' untranslated leader sequence. A putative polyadenylation signal similar to the consensus for higher eukaryotes (AATAAA, (Proudfoot and Brownlee, 1976)) was found 15bp from the poly(A) tail of the message (Fig.11) yielding a 529bp long 3' untranslated trailer. With the coding sequence being 696bp (231aa+ stop) the size of the expected mature message is 1294bp exclusive of the poly(A) tail, in good agreement with the estimate of 1.3kb to 1.4kb for the natural mRNA (Fig. 12A). In none of the experiments (exhaustive genomic and cDNA cloning, PCR on genomic DNA and total RNA, Northern, Primer extension, and a low-stringency genomic Southern analysis (with *S. cerevisiae* and *S. pombe* IID specific probes, data not shown) has there been any evidence for highly homologous multiple genes or alternative RNA splicing of TFIID in *S. pombe*.

Functional analysis

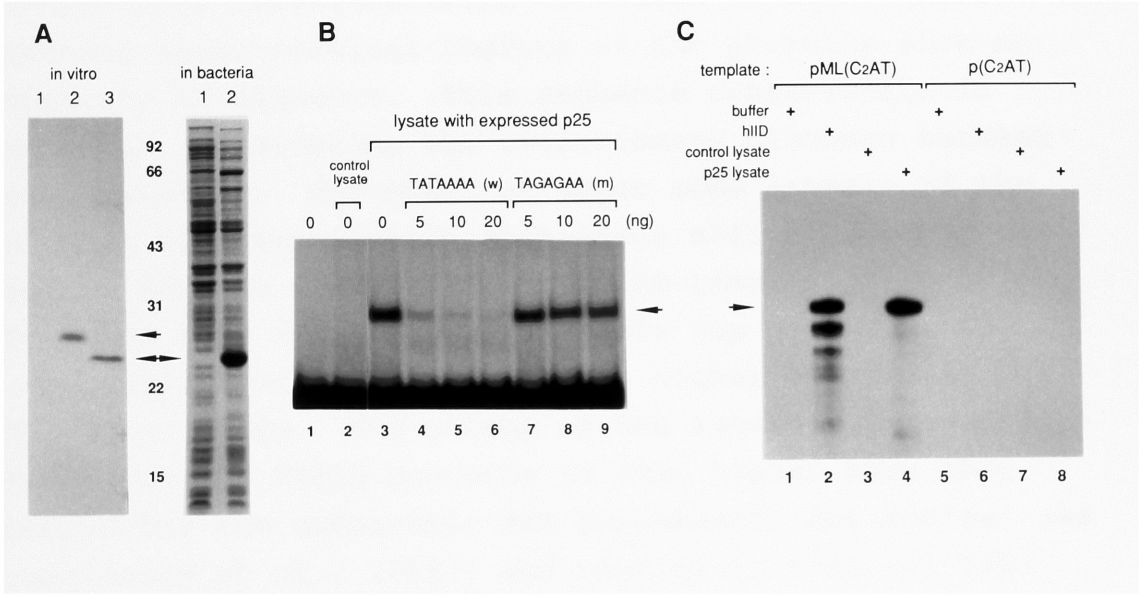
To demonstrate that the open reading frame described above encodes a protein with TFIID activity, a corresponding cDNA was cloned into pGEM-7Zf(-) and T7 polymerase-transcribed products were translated in rabbit reticulocyte lysates. SDS-polyacrylamide gel analysis (Fig. 13A, left) revealed a single band with an apparent size of 25kd, in good agreement with the calculated molecular weight of 25,427 Dalton. The *S. pombe* RNA-encoded protein also bound specifically to the TATA box in the adenovirus major late promoter, as shown by competition with wild-type versus mutant TATA box-containing oligonucleotides (Fig. 13B). As expected, the mobility of the resulting complex was slightly faster than that observed with the 27kd *S. cerevisiae* protein (data not shown). The *S. pombe* cDNA-encoded protein (p25) was also expressed in *Escherichia coli* (Fig. 13A, right) and showed promoter

Figure 13. Expression and functional analysis of cloned *Schizosaccharomyces pombe* TFIID.

A. SDS-PAGE analysis of (^{35}S)methionine-labelled TFIID expressed in rabbit reticulocyte lysates and visualized by autoradiography (*left*) or Coomassie Blue-stained TFIID expressed in bacteria (*right*). (*Left*) Reticulocyte lysates were programmed with no RNA (lane 1) or with *in vitro*-transcribed TFIID mRNA from *S. cerevisae* (lane 2) or *S. pombe* (lane 3) cDNAs. (*Right*) *E. coli* lysates were prepared 3 hr postinduction of cells containing the pET3a plasmid alone (lane 1) or with the *S. pombe* TFIID cDNA insert (lane 2).

B. Specific binding of *S. pombe* TFIID to the TATA box. Gel mobility shift assays with a 184bp adenovirus major late promoter fragment used in no protein (lane 1) and reticulocyte lysates programmed with no RNA (lane 2) or with *S. pombe* TFIID RNA (lanes 3-9). Competing oligonucleotides (containing base pairs -45 to -15 of the adenovirus major late promoter) contained either wild type or mutant TATA sequence (as indicated) and were added at the levels (0-20ng) indicated.

C. Transcription reaction with *S. pombe* TFIID. Transcription assays containing partially purified TFIIB, IIE/F, and RNA polymerase II were complemented with buffer (lanes 1 and 5), with partially purified human TFIID (lanes 2 and 6), with lysate from *E. coli* containing the pET3a plasmid eith no insert (lanes 3 and 7), and with lysate from *E. coli* containing the pET3a plasmid with the *S. pombe* TFIID cDNA insert (lanes 4 and 8). The pML(C_2AT) template in lanes 1-4 contained the adenovirus major late promoter (sequences -400 to +10) attached to the G-less cassette construct, whereas the p(C_2AT) template in lanes 5-8 contained only the G-less cassette (Sawadogo and Roeder 1985a). Templates containing only promoter sequences from -50 to +10 attached to the G-less cassette behaved exactly like the pLM(C_2AT) templates (data not shown). The arrow indicates the transcript resulting from specific initiation at the +1 site.


A
in vitro
1 2 3
in bacteria
1 2
92
66
43
31
22
15
B
lysate with expressed p25
control lysate
TATAAAA (w)
TAGAGAA (m)
0 0 0 5 10 20 5 10 20 (ng)
1 2 3 4 5 6 7 8 9
C
template : pML(C2AT) p(C2AT)
buffer
hIID
control lysate
p25 lysate
1 2 3 4 5 6 7 8

dependent transcriptional activity when assayed in a TFIID deficient complementation system from human cells (Fig. 13C). This ability to substitute for the human TFIID provides the final proof that the cloned *S. pombe* gene encodes a functional protein.

The conserved TFIID core domain

Sequence comparisons reveal an extraordinary degree of homology (93% sequence identity) between the 180-residue carboxy-terminal region of *Schizosaccharomyces pombe* TFIID (residues 52-231) and the corresponding region of *Saccharomyces cerevisae* TFIID (residues 61-240), whereas the remaining amino-terminal regions of the proteins show no similarity in sequence. This sequence conservation is remarkable, considering the evolutionary distance between these organisms. Moreover, at least some aspects of the mechanism of transcription initiation differ, because the distance between the TATA box and the transcription initiation site is large and variable (up to 120bp) in *S. cerevisiae* but closer to that in higher eukaryotes (20-40bp) in *S. pombe*. The extent of the sequence conservation between the two TFIID proteins is even higher than that reported for the eukaryotic RNA poymerases (for review, see (Cornelisson et al., 1988)) and similar to that (91-92%) reported for histones H3 and H4 (Matsumoto and Yanagida, 1985).

This sequence conservation suggests a highly conserved TFIID 'core' structure that must satisfy a number of essential functions on the promoter. These include both site-specific binding to the TATA element by apparently unique types of interactions (Nakajima et al., 1988) and the subsequent recruitment (by direct or indirect interactions) of RNA polymerase II and other general initiation factors into the a functional preinitiation complex (van Dyke et al., 1988; Buratowski et al., 1989). The extent of sequence

conservation between *S. pombe* TFIID and *S. cerevisae* TFIID is also notably consistent with recent mutagenesis studies (Horikoshi et al., 1990), which show that amino-terminal residues 1-62 of *S. cerevisae* TFIID are completely dispensible for basal level transcription, whereas the remainder of the molecule is absolutely essential. Even more striking than the overall conservation of amino acid sequence in carboxy-terminal region is the near perfect conservation of the residues comprising previously described structural motifs, including the lysine repeat in the central basis core, the flanking direct repeats and the sigma homology (Fig. 14 and 15). Altogether, these studies argue strongly for a TFIID core that contains potentially overlapping domains that are essential for these general functions and are highly conserved in evolution. Below, we consider the possible roles of these domains in light of these studies. Because TFIID can also be a target for various regulatory factors (for review, see (Horikoshi et al., 1989a; Horikoshi et al., 1990), considerations of these interactions are also relevant.

The direct repeats and the σ homology Interrupted direct repeats, indicative of an ancient duplication event, have been noted in *S. cerevisae* TFIID (Cavallini et al., 1989; Hoeijmakers, 1990; Nagai, 1990; Stucka and Feldman, 1990). Mutational studies (Horikoshi et al., 1990) have implicated both of these repeats in DNA binding and have led to the suggestion that they could give the TFIID structure a twofold symmetry that might be important for overall binding specificity and strength (note the inherent symmetry of the TATA element consensus T A T A T/A A T/A). The fact that only 2 of the 25 amino acids in the more strictly defined repeat (shaded residues in Fig. 14) differ between *S. pombe* and *S. cerevisiae* is indicative of the strong evolutionary pressure to preserve this structure. However, the significance of the larger, more broadly defined direct

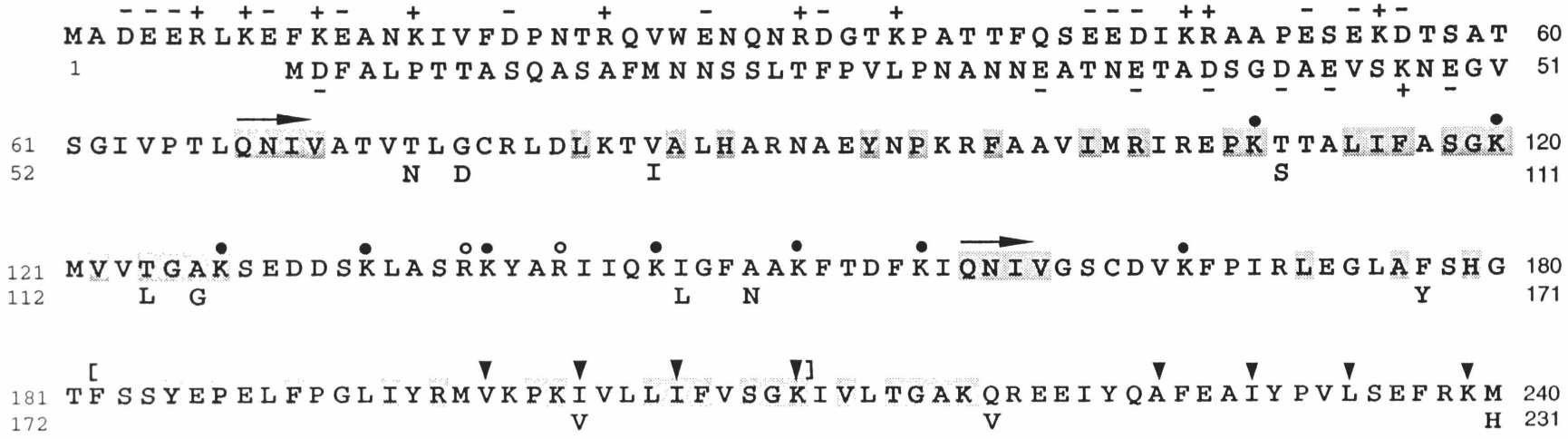
repeats in *S. cerevisae* (Cavallini et al., 1989; Hoeijmakers, 1990) is questionable, because part of the first repeat falls into the TFIID amino-terminal region that is not conserved between the two yeasts. We note that many of the invariant residues within the smaller repeats are hydrophobic and therefore could be involved in stabilizing interactions between the two domains.

We have noted previously weak putative homologies between the carboxy-terminal regions of *S. cerevisae* TFIID (residues 180-240) and a region of bacterial σ-factor involved directly in recognition of the -10 promoter element (TATAAT consensus). Consistent with the possibiblity that these regions are functionally related, the present analysis shows that a near absolute conservation of the relevant residues (the exception being a conservative change at position 202 in *S. pombe*), as well as a mutational analysis (Yamamoto et al., 1992) has also implicated specific residues of the σ homology in TFIID binding. Because the σ homology region lies partially within the second direct repeat and is not conserved in the first, the regions of TFIID containing the repeats may not be functionally equivalent.

Basic repeat region As noted previously, *S. cerevisiae* TFIID contains an interesting repeat of basic residues, especially lysines, in the central region of TFIID (Horikoshi et al., 1989a). This sequence appears capable of forming an α -helical structure, and deletion mutants in this region are non-functional (Horikoshi et al., 1990). Strikingly, all of the charged amino acids (mainly lysines) within this 37-residue region are identical between *S. cerevisae* and *S. pombe* TFIID, indicating that a specific structure is important for function. More recent studies with point mutants (Yamamoto et al., 1992) have shown that most of the lysine residues between the direct repeats, presumably on one face of an α -helix (Horikoshi et al., 1989a; Horikoshi et al., 1990), are not important either for DNA binding or for

Figure 14. Sequence comparison of TFIID genes from *Saccharomyces cerevisae* and *Schizosaccharomyces pombe*.

The sequence of *S. cerevisiae* TFIID (Horikoshi et al. 1989a) is shown in the upper line, whereas only the *S pombe* TFIID residues differing between the two yeasts are shown in the second line. The + and - symbols denote the basic and acidic residues in the highly divergent amino termini. The arrows indicate the beginning of each of the two interrupted direct repeats (Cavallini et al. 1989), with the individual residues comprising these repeats shaded. Lysines (solid circle) and arginines (open circle) indicate residues in the central basic core (Horikoshi et al. 1989a); this core has the potemtial for forming an α-helix with the positively charged residues (especially the lysine repeat) on one face. The σ -homology region (Horikoshi et al. 1989a) is comprised of a region (denoted by brackets) with a weak overall sequence identity to the σ 2.3 and 2.4 regions and two small regions containing residues (indicated by solid inverted triangles) that are the most highly conserved in the 2.4 regions of various sigma factors.



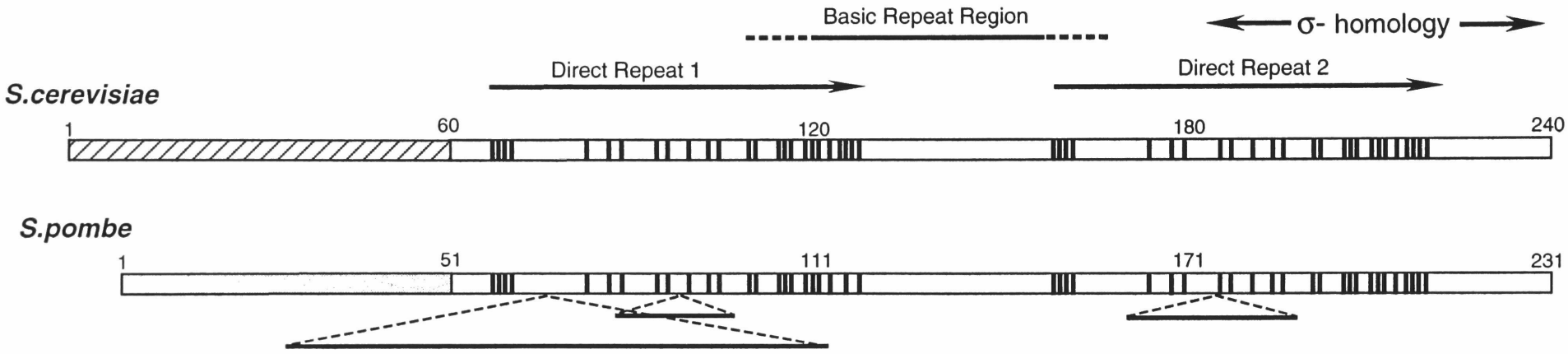
basal promoter activity. Given the strong conservation of these residues, these results are consistent with the earlier suggestion (Horikoshi et al., 1989a) that this repeat may play a role in mediating the action of certain types (e.g. acidic) of activators, that are functional in yeast and human.

Amino-terminal regions Contrasted with these conserved structural features are the divergent amino-termini. Although both hydrophilic, the N-termini are not only unrelated in sequence but also exhibit different amino acid compositions. The 60 residue *S. cerevisiae* amino terminus has a much higher density of charged residues (14 acidic and 10 basic) than the *S. pombe* amino terminus (7 acidic and 1 basic). On the other hand, the 51 residue *S. pombe* amino terminus contains high concentrations of serines (6) and threonines (5), as well as alanines (8) and asparagines (7), which are not found in the *S. cerevisae* proteins.

The precise separation of completely divergent amino termini from the highly conserved core domain is conspicuous and suggests a functional difference for these regions. The greater and more variable distance between the TATA box and the initiation site in *S. cerevisae* raises the possibility of a different initiation mechanism that might involve interactions of general factors (or other intermediates) with the amino-terminal region. Because the TATA box - initiation site distance on *S. pombe* promoters is closer to that observed for mammalian genes, it may be relevant that the amino terminus of the human TFIID, although larger, resembles that of *S. pombe* more closely (Hoffmann et al., 1990). On the other hand, and especially because the TFIID core is sufficient for basal promoter activity (Horikoshi et al., 1990), it seems equally probable that the amino termini may be involved in differential interactions with regulatory factors. The availability of active regulatory factors and

Figure 15. Schematic of the sequence motifs found in the newly identified TFIID core domain.

The schematic shows both the large carboxy-terminal core that is highly conserved (93% sequence identity) between *S. pombe* and *S. cerevisae* and the conserved motifs within this core. Vertical bars in the cores indicate the residues that are identical within the direct repeats; the lower horizontal bars indicate the positions and relative sizes of the introns in the *S. pombe* gene.


Basic Repeat Region
σ- homology
Direct Repeat 1
Direct Repeat 2
S.cerevisiae
1
60
120
180
240
S.pombe
1
51
111
171
231

reconstituted cell free systems from several organisms, including *S. cerevisiae*, should allow these possibilities to be tested.

(ii) TFIID from higher Eukaryotes

This section represents a summary of previously published papers (Hoffmann et al., 1990; Gasch et al., 1990; Muhich et al., 1990; Tamura et al., 1991) describing the isolation of cDNAs encoding 'TFIID' from various organisms, which I undertook personally or was involved in as the principal collaborator.

Isolation of human cDNAs

Extensive studies by low-stringency genomic Southern blots had failed to reveal a DNA fragment that cross-hybridized specifically and reproducibly with the previously cloned TFIID gene from *Saccharomyces cerevisiae* (data not shown), despite the apparent functional conservation between yeast and human TFIIDs. Subsequent cloning of the TFIID gene from the fission yeast *Schizosaccaromyces cerevisiae* (Hoffmann et al., 1990) as well as deletion analyses of the *S. cerevisiae* TATA factor (Horikoshi et al., 1990) identified a so-called 'TFIID core region' that is highly conserved in the two divergent yeasts and is necessary and sufficient for TATA box binding and transcriptional activities.

On the basis of these findings, avoiding obviously non-conserved amino-acid residues, all-inclusively degenerate oligonucleotides were designed coding for those amino acid sequences of the TFIID core region that would minimize the number of degenerate primers. A graphic analysis of the *S. cerevisiae* TFIID amino acid sequence for this purpose (Fig. 16a) shows up to five such short sequence segments suitable for designing primers used in a polymerase chain reaction (PCR). These were synthesized and used in appropriate combinations to amplify DNA fragments from a cDNA library at low stringency. From the many resulting fragments (Fig. 16b) putatively positive bands upon agarose gel size fractionation were identified by satisfying two criteria: (i) matching the expected size for a specific primer pair, based on the assumption that the 'TFIID core region' is conserved

Figure 16. Cloning human TFIID by polymerase chain reaction (PCR).

A. oligonucleotide primer design for PCR was guided by a previously mapped homology region in TFIID (Hoffmann et al., 1990a), as well as an analysis of the *S. cerevisiae* TFIID amino acid sequence to minimize the degeneracy of all-inclusive primers. The graph shows the degeneracy (Y-axis) of an all-inclusive 18 nucleotide oligomer coding for a 6 amino acid stretch in the *S. cerevisiae* TFIID sequence whose position within the protein is mapped onto the X-axis. Five polypeptide segments within the TFIID core domain have relatively low degeneracies and are designated A, B, C, D, and E. All-inclusively degenerate primers are synthesized in both sense (+) and anti-sense (-) directions as indicated. Primers were subsequently used in PCR singly as negative control reactions (data not shown) and in appropriate pairs. Based on the assumption of a collinear evolutionary conservation of the TFIID polypeptide, products with the following sizes were expected (A+B- : 212bp; A+C- : 239bp; A+D- : 328bp; A+E- : 413bp; B+D- : 152bp; B+E- : 224bp; C+D- : 119bp; C+E- : 194bp; D+E- : 95bp)

B. PCR products generated from indicated primer pairs are fractionated on a 2% agarose gel and stained with ethidium bromide after altogether 100 cycles of amplification with an intermediate replenishment of primers and Taq polymerase.

C. Low stringency Southern blot of the agarose gel in B. probed with a random primed probe made from the 738bp HincII-HinDIII fragment of *S. pombe* TFIID (Hoffmann et al., 1990a). Promising DNA fragments identified on the basis of expected size as well as cross-hybridizability at low stringency conditions are marked with "*" (lanes 1,2, and 6). The only specific TFIID PCR product is the one in lane 1.

(These experiments were carried out in collaboration with E. Sinn.)

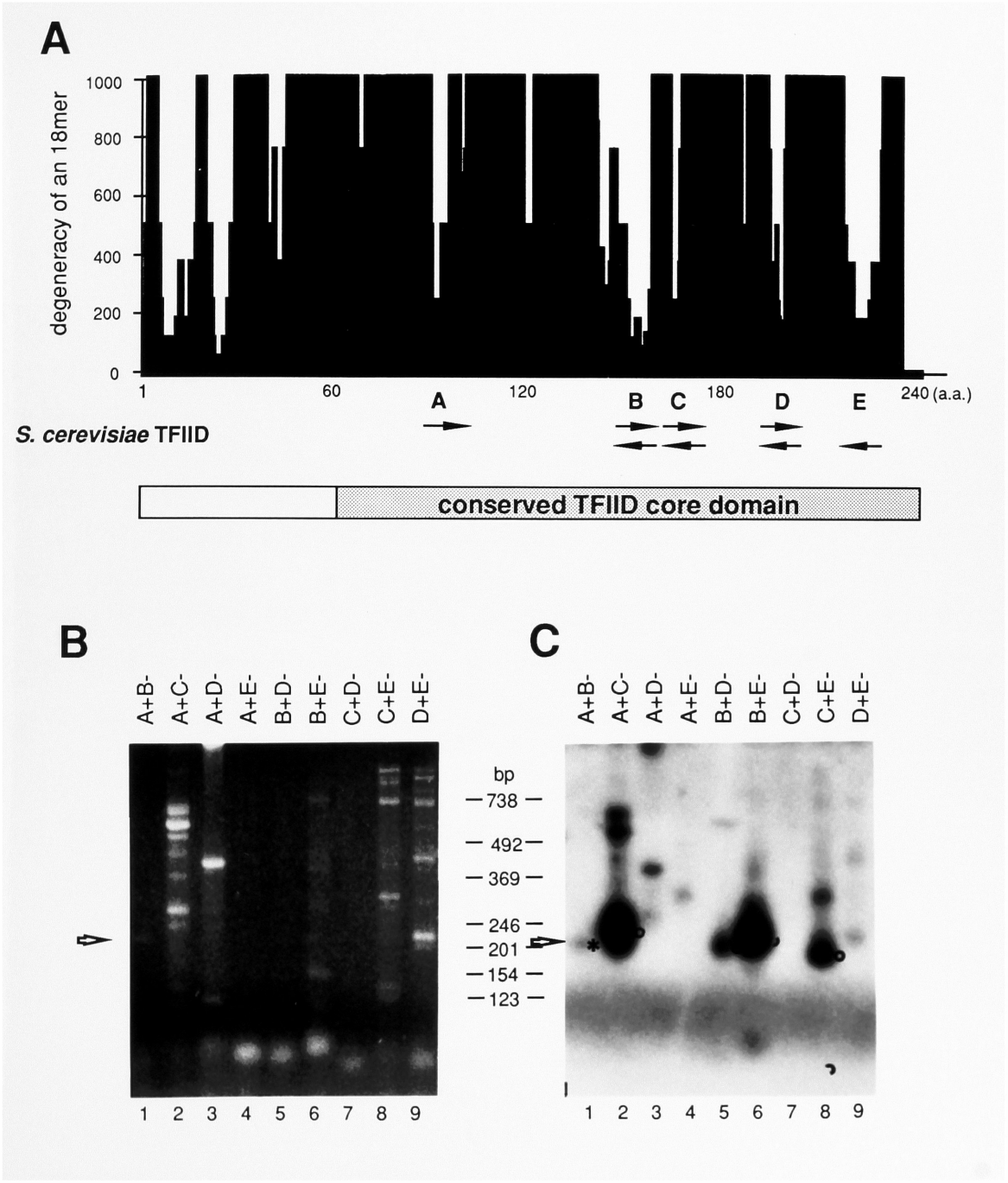
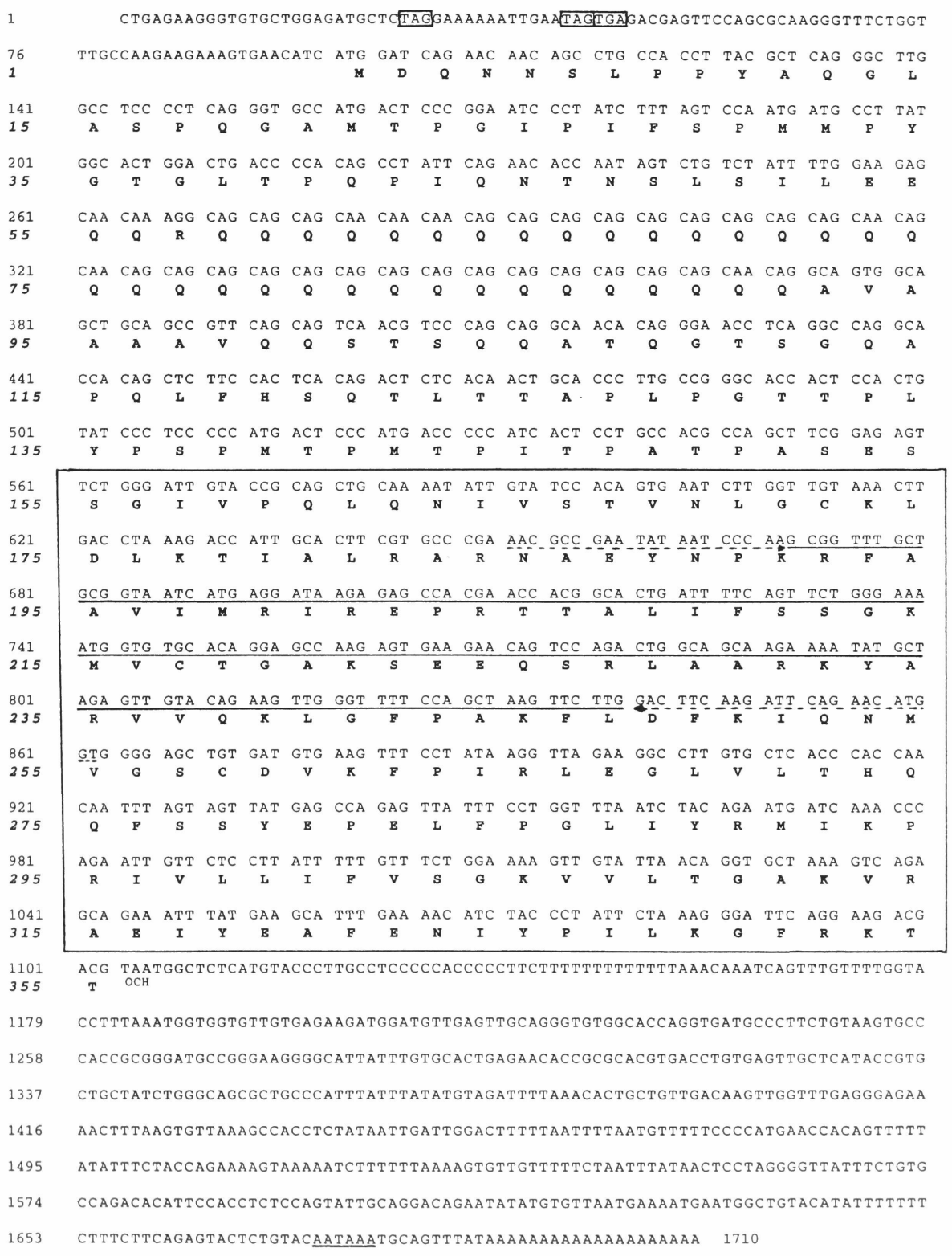

A
degeneracy of an 18mer
1000
800
600
400
200
0
1
60
120
180
240 (a.a.)
A
B
C
D
E
S. cerevisiae TFIID
conserved TFIID core domain
B
A+B-
A+C-
A+D-
A+E-
B+D-
B+E-
C+D-
C+E-
D+E-
1 2 3 4 5 6 7 8 9
C
A+B-
A+C-
A+D-
A+E-
B+D-
B+E-
C+D-
C+E-
D+E-
bp
738
492
369
246
201
154
123
1 2 3 4 5 6 7 8 9

Figure 17. Nucleotide and predicted amino acid sequence of human TFIID cDNA.

An open reading frame of 1005bp was found in three independent, but overlapping cDNAs (all greater than 1.2kb in length) whose combined sequence is shown here. It encodes a polypeptide containing the characteristic TFIID homology core domain (amino acids 155 - 335, boxed, Hoffmann et al., 1990a) as well as interesting sequence motifs in its N-terminal half that are unique to this human polypeptide. Translation start and stop codons are indicated by boxes in the DNA sequence; a putative polyadenylation signal near the 3'end is underlined. The original TFIID specific PCR product correponds to sequences 651-862 and is undelined flanked by arrows indicating the primers.

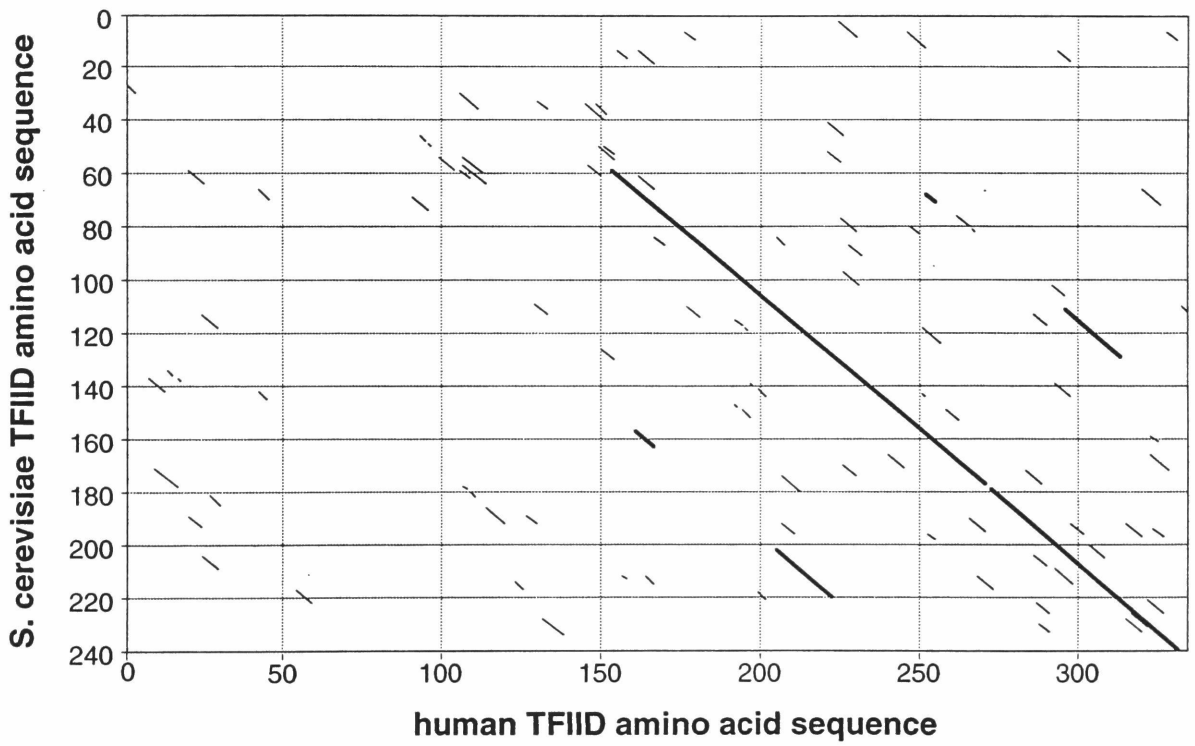


collinearly in evolution, as exemplified by the two available yeast sequences; and (ii) a degree of cross-hybridizability to a DNA fragment of the yeast TFIID genes (Fig. 1c). Three such fragments were initially identified, gel purified and cloned into a plasmid vector upon further PCR-mediated amplification. DNA sequencing of the inserts revealed an open reading frame (ORF) in one of these (original PCR product of 212bp in lane 1) that had indeed a high degree of amino-acid similarity to the yeast TFIID sequences in the expected location (from primer positions) within the TFIID core region. The other two inserts (oiginal PCR products in lanes 2 and 6) though being positive on a low stringency Southern did not contain any ORFs nor were they of the exact expected size as determined by DNA sequencing. Their origin or significance is not known and their occurrence can only underline the importance of a good selection system when PCR is used at low stringency conditions where artifacts and noise can produce seemingly higher signals than the specific product.

Using this 212bp insert the same library was screened, and after several rounds of plaque purification three positive clones were isolated. Subsequent sequencing showed that they were all independent cDNAs from messages of the same gene that also contained the exact fragment used for the screen (apart from a few expected mismatches within the primer annealing regions).

The complete ORF (Fig. 17) reconfirms the validity of conclusions based earlier on sequence comparisons between *S. cerevisiae* and *S. pombe* TFIID (Hoffmann et al., 1990) and structure/function analyses (Horikoshi et al., 1990): human TFIID also consists of an extraordinarily highly conserved TFIID core region in the C-terminal part of the protein as shown by matrix comparison with *S. cerevisiae* TFIID (Fig. 18) whereas the N-terminus apparently bears no homology at all to equivalent regions in either yeast TFIIDs. The N-terminal region of human TFIID in fact is much larger than the

Figure 18. Matrix comparison of amino acid sequences of TFIID from *Saccharomyces cerevisae* and human.
Thick vertical lines indicate significant sequence similarities between teast and human TFIID amino acid sequences. The main line demonstrates complete collinearity in conservation of the characteristic TFIID core domain; flanking shorter line are indicative of the direct repeat structure of this core domain.



Pustell protein matrix

7 aa window

20% identity

previously isolated TFIIDs from other species and contains an interesting primary structure consisting of a conspicuous glutamine repeat (termed 'Q-run') flanked by two domains rich in serines, threonines, and prolines organized into a characteristic repeat structure (discussed below). The predicted molecular weight of the human TFIID polypeptide is 37,166 Daltons, larger than its yeast counterparts but signficantly less than previous size estimates of glycerol gradient or gel filtration fractions with TFIID activity. An interesting possible explanation for this discrepancy is that TFIID in its native state in the nucleus is part of a multiprotein complex.

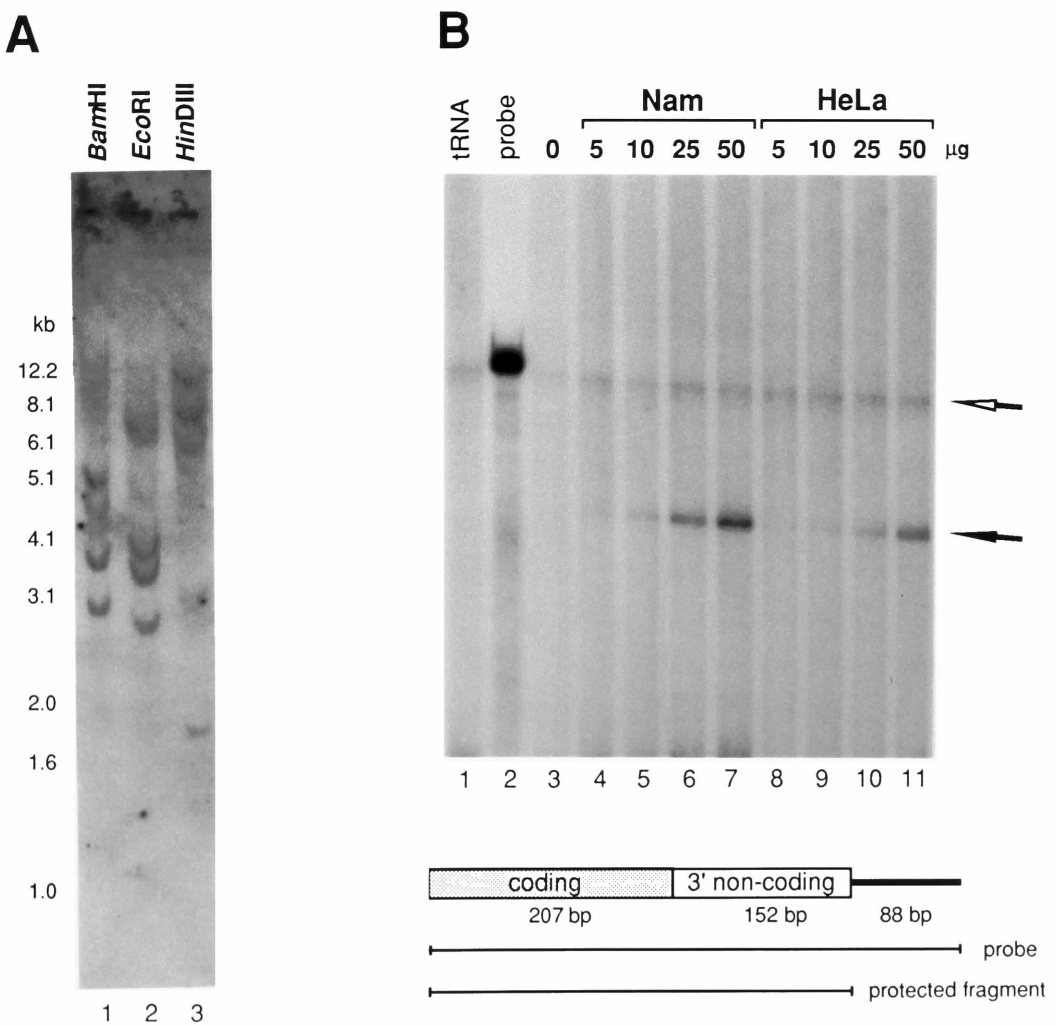
A genomic Southern analysis with an anti-sense RNA fragment from part of the TFIID core region synthesized to a very high radio-specific activity (Fig. 19a) hybridized to several high molecular weight human placental DNA fragments at high stringency conditions. Longer DNA probes or probes with less specific activity failed to reveal any appreciable signal above background under conditions when other single copy genes used as controls were successfully displayed (data not shown). These observations can be interpreted to point to a highly complex genomic organization of the TFIID gene. This might involve multiple highly homologous functional or non-functional (pseudogenes) copies of the gene, and/or a complex exon-intron structure. Such a scenario has previously been observed for the TFIID genes in *Schizosaccharomyces pombe* (Hoffmann et al., 1990) and *Arabidopsis thaliana* (Gasch et al., 1990) where 3 and 8 introns respectively are considered high numbers of introns for such a short message in the respective organisms. As little is currently known about the origin and function of intervening sequences in eukaryotic genomes and only in a few cases correlations between exon-intron structure and functional domain organization of the protein have been convincingly shown, it remains to be seen what possible significance such a complex TFIID gene structure might have.

Figure 19. The human TFIID gene and mRNA.

A. Human placental DNA probed on a Southern blot with an RNA probe synthesized from a fragment of the TFIID cDNA shows several bands when digested with either *Bam*HI (lane 1), *Eco*RI (lane 2), or *Hin*dIII (lane 3), providing evidence that the isolated cDNAs originate indeed from a human gene.

B. An RNase protection assay with the same RNA probe detects TFIID messenger RNA in both Namalwa (lanes 4-7) and HeLa cells lanes (8-11), while no such signal is present in indicated control lanes (lanes 1-3).

(These experiments were carried out by E. Sinn.)


A
BamHI
EcoRI
HinDIII
kb
12.2
8.1
6.1
5.1
4.1
3.1
2.0
1.6
1.0
1 2 3
B
tRNA
probe
Nam
HeLa
0 5 10 25 50 5 10 25 50 µg
1 2 3 4 5 6 7 8 9 10 11
coding
3' non-coding
207 bp
152 bp
88 bp
probe
protected fragment

Using the same anti-sense RNA probe as above, an RNase protection assay (Fig. 19b) shows that the gene whose cDNA has been cloned from a Namalwa (NAM) cell library is expressed both in NAM (lanes 4-7) and in HeLa (lanes 8-11) cells in almost equal amounts. However, to detect such a signal the most sensitive method had to be employed; Northern blots with either total or poly(A)$^+$ selected RNA did not show any signal above background pointing towards a relatively small number of transcripts of a well-defined length, possibly indicating a tight regulation of TFIID gene expression either at the level of RNA synthesis, processing or degradation.

TFIID cloning from mouse, insect, and plant species

Given the identification of a highly conserved TFIID core domain (Hoffmann et al., 1990; Hoffmann et al., 1990), it was not surprising that a human TFIID cDNA-derived probe encoding 132 residues within the core domain specifically hybridized at low stringency to DNA fragments contained in *Bam*HI and *Hin*DIII restricted mouse genomic DNA (T. Tamura, personal communication). The same probe was then used at similar hybridization conditions to screen oligo(dT)- and random hexamer-primed mouse brain libraries yielding overlapping cDNA clones that were apparently derived from a single RNA species and was used to construct a 1652bp cDNA (Tamura et al., 1991). The ORF beginning at position 63 encodes a 316 amino acid polypeptide with 92% overall sequence identity to the human TFIID sequence. The C-terminal 221 amino acid sequence, including the previously identified 180 residue TFIID core domain, was identical to the corresponding region in human TFIID. Nucleotide sequences between mouse and human TFIID cDNAs revealed a high (90%) sequence identity in the coding region and the most distal 3'-untranslated region (downstream from 1200), but weak similarities in the proximal 5' and 3' non-coding regions.

Several genomic clones from *Arabidopsis thaliana* as well as a large number of cDNAs were isolated (Gasch et al., 1990) using above-described PCR-mediated cloning (A. Gasch, personal communication) as well as direct hybridization techniques probes derived from the *S. cerevisae* TFIID gene. Interestingly, the sequenced clones fell into two classes coding for very similar but distinct amino acid sequences (At TFIID-1 and At TFIID-2). Genomic Southern analysis confirms that there are at least two closely related genes present in the *Arabidopsis* genome. Northern analysis using gene-specific probes reveals that the At-1 and At-2 messenger RNAs are 1.4kb and 1.3kb in size, respectively and that they are present in roughly equal amounts. No evidence for tissue- or light-specific regulation could be detected (A. Gasch, personal communication). The open reading frames of At TFIID-1 and At TFIID-2 both code for proteins of 200 amino acids containing a C-terminal TFIID core domain with 95% sequence identity to each other and 85% sequence identity to the yeast TFIID proteins. The short 18 amino acid N-termini exhibit 75% sequence identity to each other and no similarity to TFIID sequences from other species. As expected, both proteins, when expressed in bacteria, are functional in basal transcription assays on a template containing the adenovirus major late promoter when complemented with TFIID deficient HeLa nuclear extract-derived fractions. Rabbit reticulocyte or wheat germ lysate produced At TFIID proteins shift TATA box-containing probes in multiple complexes, one of which is specific to At TFIID-2 indicating suspected differences in the protein interaction surfaces of the two proteins.

A Drosophila TFIID cDNA clone was isolated (Muhich et al., 1990) with a probe obtained by PCR amplification of Drosophila DNA, with primers selected from regions of low codon degeneracy within the yeast TFIID gene (as described above, Fig. 16). A 173bp PCR product revealed a high level of amino acid sequence identity to the *S. cerevisae* TFIID gene and was employed in Southern hybridization which

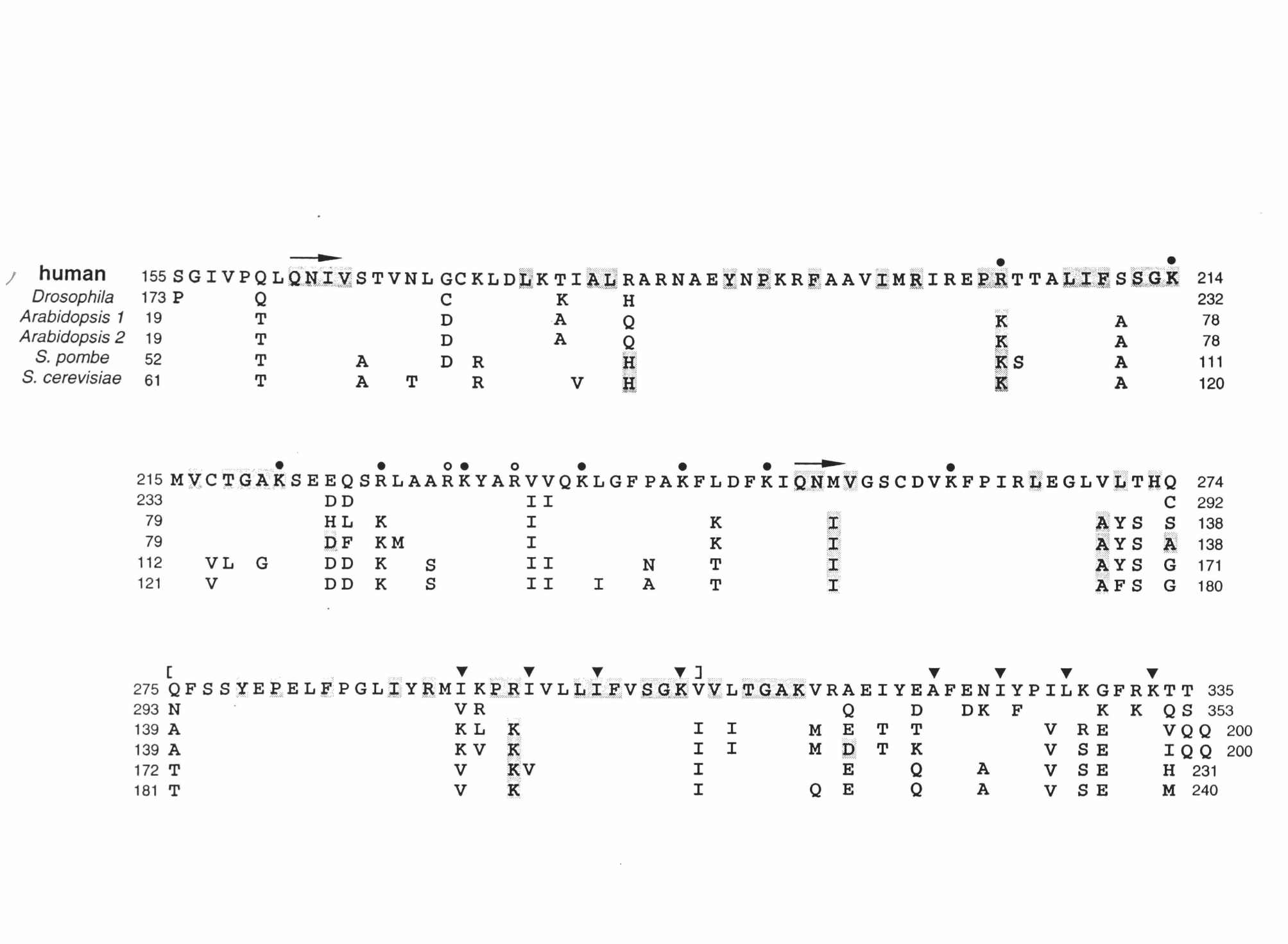
indicated a single genomic locus as well as in an embryonic cDNA library screen. A 1.5kb cDNA contained an open reading frame of 353 amino acids, which corresponds to a calculated mass of 38,407 Da. A TFIID core domain displaying 81% sequence identity to yeast and 88% identity to human TFIID core domains is located in the C-terminal portion of the polypeptide. Functional assays with this domain (196aa) produced in bacteria demonstrates its activity in basal transcription with HeLa derived complementing factors as well as its ability to form a DNase I resistant complex with the TATA element of the *Drosophila* fushi tarazu promoter.

Homologies in TFIIDs across the eukaryotic spectrum

The functional conservation of the TATA binding and the transcriptional initiation activities between yeast and human (Buratowski et al., 1988; Cavallini et al., 1988) is correlated with a remarkable degree of sequence conservation and complete collinearity (Fig. 18) between the ≈180-residue C-terminal regions of the TATA factors from different organisms (Fig.20). Moreover, this region correlates exactly with the region of *S. cerevisae* TFIID previously shown to be necessary and sufficient for TATA box-specific DNA binding and core promoter transcription (Horikoshi et al., 1990). With examples of TFIID amino acid sequences from organisms representing four different eukaryotic kingdoms we can identify several overlapping structural domains within this region that are particularly well conserved in evolution and are likely therefore to be important for the functioning of the TATA factor. Figure 20 details and Figure 22 summarizes schematically, a comparison of the human TFIID sequence (Hoffmann et al., 1990; Kao et al., 1990; Peterson et al., 1990) with TFIID sequences from *Saccharomyces cerevisiae* (Cavallini et al., 1989; Eisenmann et al., 1989; Hahn et al., 1989a; Horikoshi et al., 1989a; Schmidt et al., 1989),

Figure 20. Sequence comparison of the highly conserved TFIID core domain.

The sequence of the human TFIID core domain is displayed on the top line, whereas the lower lines show only amino acids that differ in the TFIID sequences from other species. Putative structural domains that are discussed in the text are indicated as follows: (1) direct repeats, long arrows indicate regions containing the two interrupted repeats whose constituent residues are shaded; (2) basic repeat, solid circles (lysines) and open circles (arginines) indicate residues in the central basic core which lie on one face of potential α-helical structures; (3) sigma homology, the region with an overall sequence similarity to the sigma 2.3 and 2.4 domains is indicated by brackets while solid inverted triangles denote residues, in two small regions, that correspond to those most highly conserved in the 2.4 regions of various sigma factors. The regions with sequence similarities to the helix-loop-helix proteins (Gasch et al., 1990) extend from residues 201-220 in direct repeat 1 and from 260 to 314 in direct repeat 2.


human 155 SGIVPQLQNIVSTVNLGCKLDLKTIALRARNAEYNPKRFAAVIMRIREPRTTALIFSSGK 214
Drosophila 173 P Q C K H 232
Arabidopsis 1 19 T D A Q K A 78
Arabidopsis 2 19 T D A Q K A 78
S. pombe 52 T A D R H K S A 111
S. cerevisiae 61 T A T R V H K A 120
215 MVCTGAKSEEQSRLAARKYARVVQKLGFPAKFLDFKIQNMVGSCDVKFPIRLEGLVLTHQ 274
233 DD II C 292
79 HL K I K I AYS S 138
79 DF KM I K I AYS A 138
112 VL G DD K S II N T I AYS G 171
121 V DD K S II I A T I AFS G 180
275 QFSSYEPELFPGLIYRMIKPRIVLLIFVSGKVVLTGAKVRAEIYEAFENIYPILKGFRKTT 335
293 N VR Q D DK F K K QS 353
139 A KL K I I M E T T V R E VQQ 200
139 A KV K I I M D T K V S E IQQ 200
172 T V KV I E Q A V S E H 231
181 T V K I Q E Q A V S E M 240

Schizosaccharomyces pombe (Fikes et al., 1990; Hoffmann et al., 1990), *Arabidopsis thaliana* (Gasch et al., 1990), *Drosophila melanogaster* (Hoey et al., 1990; Muhich et al., 1990), and mouse (Tamura et al., 1991). Altogether, these observations clearly argue for a conserved TFIID 'core' which explains the evolutionary conservation of TFIID function.

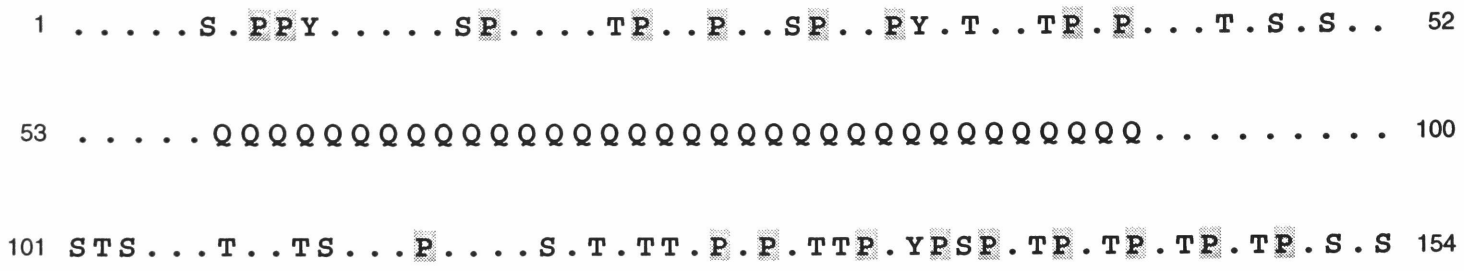
The TFIID core also contains the structural motifs (Fig. 20, 22) recently noted (Cavallini et al., 1989; Horikoshi et al., 1989a; Gasch et al., 1990; Hoeijmakers, 1990; Nagai, 1990; Stucka and Feldman, 1990). The residues comprising these elements show an even more striking (near absolute) conservation from yeast through to human, which argues strongly for a conservation of functional roles as follows.
(i) The direct repeat structure first noted in the yeast TATA factor protein sequence (Cavallini et al., 1989; Hoeijmakers, 1990; Nagai, 1990; Stucka and Feldman, 1990) is a perfectly conserved structural feature of the protein and which allow an element of symmetry in the folded TFIID molecule suggesting a tertiary structure consisting of two domains originating from an ancient duplication event. Such an intra-molecular dimeric structure with its inherent symmetry might be important in binding DNA, similar to the inter-dimeric activator complexes analyzed in recent years in prokaryotes (eg. cro, lambda repressor) and eukaryotes (eg. leucine zipper, helix-loop-helix proteins).
(ii) There is a putative homology to bacterial sigma factors (Horikoshi et al., 1989a), which overlaps the the C-terminal direct repeat and introduces an element of asymmetry to the direct repeat structure. The fact that this similarity is to amino acids within the 2.3 and 2.4 regions that are also conserved among diverse sigma factors and have been shown to be responsible for binding the -10 box (consensus TATAAT) might suggest that these amino acids in the TATA factor are important in determining the specificity of DNA binding, an interpretation consistent with recent point mutation results (Yamamoto et al., 1992).

(iii) There is a repeat of basic residues (notably lysines) in the central basic core connecting the direct repeats that has the potential for formation of an α-helix with basic residues on one face (Horikoshi et al., 1989a). Appropriate mutagenesis studies indicate that individual basic residues are required neither for DNA binding nor for basal transcription (Yamamoto et al., 1992), consistent with the earlier suggestion that the function of the basic repeat helix might be one of interaction with other members of the multi protein initiation complex, or in fact with acidic activators which have been shown to interact directly with the TATA factor (Stringer et al., 1990).
(iv) There are homologies to the helix-loop-helix domain of myc related proteins, especially the Id protein, first noted by Gasch et al. (1990) for the *Arabidopsis* and yeast proteins. As discussed by these authors, this raises the possibility that these regions are important either for homotypic interactions within TFIID or for heterotypic interactions with distinct regulatory factors.

In striking contrast to the conserved C-terminal regions, the N-terminal regions of TFIID differ markedly both in size and sequence (Fig. 22) across the eukaryotic spectrum, yet show similarity within the grade of *bilateria*. The human N-terminus, for example, lacks the high density of charged residues previously noted in yeast TFIID, but contains obvious structural features that could be functionally related to similar structural motifs in other regulatory factors. These include (see Fig. 21): a region containing 34 consecutive glutamines (the Q-run) flanked by two regions (STP-1 and STP-2) rich in serine (S), threonine (T) and proline (P) residues. The latter show a general similarity to Ser, Thr, (Pro)-rich regions in regulatory proteins like Sp1 (Courey and Tjian, 1988) and GAL11 (Suzuki et al., 1988), whereas the Ser-Pro, Thr-Pro and Tyr-Pro/Pro-Tyr repeats (usually preceded, followed by or interspersed with an Ala,

Figure 21. Sequence motifs in the N-terminal structure of the human TATA factor.

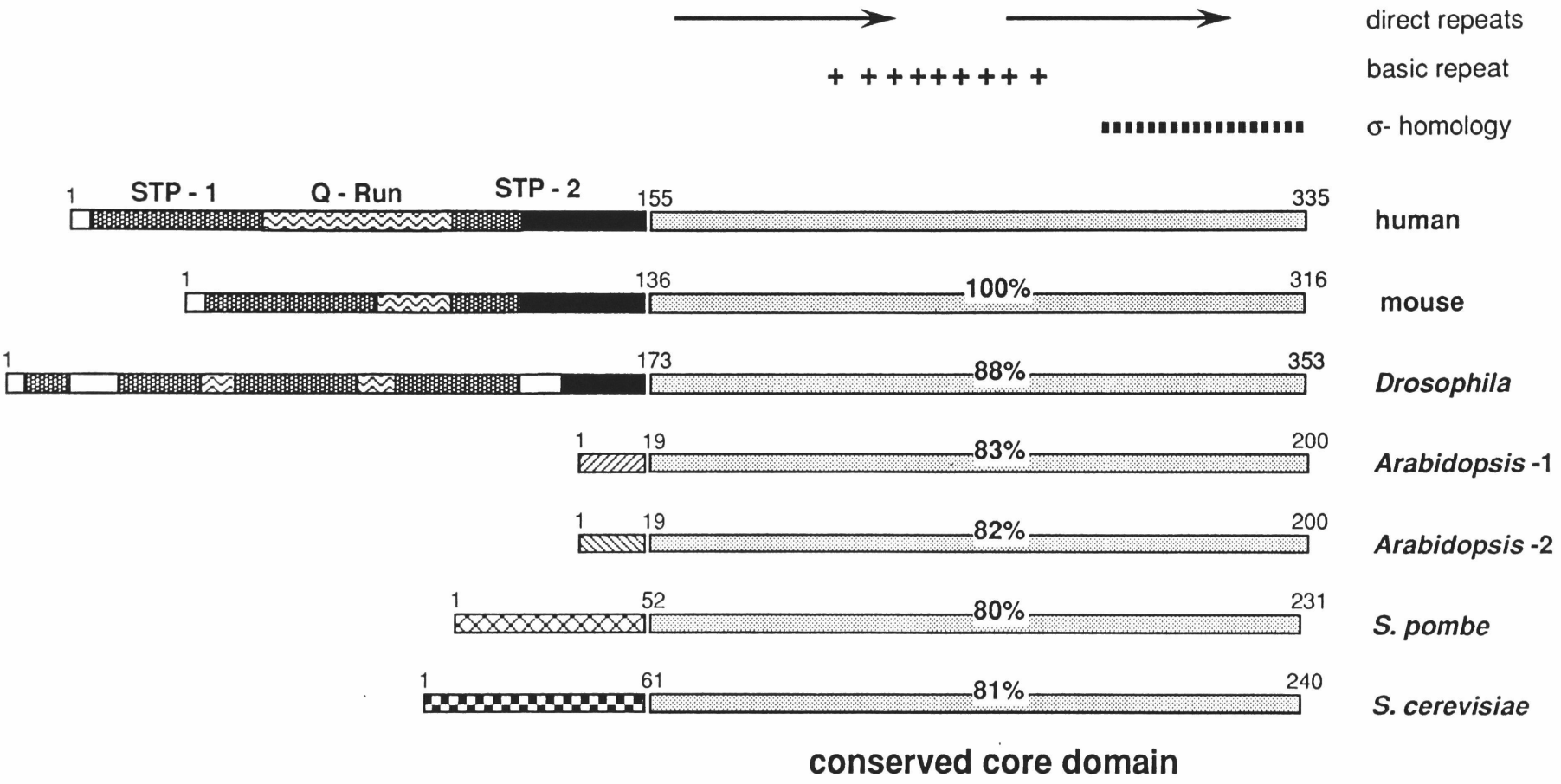
The potentially interesting primary structure motifs highlighted include a glutamine repeat which is flanked by two regions (STP-1 and 2) that are rich in serine, threonine and proline and characterized by triplets containing Ser-Pro, Thr-Pro, and Pro-Tyr/Tyr-Pro pairs preceded or followed by a Met, Ile, Leu, or Ala residue (not shown).



RNA pol II large subunit heptad repeat consensus Y S P T S P S

Ile, Leu or Met residue) are reminiscent of the reversibly phosphorylated heptad repeat (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) of the C-terminal domain of the largest subunit of RNA polymerase II (Payne et al., 1989). Similarly the Q-run and adjacent Gln-rich regions resemble polyglutamine and Gln-rich regions present in other factors (Courey and Tjian, 1988; Suzuki et al., 1988; Mitchell and Tjian, 1989) and implicated, in some cases, as activation domains (Courey and Tjian, 1988; Gerster et al., 1990). These regions might well serve as important interfaces for intermolecular protein-protein interactions. Given the lack of conservation of this domain to plants or yeasts, it is not surprizing that structure-function studies in the yeast protein suggest that this domain is not essential for basal transcription functions *in vitro* (Horikoshi et al., 1990). However the human N-terminus might be involved in mediating regulatory factor interactions, either with TFIID itself or with other general factors, and the possibility is raised that analogous functions might be encoded in separate factors (such as GAL11) in lower eukaryotes.

Figure 22. Evolutionary conservation of the TATA factor. The degree of sequence identity (relative to human) within the conserved C-terminal core domains of the TFIIDs from human, mouse, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* are summarized. Also indicated are the positions in the core of the structural domains discussed in the text. The N-terminal regions of these proteins differ markedly both in length and amino-acid composition and sequence, except that the Drosophila N-terminus shows some sequence similarity to the mammalian N-termini which are almost identical. Black boxes indicate highly ordered triplet arrangement discussed in the text and Figure 21, dark grey shading indicate other parts of the STP-rich regions. Runs of glutamine residues are indicated by wavy shading.


direct repeats
basic repeat
σ- homology
1
STP - 1
Q - Run
STP - 2
155
335
human
1
136
100%
316
mouse
1
173
88%
353
Drosophila
1
19
83%
200
Arabidopsis -1
1
19
82%
200
Arabidopsis -2
1
52
80%
231
S. pombe
1
61
81%
240
S. cerevisiae
conserved core domain

(iii) Comparing recombinant vs. native TFIID

Functional analysis of the human cDNA-encoded protein

To assay the activity of the protein encoded by the human cDNA, a fragment extending from nucleotides -95 to +1170 relative to the first ATG of the longest open reading frame (Fig. 17) was *in vitro* transcribed by T7 RNA polymerase and subsequently expressed in rabbit reticulocyte lysate. The resulting polypeptide had a mobility on an SDS-polyacrylamide gel equivalent to 38kd (Fig. 23a), being in good agreement with the predicted relative molecular weight of 37,166 Daltons.

Mobility shift assays (Fig. 23b) with these same lysates showed TATA box-specific DNA binding as demonstrated by the relative sensitivities to unlabelled wild-type (lanes 4-6) and mutant (lanes 7-9) TATA box containing oligonucleotide competitors. Interestingly, the *in vitro* expressed cloned cDNA p38 produced a mobility shift of the same order as partially purified human TFIID (lanes 10-12) whose native size has previously been estimated at 110kd by gel filtration. Whether this complex with unusually slow mobility (relative to the cloned yeast TFIIDs, (Hoffmann et al., 1990)) represents a protein-DNA complex involving polypeptides other than p38 that might be present in the reticulocyte lysate is an interesting possibility. What is apparent, however, is that a second p38-dependent, TATA box-specific complex is present only with recombinant TFIID. A third, somewhat broader band of even lower mobility is evidently independent of mRNA exogenously added to the lysate.

A second criterion for TFIID activity is transcriptional initiation in a TFIID-dependent transcription system. This can be in a reconstituted transcription system (Fig. 23c, lanes 1-4) where other essential basic factors (TFIIB, TFIIE/F, polII) are used in partially purified form or in heat-treated nuclear extract (Fig. 23c, lanes 5-9).

Figure 23. ***In vitro*** **expression and function of the human cDNA-encoded TFIID in TATA box binding and basal transcription.**

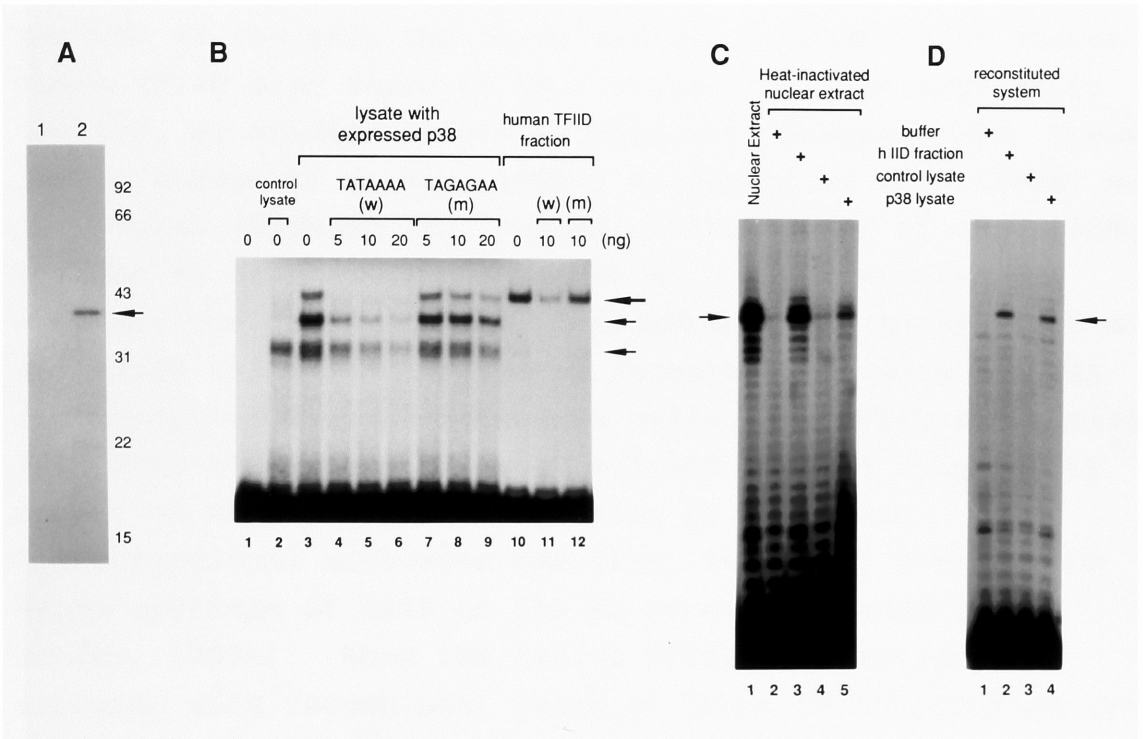
A. SDS-PAGE analysis of (^{35}S)methionine-labelled TFIID produced in reticulocyte lysate. Lysates were programmed with no RNA (lane 1) or T7 RNA polymerase produced RNA from a human TFIID cDNA (lane 2).

B. The expressed human cDNA binds the TATA sequence specifically. Gel shift assays with an adenovirus major late promoter (MLP) fragment contained : lane 1, no protein; lane 2, negative control lysate (without RNA), lanes 3-9, lysate programmed with human TFIID mRNA; lanes 10-12, partially purified human TFIID. Unlabelled oligonucleotides (MLP, positions -45 to -15) contain either a wild type (w) or a mutant (m) TATA box and were added as competitors in the ng amounts indicated above each lane. The bold arrow indicates the human TFIID specific complex whereas the significance of the secondary bands indicated by the thin arrows is discussed in the text.

C. Expressed human TFIID cDNA can complement heat sensitive TFIID activity. Transcription assays either nuclear extract ('+'control, lane 1) or heat-treated nuclear extracts complemented with either buffer (lane 2), partially purified human TFIID (lane 3), control (no RNA) reticulocyte lysate (lane 4), and lysate programmed with the hTFIID cDNA. The primer extension product resulting from accurate transcription initiation is indicated by the arrow.

D. Expressed hTFIID cDNA is essential for transcription. Using partially purified basic transcription factors TFIIB, TFIIE/F and RNA polymerase II, an *in vitro* reconstituted transcription system is complemented with buffer (lane 1), partially purified TFIID (lane 2), unprogrammed lysate (lane 3), and lysate with expressed hTFIID cDNA (lane 4).

(These experiments were carried out to a large part by T. Yamamoto.)

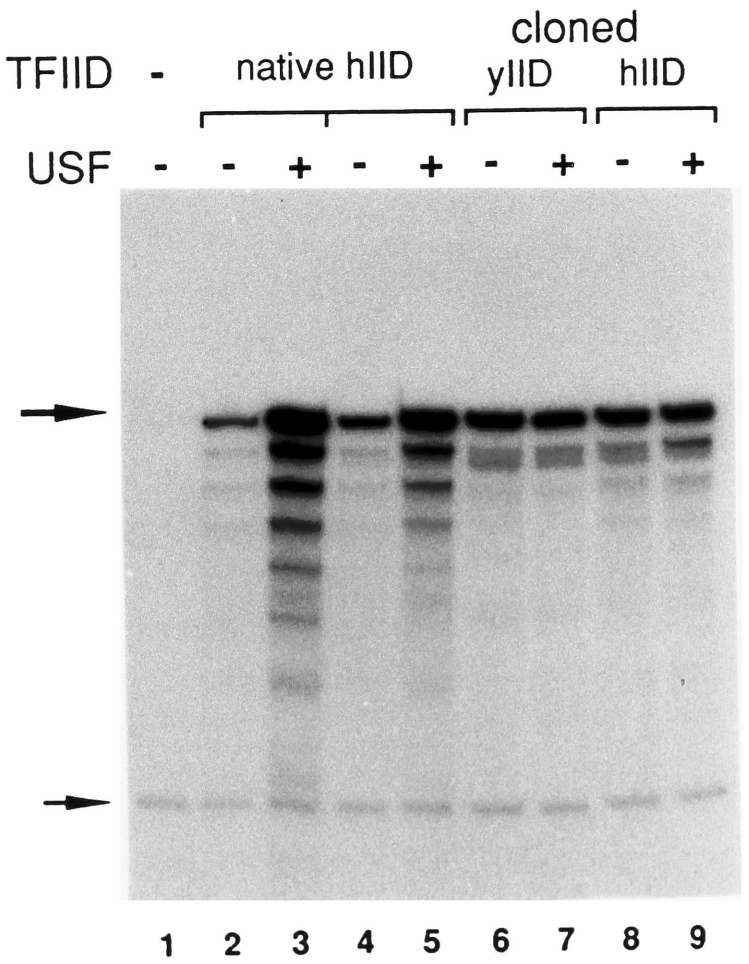

A
1 2
92
66
43
31
22
15
B
lysate with expressed p38
human TFIID fraction
control lysate
TATAAAA (w)
TAGAGAA (m)
(w) (m)
0 0 0 5 10 20 5 10 20 0 10 10 (ng)
1 2 3 4 5 6 7 8 9 10 11 12
C
Heat-inactivated nuclear extract
Nuclear Extract
1 2 3 4 5
D
reconstituted system
buffer
h IID fraction
control lysate
p38 lysate
1 2 3 4

The expressed p38 (lanes 4 and 9) can evidently substitute for a native, partially purified TFIID preparation (lanes 2 and 7) despite the fact that reticulocyte lysate can have a repressing effect in such an assay (data not shown) as well as containing residual amounts of TFIID-like activity itself (compare lanes 6 and 8). These results provide convincing evidence that the cloned cDNA encodes a functional protein with TATA box-binding and core promoter transcription activities.

Consistent with its key role in the promoter activation pathway as the only DNA-bound general transcription factor, human TFIID also seems to be a target for some regulatory factors, as evidenced both by physical (Sawadogo and Roeder, 1985b; Horikoshi et al., 1988a; Horikoshi et al., 1988b) and functional (Workman and Roeder, 1987; Abmayr et al., 1988; Workman et al., 1988; Workman et al., 1990) studies of interactions between these components. When analysed in a cell-free transcription system reconstituted with general transcription factors from HeLa cells, two TFIID-containing fractions of differing purity allowed for the stimulatory effect on transcription initiation by the human transcriptional activator USF (Fig, 24, lanes 2-5), which binds upstream of TATA on the ML promoter (Sawadogo and Roeder, 1985b). When the native TFIID preparation was replaced with recombinant yeast or human TFIID purified from overexpressing bacteria, harboring the TFIID ORF on an inducible promoter plasmid, no detectable USF-dependent enhancement was observed (Fig. 24, lanes 6-9). Possibly related to this difference, the native human TFIID showed transcripts in the guanosine(G)-less cassette assay indicative of multiple initiations on the same template (Fig. 24 legend) whereas the bacterially expressed proteins did not. This experiment, by contrasting recombinant with native TFIID, provides direct evidence to date that natural TFIID is directly involved in mediating activator effects. Given the inability of the recombinant protein to mediate

Figure 24. USF-stimulated transcription with TFIID.
Reactions containing an adenovirus MLP template with upstream sequences (-400 to +10) attached to the 380-bp G-less cassette pML(C_2AT) and HeLa factors TFIIB, TFIIE/F, and RNA polymerase were complemented with buffer (lane 1), DE52 fraction of native TFIID (lanes 2,3), amino-octyl fraction of native TFIID (lanes 4,5) and purified bacterially expressed yeast (lanes 6,7) or human (lanes 8,9) TFIID. A DE52 fraction of human activator USF was present in lanes 3,5,7 and 9. Transcription from the G-less cassette was in the absence of GTP; the upper arrow indicates the full length cassette transcript resulting from accurate initiation at +1 whereas the shorter ≈30-bp spaced bands (lanes 2-5) are indicative of multiple initiations on the same template. The lower arrow indicates the position of a labelled DNA fragment added as an internal standard.

(This experiment was carried out by M. Horikoshi.)


TFIID
-
native hIID
cloned
yIID
hIID
USF
- - + - + - + - +
1 2 3 4 5 6 7 8 9

transcriptional activation, several possible explanations can now be addressed such as essential post-translational modifications, TFIID-associated factors, or simply co-fractionating, but essential, factors present in the native TFIID preparation.

TFIID gel shift: interacting factors

While bacterially expressed recombinant human TFIID was apparently incapable of mediating the effect of a *trans*-activator, mobility shift assays with reticulocyte lysate produced recombinant human TFIID demonstrated a DNA-protein complex of equal mobility as the partially purified preparation from native sources. The mobility of this complex, furthermore, was remarkably low compared to those obtained with either yeast (*S. cerevisae* or *S. pombe*) TFIID with the same probe. Further indication for the presence of human TFIID interacting proteins present in the lysate was obtained in a comparison of the relative mobilities of human and yeast TFIIDs and an N-terminally truncated protein, consisting of the human TFIID core domain (residues 154-181). While the mobility in SDS-polyacrylamide electrophoresis of these reticulocyte lysate produced proteins reflected their relative molecular mass (Fig. 25A), gel shift analysis indicated a much slower complex containing the human 180 residue core domain than the 240 residue yeast protein (Fig. 25B). Competition with unlabelled oligonucleotides demonstrated the specificity of the DNA-protein complex, and a co-translation experiment was not indicative of dimerization (Fig. 25C). Thus it seemed likely that rabbit reticulocyte lysate contained a protein that stably interacts under the conditions employed only with human TFIID, and, somewhat surprisingly via the highly conserved core domain. Such an interpretation is consistent with subsequently published yeast *in vivo* analysis demonstrating that functional differences between yeast and human TFIID,

Figure 25. Gel mobility shift assays with *in vitro* translated yeast and human TFIIDs.

A. SDS-PAGE analysis of (^{35}S)methionine-labelled TFIIDs produces in reticulocyte lysates. Lysates were programmed with RNAs derived from indicated sources. Arrows with numbers corresponding to lane numbers indicate specific full length polypeptides.

B. Gel shift assays with proteins shown in A. bound on a probe containing the MLP proximal promoter region (nucleotides -138 to +46). Numbered arrows indicate specific shifts in corresponding lanes. The asterisk indicates a non-specific band due to components endogenous to the lysate.

C. DNA-protein complexes containing hIID (lanes 1-3) and hcore (lanes 4-6) are specific to the TATA box. Unlabelled oligonucleotides (MLP, positions -45 to -25) contained either a wild type (wt) or mutant (mu) TATA box as indicated and were added as competitors. A reaction shown in lane 7 contained equal amounts of hIID- and hcore-programmed lysate.

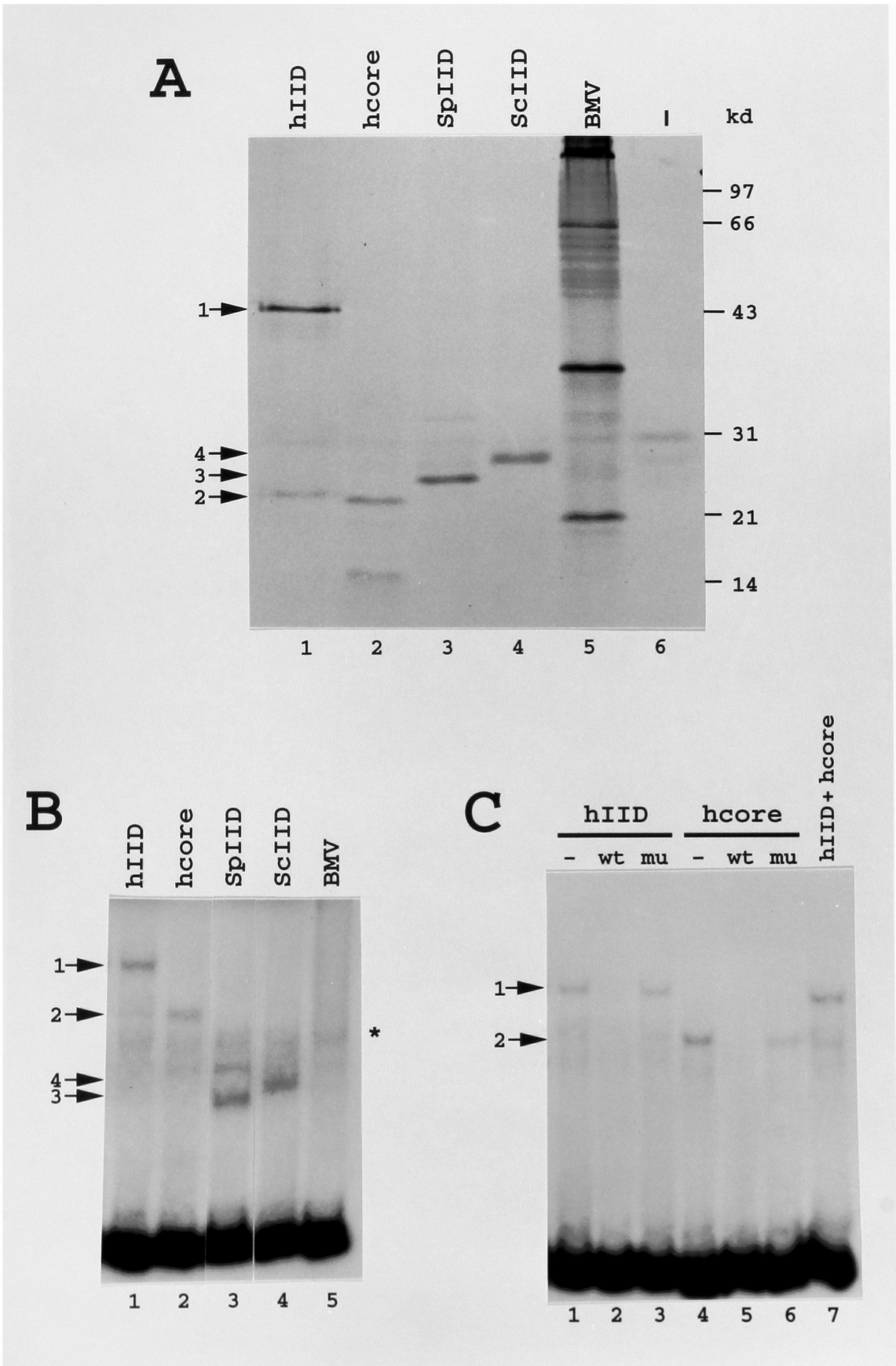
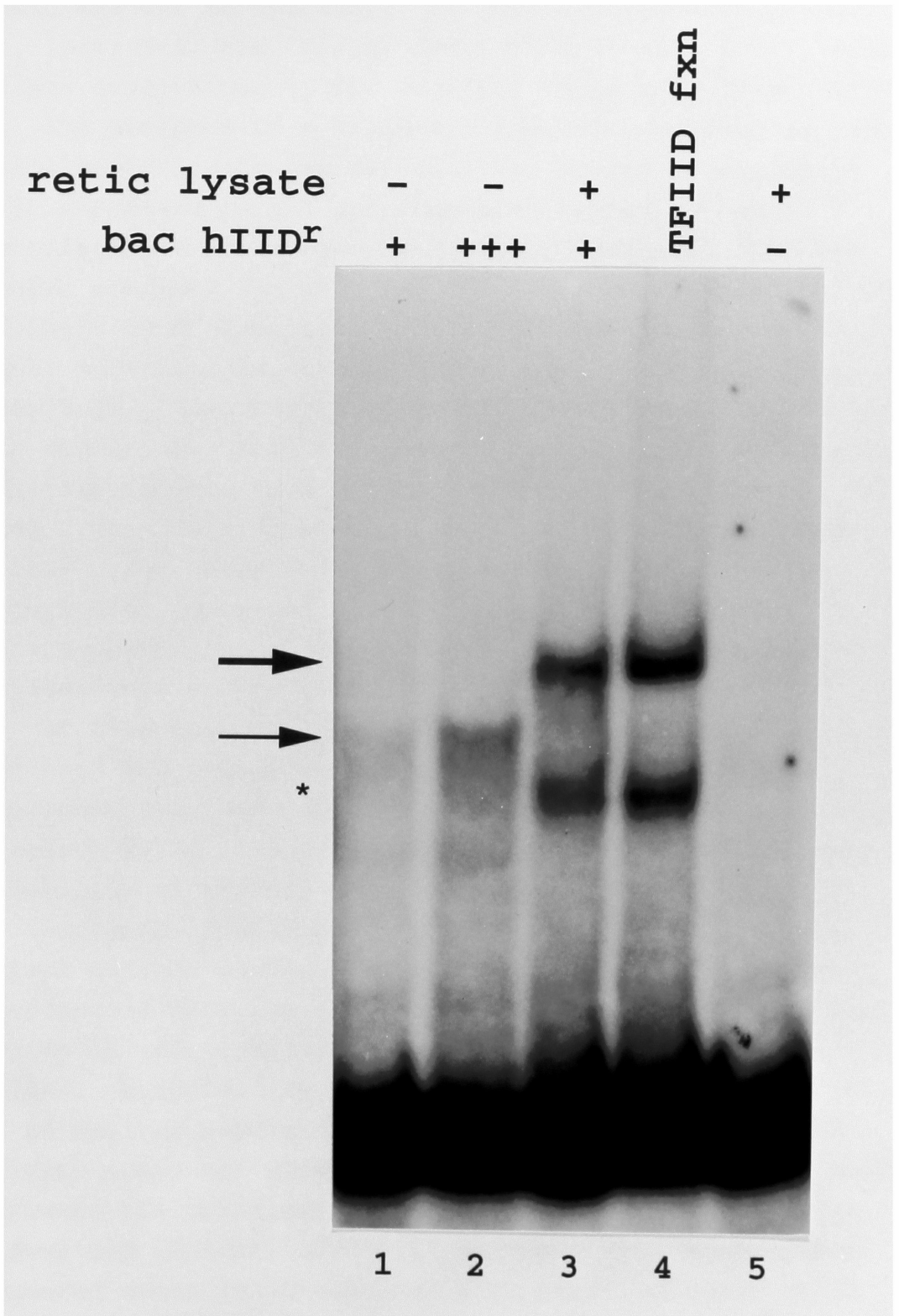

A
hIID
hcore
SpIID
ScIID
BMV
-
kd
97
66
43
31
21
14
1
4
3
2
1 2 3 4 5 6
B
hIID
hcore
SpIID
ScIID
BMV
1
2
4
3
*
1 2 3 4 5
C
hIID
hcore
hIID+hcore
- wt mu - wt mu
1
2
1 2 3 4 5 6 7

Figure 26. Gel mobility shift assay with bacterially expressed human TFIID.

A radio-labelled TATA box-containing probe was incubated with bacterially expressed and purified hIID by itself (lanes 1 and 2) or in combinaion with unprogrammed reticulocyte lysate (lane 3), or with amino octyl fraction of native human TFIID (lane 4), or just with unprogrammed lysate (lane 5). The thick arrow indicates the previously characterized TATA box-specific protein-DNA complex, the thin arrow a novel shift specific to bacterially expressed hIID, and the asterisk a non-specific band.


retic lysate
bac hIID^r
TFIID fxn
- - + +
+ +++ + -
*
1 2 3 4 5

so-called 'species-specificity', as defined by the lethality of the human protein in yeast, reside in the conserved core domain and not in the highly divergent N-terminus (Cormack et al., 1991; Gill and Tjian, 1991; Poon et al., 1991), despite earlier suggestions to the contrary (Pugh and Tjian, 1990).

The presence of a TFIID-interacting component in rabbit reticulocyte lysate was definitively proven by employing TFIID expressed in and purified from bacteria. DNase I footprinting analysis had demonstrated specific TATA box binding activity for this protein (data not shown), but under gel shift conditions used here little DNA-TFIID complex (of higher mobility) is detected after electrophoresis (Fig. 26, lanes 1,2). However, in combination with reticulocyte lysate this bacterially produced protein caused a gel mobility shift indistinguishable from the native TFIID fraction (Fig. 26, lanes 3,4). This effect was sensitive to high salt and sarkosyl, and independent of an incubation at elevated temperatures (data not shown), suggesting that stable protein-protein interactions rather than post-translational modifications are required.

At this point, both functional and physical data indicated the requirement for additional factor(s) to complement the human cDNA-encoded protein for the full range of native TFIID-like activities. Given the relative ease and reliability of the gel shift assay (cf. activation in Fig. 24; (Pugh and Tjian, 1990)), I screened available HeLa nuclear extract-derived column fractions for a TATA box-specific and bacterially produced TFIID-dependent gel shift complex of identical mobility to the native TFIID-DNA complex. Surprisingly, phosphocellulose flow-through (Fig. 27) as well as second column fractions containing TFIIA activity (data not shown), scored positive in such an assay. Following the purification of TFIIA to near homogeneity (DeJong and Roeder, 1993), I confirmed that bacterially expressed human TFIID combined with purified human TFIIA bound to TATA box-containing template in such a complex of a

Figure 27. Rabbit reticulocyte lysate and $P11_{0.1}$ fraction contain a factor that stimulates bacterially expressed human TFIID binding to the TATA box.

A radiolabelled TATA box containing probe was incubated with $P11_{0.1}$ fraction (lanes 1-6), bacterially expressed and purified hIID (lanes 4-11), unprogrammed reticulocyte lysate (lanes 7-9), and/or amino-octyl fraction of native human TFIID. Unlabelled oligonucleotides (MLP, positions -45 to -15) contained either a wild type (wt) or a mutant (mu) TATA box. The arrow indicates a TATA box specific complex; asterisks indicate non-specific complexes.

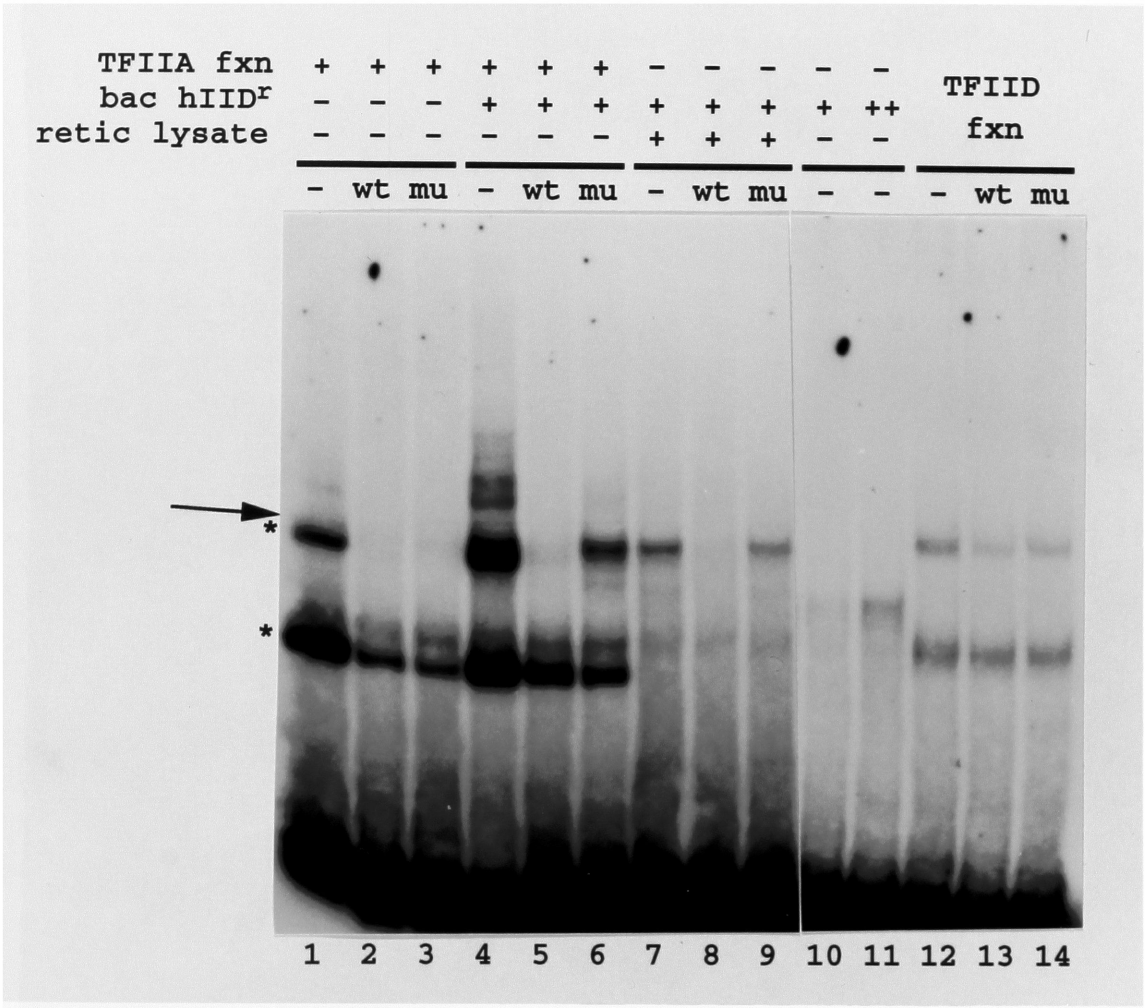

TFIIA fxn + + + + + + - - - - -
bac hIID^r - - - + + + + + + + ++
retic lysate - - - - - - + + + - -
TFIID fxn
- wt mu - wt mu - wt mu - - - wt mu
*
*
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 28. Bacterially expressed human TFIID combined with purified human TFIIA form complexes with TATA box-containing DNA fragments of the same mobility as partially purified human TFIID.

A radio-labelled TATA box containing MLP-derived DNA fragment was incubated with bacterially expressed and purified IID from S. cerevisiae (lanes 2 and 3) or human (lanes 4 and 5), alone (lanes 2 and 5) or in combination with a highly purified TFIIA fraction (DEAE-5PW, lanes 3 and 4). The effect of TFIIA by itself and an amino-octyl fraction of native human TFIID are shown in lanes 1 and 6, respectively. Arrows indicate TATA box-specific complexes, asterisks indicate non-specific bands.

TFIIA + - + + -
bac IID^r - Sc Sc h h
TFIID fxn
*
*
1 2 3 4 5 6

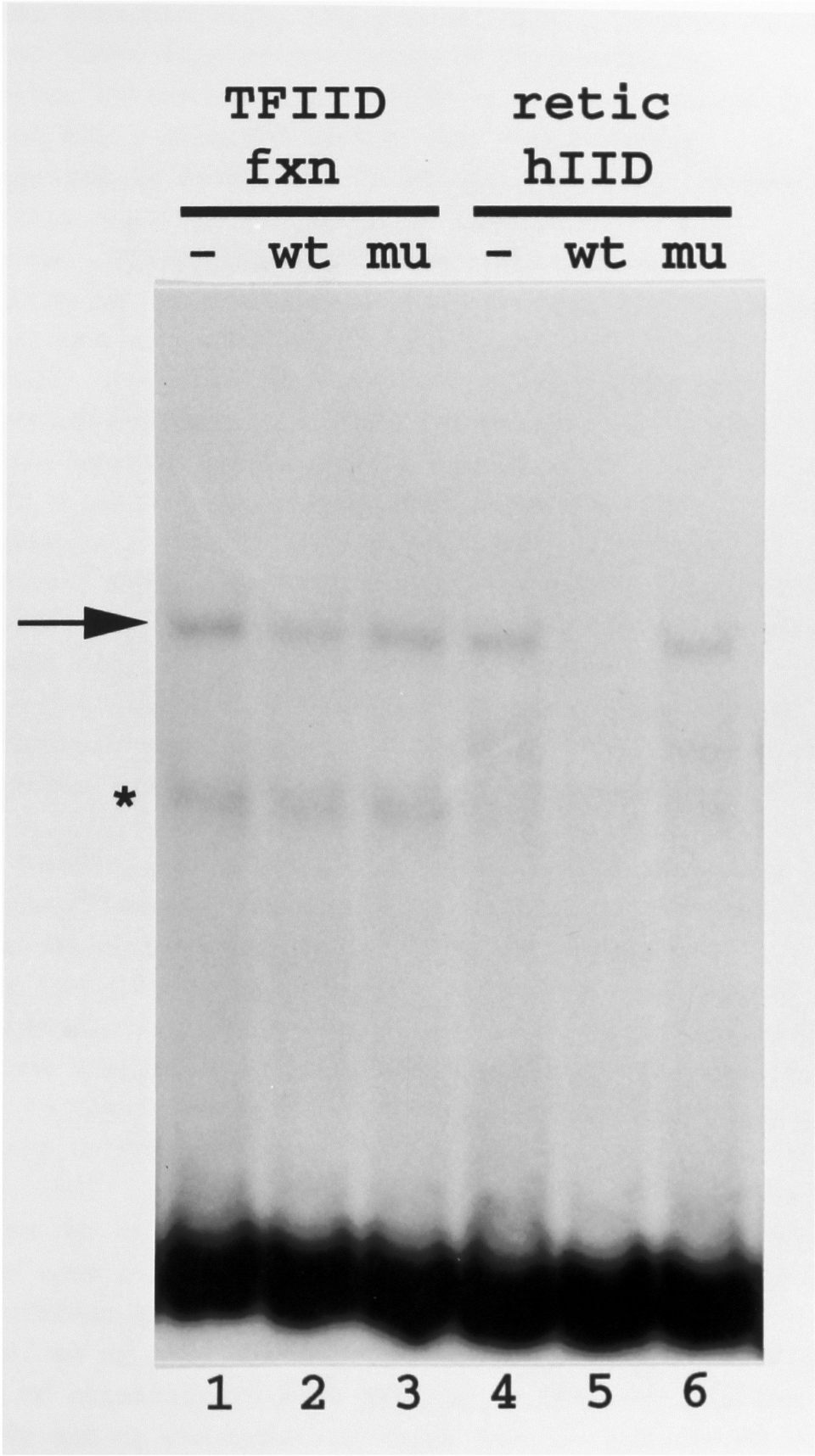
mobility indistinguishable from the native TFIID-DNA complex (Fig. 28), previously described (H. Fujii and M. Horikoshi, personal communication; (Hoffmann et al., 1990)).

This finding was apparently paradoxical to the observed activity of the native TFIID preparation in mediating upstream activator effects, while recombinant human TFIID combined with TFIIA does not have such an effect. However, it has been known for some time that TFIIA enhances remarkably, and possibly modulates, the TATA-box binding activity of TFIID (Fire et al., 1984; Reinberg et al., 1987; Reinberg and Roeder, 1987; Lee et al., 1992). Furthermore, TFIIA interacts with TFIID not only on the DNA but also in solution, providing the basis for the first successful purification of this factor (DeJong and Roeder, 1993) and indeed is observed to be present in the TFIID fraction (J. deJong, personal communication; (Yokomori et al., 1993a)). Thus it is conceivable that the partially purified TFIID preparation contains multiple populations of p38 TFIID, some associated with TFIIA, giving rise to the prominent gelshift complex. Other populations, however, may be responsible for mediating the *trans*-activation effect and may not bind DNA in such a manner that the electrophoresis conditions employed will allow detection of the corresponding complexes.

Such an interpretation is consistent with the great variation in estimates for the native size of the protein complex responsible for TFIID activity. While recently cited estimates based on glycerol gradient sedimentation put the native size at 750kd (Conaway and Conaway, 1990; Conaway et al., 1991), earlier size exclusion chromatography experiments indicated a size of 130kd (Samuels et al., 1982; Reinberg et al., 1987). Though the native size of TFIIA or a p38 TFIID-TFIIA complex has not been accurately estimated, it might be relevant to recall that human TFIIA subunit composition is 35kd, 19kd, and 12kd (DeJong and Roeder, 1993): Assuming the

Figure 29. Comparison of oligonucleotide competition behaviour of native TFIID with *in vitro* expressed TFIID.

A radio-labelled TATA box-containing, MLP-derived (positions -138 to +46) DNA fragment was incubated with amino-octyl fraction of native TFIID (lanes 1-3) or reticulocyte lysate programmed with hIID RNA (lanes 4-6). Unlabelled oligonucleotides (MLP, positions -45 to -15) contained either a wild type (wt) or a mutant (mu) TATA box. The arrow indicates the TATA box specific complex, the asterisk a non-specific band.


TFIID fxn
retic hIID
- wt mu - wt mu
*
1 2 3 4 5 6

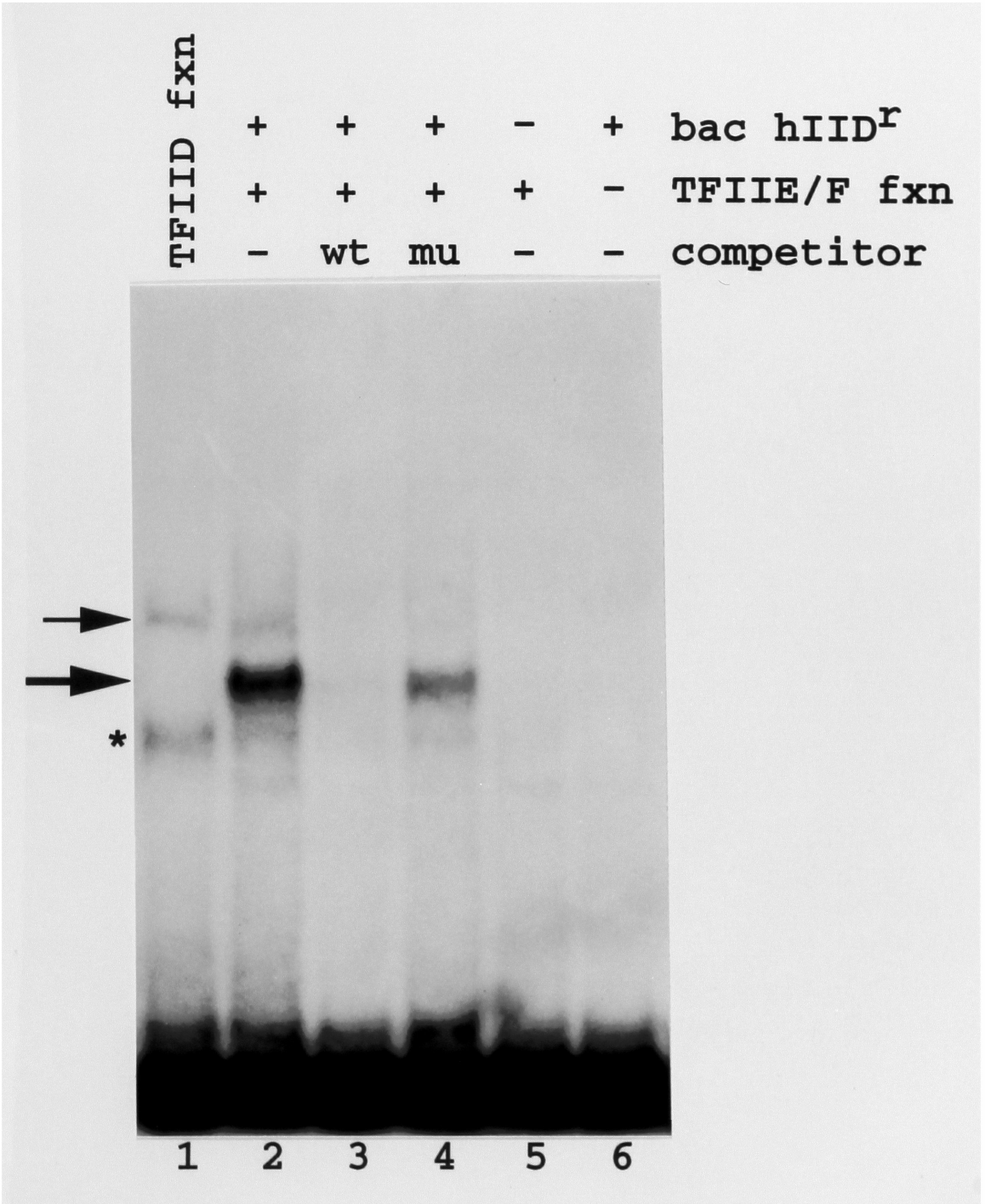
simplest stoichiometry, the size of such a complex would be close to those earlier estimates of native TFIID.

Other assays can distinguish the native, partially purified TFIID fraction is not only functionally distinguishable from a p38 TFIID-TFIIA complex. DNase I protection analysis on the major late promoter with recombinant TFIID only exhibits a specific footprint restricted to the TATA box region, whereas at least some preparations of native TFIID have given rise to the previously described extended footprints. Similarly, careful competition analysis with TATA box-containing 30 base pair oligonucleotides demonstrates a differential stability of the two gel shift complexes (Fig. 29) indicating that they are not equivalent despite indistinguishable mobility: while the recombinant TFIID-TFIIA-DNA complex is easily competed with a ten-fold excess of short wild-type competitor, the native TFIID-DNA complex is more resistant, suggestive of additional protein-DNA contacts outside the 30 base region covered by the oligonucleotide competitor (seen in footprint analysis), that have a stabilizing effect on this protein-DNA complex.

An additional protein that specifically interacts with recombinant human TFIID was found in HeLa nuclear extract derived fractions (Fig. 30), though the mobility of the protein DNA complex was markedly different from that of native TFIID. Further analysis with more highly purified fractions (unpublished data) indicated that this particular factor is likely to be NC2/Dr1 (Inostroza et al., 1992), a component of the co-activator fraction USA (Meisterernst and Roeder, 1991). Interestingly, this factor, which locks TFIID onto the DNA in a non-productive complex, competes with TFIIA for the same protein interaction surface of recombinant TFIID (Meisterernst and Roeder, 1991; Inostroza et al., 1992). However, as my goal has been to characterize native TFIID, I turned my attention to more strategies that would allow me to identify and to characterize those factors associated with

Figure 30. The $P11_{0.5}$ fraction contains another TFIID-interacting factor.

A radio-labelled TATA box-containing, MLP-derived DNA fragment was incubated with amino-octyl native human TFIID (lane) or with bacterially expressed and purified hIID in combination with a $P11_{0.5}$ fraction (lanes 2-4) or alone (lane 6). The effect of the $P11_{0.5}$ fraction itself is shown in lane 5. Wild-type (wt) and mutant (mu) TATA box-containing competitors are included in lanes 3 and 4, respectively. Arrows indicate TATA box-specific complexes, the asterisk a non-specific band.


TFIID fxn
+ + + − + bac hIID^{r}
+ + + + − TFIIE/F fxn
− wt mu − − competitor
*
1 2 3 4 5 6

the recombinant TATA box binding protein that function together to account for the classically defined TFIID activity.

Chapter II: Identifying other subunits of TFIID

Introduction

Significant differences in both functional and physical assays (presented in the previous chapter) between the human cloned protein (p38) and the previously characterized TFIID activity present in HeLa-derived chromatographic fractions, led to the hypothesis that TFIID is a multi-protein complex. According to this model, the cDNA-encoded polypeptide represents only one of several subunits of a multi-protein complex that as a whole is responsible for the physical and functional characteristics classically attributed to TFIID. Figure 31 is intended to summarize some of the important characteristics and criteria discussed in more detail in previous chapters.

The cloned protein (p38) is capable of carrying out key functions by itself: Most notably, it specifically recognizes and binds the TATA box to form a DNA-protein complex that is resistent to nucleosome assembly and represents the molecular basis of the functionally observed 'template commitment'. It is for this reason that this polypeptide was called "TFIIDτ", as the TATA box binding subunit of TFIID, or more popularly "TBP" for TATA binding protein (Beware of the potential confusion with Tat-binding protein, TBP-1 (Ohana et al., 1993)). Secondly, it is evidently capable of nucleating preinitiation complex formation once bound to the promoter region by recruiting TFIIB (van Dyke et al., 1988; Buratowski et al., 1989). This event in turn leads to the assembly of the other general transcription factors and results in the observation that TFIID subunits other than TBP are not required for basal level transcription from TATA box-containing promoters *in vitro* (e.g. Fig. 23, 24) or *in vivo* (Colgan and Manley, 1992). However, significant differences between TBP and TFIID (for a summary, see Figure 31) are observed with regards to activator responsive transcription,

Figure 31. Contrasting native TFIID with TBP (TFIIDτ).
Summary of some key characteristics of native human TFIID and the TATA box-binding protein encoded by the cDNAs described in chapter I. See text for further details.

Contrasting	native TFIID	and	TBP/TFIIDτ
1. basal transcription	✓		✓
2. activated transcription	✓		–
3. TATA-box binding	✓		✓
oligo competition	weak		strong
footprint	-47 to +35		-35 to -18
4. template commitment	✓		✓
5. molecular weight	≈750,000		38,000

native molecular mass, and footprinting characteristics when bound to the DNA, as discussed in the previous chapter.

Despite TBP's remarkable functional capabilities as a single protein *in vitro*, it is now generally believed that it exclusively functions *in vivo* within a multi-protein complex. Comparative studies with recombinant (bacterially derived) and native (HeLa derived) fractions have shown that the chromatographic characteristics of TBP and TFIID are distinguishable (e.g. on phosphocellulose, data not shown). On this basis no evidence can be found for uncomplexed TBP present in nuclear extract (data not shown). This scenario is consistent with the widely held notion that basal transcription is not a physiologically relevant phenomenon but that it is exclusively observed *in vitro* on naked DNA templates. Consequently, it is now generally accepted that *in vitro* or *in vivo* studies utilizing uncomplexed TBP are intrinsically prone to artefactual conclusions.

A fundamental reappraisal of TBP's functional roles in transcription was triggered by the highly significant observation that TBP is a subunit of the class I specific transcription factor SL1/TIF-1B (Comai et al., 1992) and that TATA box-containing oligonucleotide competitors would inhibit transcription by RNA polymerase III (White et al., 1992). Soon after, yeast *in vivo* studies with class-specific mutants demonstrated that TBP plays an essential role in transcription by all three RNA polymerases (Cormack and Struhl, 1992; Schultz et al., 1992). Thus TBP turned out to be a more universal factor than originally thought. In retrospect, it is apparent that the reasons for TBP's original identification as a class II-specific general transcription factor lie in the identification of a less well characterized yeast IID fraction by substitution for the mammalian TFIID fraction. The resulting confusion was further complicated by the fact that TBP itself has functional activity *in vitro* only in conjunction with RNA polymerase II, whereas it has an essential functional role as

a component in one of presumably three multi-protein complexes in the transcription of all three classes of genes. So-called "TBP-associated factors (TAFs)", are in turn thought to impart the class-specific function to each of the three TBP-containing multi-protein complexes. TBP then is a common component in the initiation complex on every promoter, reminiscent of certain RNA polymerase subunits that are common to all three classes of the enzyme (e.g. ABC27, ABC23, ABC14.5, ABC10a, ABC10b, (Sentenac, 1985)).

Despite TBP's universal functional role in eukaryotic transcription, my focus in this chapter remains on the molecular characterization of TFIID, and more specifically, the identification of TBP-associated factors within the TFIID complex. To this end I pursued essentially four strategies more or less concurrently, the latter ones within this chapter being more suitable with hindsight, given the particular biochemical characteristics that the TFIID complex and its constituent proteins turned out to have. First, I will describe attempts to use TBP-affinity chromatography and *in vitro* reconstitution of the TFIID complex to identify TBP-associated factors. Immuno-affinity purification with α-TBP antibody subsequently led to the successful identification of at least some of the subunits of TFIID (in collaboration with Ritsuko Takada) and $TFIIIB_2$ (with Zhengxin Wang). In an effort to obtain purified TFIID in non-denaturing conditions suitable for functional analysis, I first attempted to utilize the vaccinia-aided expression of a tagged TBP suitable for convenient affinity chromatography and then established cell-lines constitutively expressing epitope-tagged TBP to purify functional complexes. This strategy led ultimately to the most comprehensive identification of class II and class III TAFs to date (in collaboration with Cheng-ming Chiang).

(i) Attempts at TBP-affinity chromatography and reconstitution of TFIID *in vitro*

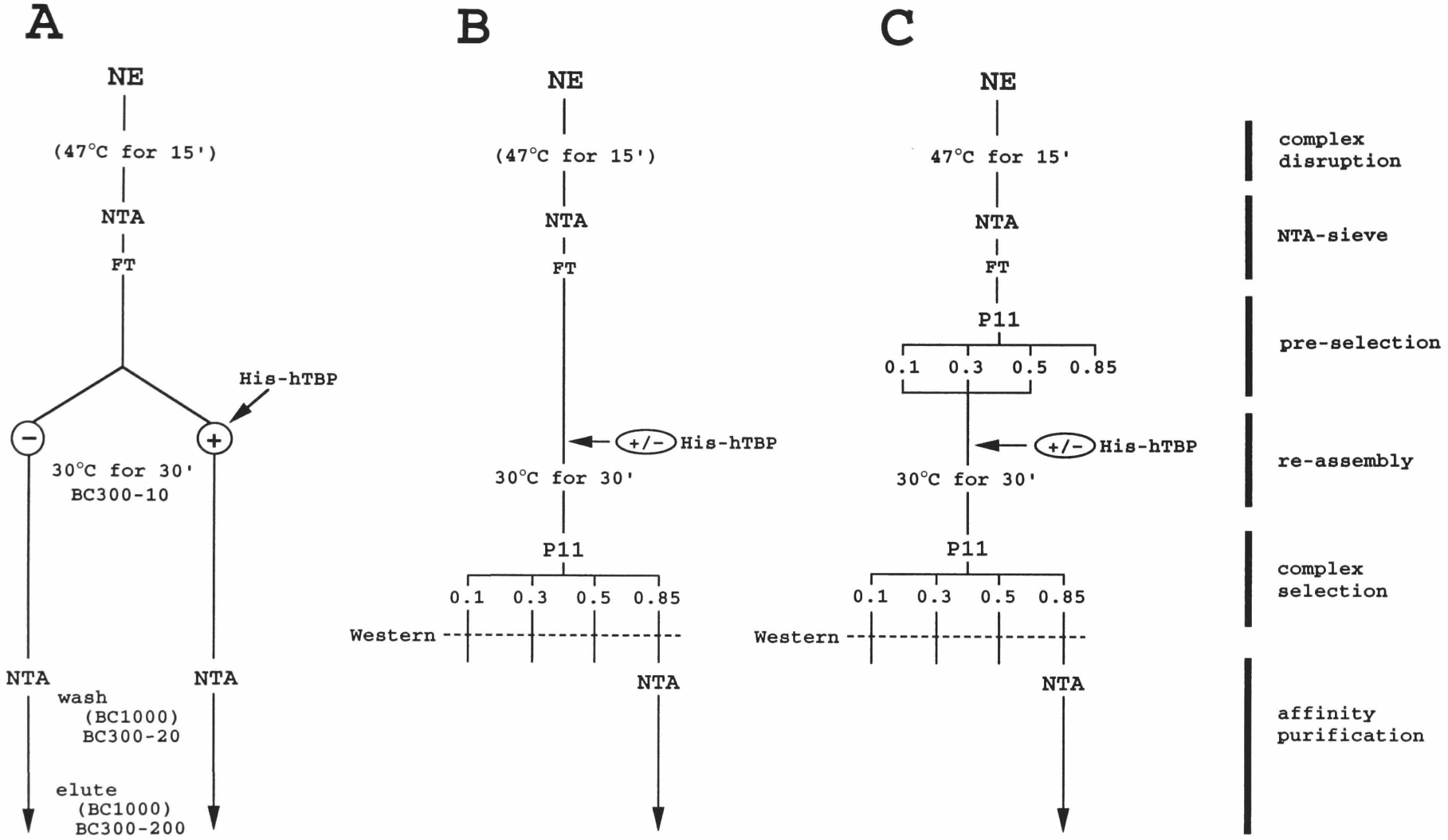
Strategies

Protein affinity chromatography has been successfully used to identify and isolate factors functional in transcription, initially from *Escherichia coli* (e.g. elongation factors NusA, NusG, (Greenblatt and Li, 1981)) and then from mammalian extracts (e.g. TFIIF, TFIIS (Sopta et al., 1985)). Classically, the procedure involves a highly purified protein ligand immobilized in large amounts on an agarose support which is then used to bind specifically proteins present in an extract or partially purified fraction. Recent technical innovations in molecular biology allowed several modifications to be made to such a protocol (provided that at least one of the proteins involved is cloned) which may increase the likelihood of recovering specifically bound factors, and assaying their function. Most importantly, the protein ligand, classically cross-linked irreversibly to the support in random orientation, can now be reversibly bound in a defined orientation with the help of an affinity tag (e.g. GST; His, chapter III; epitope tags, section iv). These modifications allow the interaction to be formed in solution, may avoid potential problems resulting from steric interference on the support, and allow for the elution of the resulting protein complex in non-denaturing conditions for functional assays.

Such modifications of the classical protocols of protein affinity chromatography seemed particularly suited for the identification of TBP-associated factors in the TFIID complex. First, the observed lability of TBP with regard to even moderate temperatures indicates that irreversible cross-linking to a support may result in denaturation of this protein. Second, given a native size estimate of 750kd, TFIID seemed likely to consist of several polypeptides some of which might not directly interact with TBP. A protocol

Figure 32. Schemes for *in vitro* reconstitution of the TFIID complex and subsequent affinity purification.

Three different schemes shown in panels A, B, and C are variations of a basic protocol for TBP-affinity purification. See text for further details.


A
NE
(47°C for 15')
NTA
FT
His-hTBP
−
+
30°C for 30'
BC300-10
NTA
NTA
wash
(BC1000)
BC300-20
elute
(BC1000)
BC300-200
B
NE
(47°C for 15')
NTA
FT
+/-
His-hTBP
30°C for 30'
P11
0.1
0.3
0.5
0.85
Western
NTA
C
NE
47°C for 15'
NTA
FT
P11
0.1
0.3
0.5
0.85
+/-
His-hTBP
30°C for 30'
P11
0.1
0.3
0.5
0.85
Western
NTA
complex disruption
NTA-sieve
pre-selection
re-assembly
complex selection
affinity purification

allowing for the assembly of the multi-protein complex in solution - in effect simply replacing the endogenous TBP with a tagged recombinant version *in vitro* - seemed to hold more promise than the expectation to successfully assemble TFIID on a support-bound TBP. In addition, such *in vitro* reconstituted TFIID can be selected for with the use of a conventional chromatographic step (e.g. phosphocellulose) before retrieving it with the tag-mediated affinity step.

Figure 32 represents a summary of a number of similar protocols - based on the considerations discussed above - that I developed to identify constituents of the TFIID complex. All protocols make use of a histidine-tagged human TBP (His-hTBP) expressed in and purified (≥98% purity) from *Escherichia coli* that can be repeatedly bound to and eluted from a Ni^{2+}-chelating resin (NTA) in non-denaturing conditions (see chapter III). The simplest scheme (Fig. 32A) involves the incubation of His-hTBP in nuclear extract and retrieval of the protein and associated factors with an NTA column. Comparison with the eluate from a control column assures specificity. Two potential problems with this protocol can be avoided with the following modifications : First, despite its high selectivity for recombinant proteins in bacterial extract, there are a significant number polypeptides present in human extract that bind to the NTA-column even at reasonably high stringency; this observation reflects the large number of *opa* repeat-like sequences found in mRNAs from *Drosophila* and mammalian species that may code hor histidine. An NTA column used on nuclear extract before the addition of His-hTBP will act as a 'sieve' (Fig. 32) to remove endogenous NTA-binding proteins and improve the specificity of the latter NTA step. The second modification to the protocol relates to the *in vitro* reassembly of the TFIID complex and was suggested by functional experiments which demonstrated that heat treated nuclear extract supplemented with recombinant TBP is capable of mediating the effect of exogenously added activators (Peterson et al., 1990; Kelleher

III et al., 1992), (R. Bernstein, personal communication). Thus mild heat treatment, which preferentially denatures TBP, was a convenient way to trigger the disruption of the highly stable TFIID complex and allow the reassembly of TAFs with a tagged form of TBP into a functional TFIID complex.

Attempts to submit imidazole-eluted complexes to functional assays were thwarted by the enormous excess of free His-hTBP that in any assay (gelshift or *trans*-activated transcription) would obscure the presumed TFIID activity. This situation could be improved by employing a phosphocellulose (P11) column following the re-assembly step with His-hTBP in order to separate the TFIID complex from free TBP or presumed $TFIIIB_2$ in a so-called "complex selection" step (Fig. 32B). A Western blot of the P11 fractions was used to monitor His-hTBP fractionation. A further variation on this strategy involved a similar P11 column after complex disruption by heat-treatment and before the complex-re-assembly step with His-hTBP (Fig. 32C). This modification held the promise of improving the specificity of the recovered proteins, was, however, based on the assumption that none of the necessary constituents of TFIID chromatograph in the high salt P11 fraction as single, uncomplexed polypeptides.

Discussion

The protein-affinity chromatography approach described above requires an efficient method for complex disruption and reassembly. Heat-treatment of nuclear extract seemed to provide a convenient means to disrupt the complex, but in fact it lacked the efficiency and reliability required, depite initial indications to the contrary from functional assays (Peterson et al., 1990; Kelleher III et al., 1992), (R. Bernstein, personal communication). This weakness of the approach was reflected in the irreproducibility of the silver stain patterns in protein gels and was also demonstrated in

Western blot and transcription assays (data not shown). Given the large set of proteins, that recognize as TFIID subunits today, that engage in multiple interactions of an uncharacterized chemical nature, reconstitution of the TFIID complex *in vitro* was indeed the major difficulty and has remained so to date. Another drawback of the strategy was limited specificity of the Ni^{2+}-chelating resin when used with mammalian extract. This was clearly improved by the 'sieving' step, before addition of His-hTBP, as well as the pre-selection step in scheme C. However, these modifications not only carried the risk that essential components in the TFIID assembly would be lost, but also complicated the experimental protocols to such an extent that technical reproducibility over so many steps became even more difficult to achieve.

Reports (Dynlacht et al., 1991; Tanese et al., 1991) published towards the end of my commitment to this strategy demonstrated a partial disruption and functional reconstitution of the TFIID complex. The so-called "TAF fraction", generated by urea or heat denaturation of TFIID bound on a chromatographic support was not, however, capable of TFIID-like functions when complemented with recombinant hTBP, but, indeed, required the endogenous TBP-containing fraction. This suggests that additional TBP-bound co-factors and/or post-translational modifications of TBP are essential for TFIID complex formation, that are not present in recombinant TBP preparations from bacteria.

In summary, and with hindsight, the tightness of the TFIID complex rule out protein affinity chromatography with TBP as a suitable method for purifying the TFIID complex. Furthermore, TBP has turned out to be a particularly 'sticky' molecule with regards to interactions with a whole host of transcription factors (see chapter III, Fig. 54), which in this experimental approach manifested itself in seemingly specifically co-purifying polypeptides that do not possess TFIID-like functions. Coupled to a Western blot assay for

known proteins, protein affinity chromatography with any of the general transcription factors, including TBP, remains a potentially fruitful approach to detect specific protein-protein interactions.

(ii) Immunopurification of TBP-containing complexes

The identification of TFIID subunits was largely accomplished and published by Takada et al. (1992), whereas the work described in the latter part of this section is the result of an equal collaborative effort between Zhengxin Wang and myself.

Immunopurification of TFIID

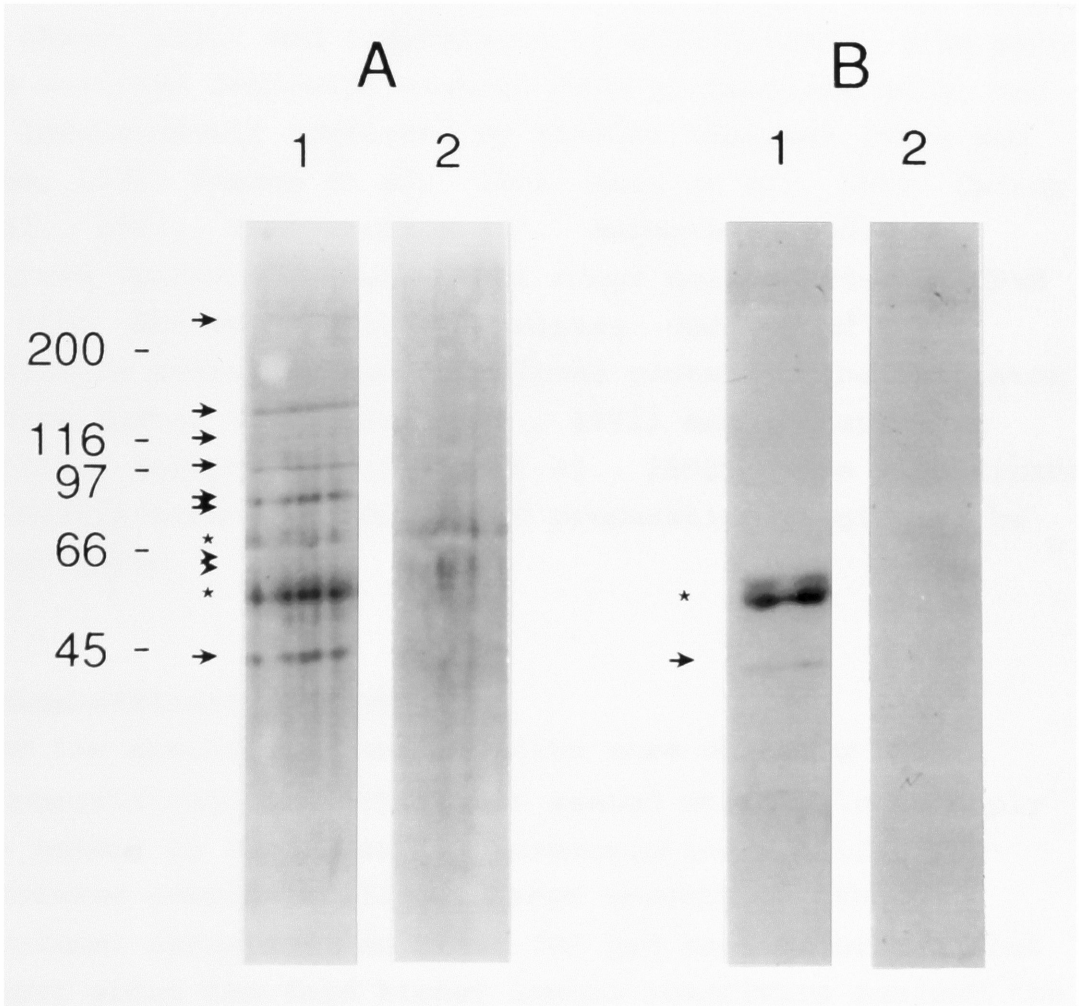
The method of choice for the identification of polypeptides subunits of a multi-protein complex is immunopurification with antibodies directed against one of its constituents whose antigenic surface is also accessible to antibodies when present in the complex. This approach has been successful for the identification of stable complex constituents as well as for the detection of reversible protein-protein interactions. The apparent stability of the protein-protein interactions within the TFIID complex held the promise of a highly efficient purification method. The N-terminus of TBP exhibits a higher level of antigenicity than the highly conserved TBP core domain (data not shown) and is also the target of choice for an immuno-purification, as the demonstration of its redundancy with respect to yeast cell viability (Cormack et al., 1991; Gill and Tjian, 1991; Poon et al., 1991; Reddy and Hahn, 1991; Zhou et al., 1991) suggests that important regulatory proteins like TAFs are likely to interact with the core domain leaving the N-terminus freely accessible to antibody binding.

For the purposes of immuno-purification of TFIID antibodies against a TBP N-terminal peptide (residues 12-29) conjugated to bovine serum albumin were raised on a large scale in rabbits, subsequently antigen purified and crosslinked to Sepharose. Immunopurification of TBP containing complexes present in the phosphocellulose high salt fraction containing classically defined TFIID activity resulted in the detection of several polypeptides

Figure 33. Immunoprecipitation of TFIID with α-hTBP antibodies.

B. Silver stain analysis of polypeptides bound from α-TBP-specific antibody (lane 1) or control (lane 2) resins. The lower arrow indicates TBP; upper arrows indicate other polypeptides specifically and reproducibly retained on the specific antibody resin. The arrowheads indicate minor (65 and 60kd) specifically bound polypeptides that were more apparent in other analyses. The stoichiometries of the specifically bound polypeptides indicated here were more equivalent when monitored by sequence analysis following elution from preparative gels. Asterisks indicate immunoglobulin heavy chain (lower) and non-specific (upper) polypeptides. Molecular masses are indicated in kDa.
B. Immunoblot analysis of TBP in bound polypeptides from specific antibody (lane 1) and control (lane 2) resins.

(This experiment was carried out by R. Takada.)


A
B
1
2
1
2
200
116
97
66
45

(approximate molecular masses of 250, 150, 120, 100, 80, 75, 65, 60) in addition to TBP that were not present in the eluate of a control column (Fig. 33A). That the 42kd polypeptide represents TBP was confirmed by Western blot analysis (Fig. 33B). The specificity of the co-purified polypeptides was further indicated by their tight association with TBP, by previous molecular mass estimates (Reinberg et al., 1987; Conaway et al., 1991; Tanese et al., 1991; Timmers and Sharp, 1991) and indications of stoichiometry from amino-acid analyses following elution from preparative gels, and was independently confirmed by similar analyses (Pugh and Tjian, 1991; Tanese et al., 1991; Zhou et al., 1992; Chiang et al., 1993), (see section iv). Amino-acid sequence analysis further indicated that minor polypeptides of 65kd and 60kd, not seen in other analyses, correspond to previously characterized and cloned proteins, the initiator binding factor YY1 (Seto et al., 1991) and the splicing auxiliary factor U2AF (Zhang et al., 1992). The significance of their presence in this TFIID preparation remains to be investigated.

Immunopurification of TFIIIB$_2$

Given the specificity and relative ease of the α-TBP immunopurification of TFIID, it seemed worthwhile to apply this method to the class III transcription-specific TBP-containing complex TFIIIB$_2$. Large amounts of rabbit polyclonal serum against human TBP had been generated that exhibit about ten-fold higher immuno-reactivity against the N-terminus than the 'core domain' of TBP. Following partial purification of IgG by conventional means, TBP-specific antibody was purified on an antigen column by elution with either high ionic strength (3.5M $MgCl_2$) or low pH (0.1M glycine pH2.5) in the hope of preparing an antibody column that could be eluable in conditions that allow recovery of functional TBP-containing complexes. Phosphocellulose

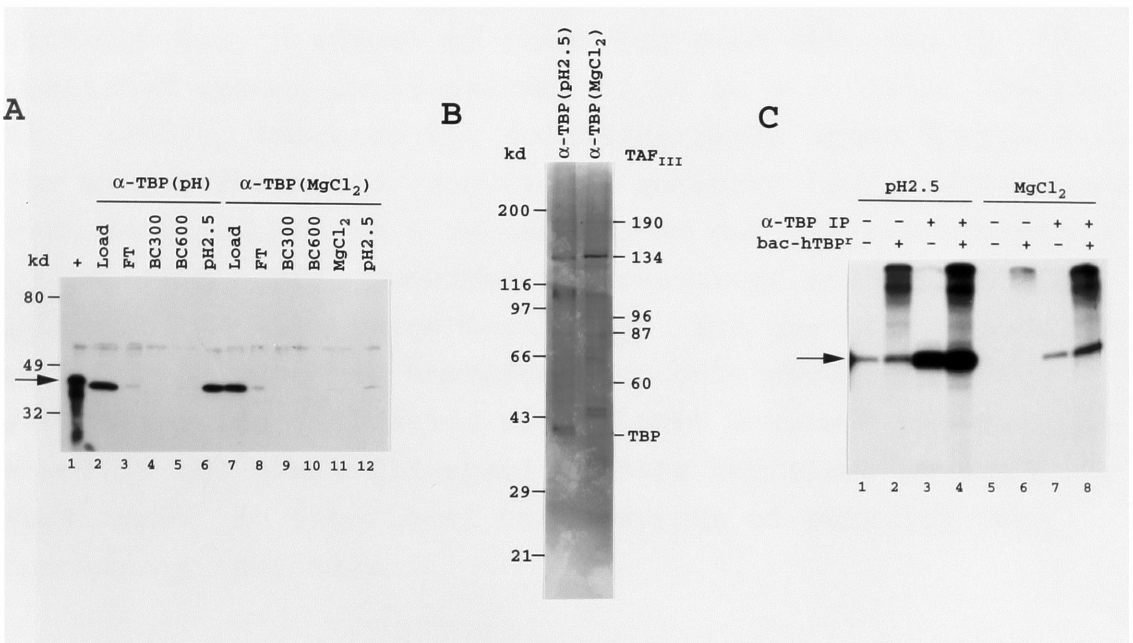
Figure 34. Immunoprecipitation of TFIIIB$_2$ with α-hTBP antibodies.

A. Western blot with α-TBP serum monitoring binding and elution efficiencies of antibodies eluted from a TBP-column by low pH (lanes 2-6) or by 3.5M $MgCl_2$ (lanes 7-12). Lane 1, bacterially expressed TBP; lanes 2 and 7, second column TFIIIB fraction; lanes 3 and 8, flow-through fraction of the antibody columns; lanes 4, 5, 9, and 10, wash fractions with indicated mM KCl concentrations; lanes 6, 11, and 12, eluate fractions with indicated elution buffers. The arrow indicates human TBP.

B. Silver stain analysis of polypeptides eluted from the two antibody columns with low pH elution buffers. The molecular weight of potential class III TAFs predicted from this and other analyses (Chiang et al., 1993) are indicated in kDa.

C. Transcription analysis of immunopurified TFIIIB$_2$. A class III transcription system consisting of second column fractions was complemented buffer (lanes 1 and 5), bacterially expressed TBP (lanes 2, 4, 6, and 8) and/or low pH eluates (lanes 3, 4, 7, and 8) from indicated antibody columns. The arrow indicates accurately initiated transcripts.

(These experiments were carried out in collaboration with Z. Wang.)


A
α-TBP(pH)
α-TBP(MgCl2)
kd
+
Load
FT
BC300
BC600
pH2.5
Load
FT
BC300
BC600
MgCl2
pH2.5
80
49
32
1 2 3 4 5 6 7 8 9 10 11 12
B
kd
α-TBP(pH2.5)
α-TBP(MgCl2)
TAFIII
200
116
97
66
43
29
21
190
134
96
87
60
TBP
C
pH2.5
MgCl2
α-TBP IP
bac-hTBPr
1 2 3 4 5 6 7 8

$P11_{0.32M}$ that contains TFIIIB activity was loaded onto each column and TBP was quantitatively bound and then recovered using low pH elution conditions but not with high ionic strength as indicated by α-TBP Western analysis (Fig. 34A). A number of high molecular weight polypeptides are, however, detected in either eluted fraction as detected by silver stained SDS-PAGE (Fig. 34B), suggesting that 3.5M $MgCl_2$ strips class III-specific TAFs from the bound TBP molecule. Identified polypeptides (p190, p135, p96, p87, p60) were also present in conventionally purified $TFIIIB_2$ (Z. Wang, personal communication) and $TFIIIB_2$ purified with the aid of a tagged hTBP-expressing cell line (section iv) (Chiang et al., 1993), although only a subset of these had been detected in published immuno-purifications (Lobo et al., 1992; Taggart et al., 1992). Importantly, following small scale dialysis of the eluate fractions, functional activity could be recovered from both columns in a reconstituted pol III transcription system (Fig. 34C), provided the reactions are adequately supplemented with recombinant TBP. The use of TBP-specific antibody columns has therefore not only resulted in the successful identification of at least a subset of class II-specific and class III-specific TAFs but also, at least in some cases, in functional preparations of purified TBP-containing complexes.

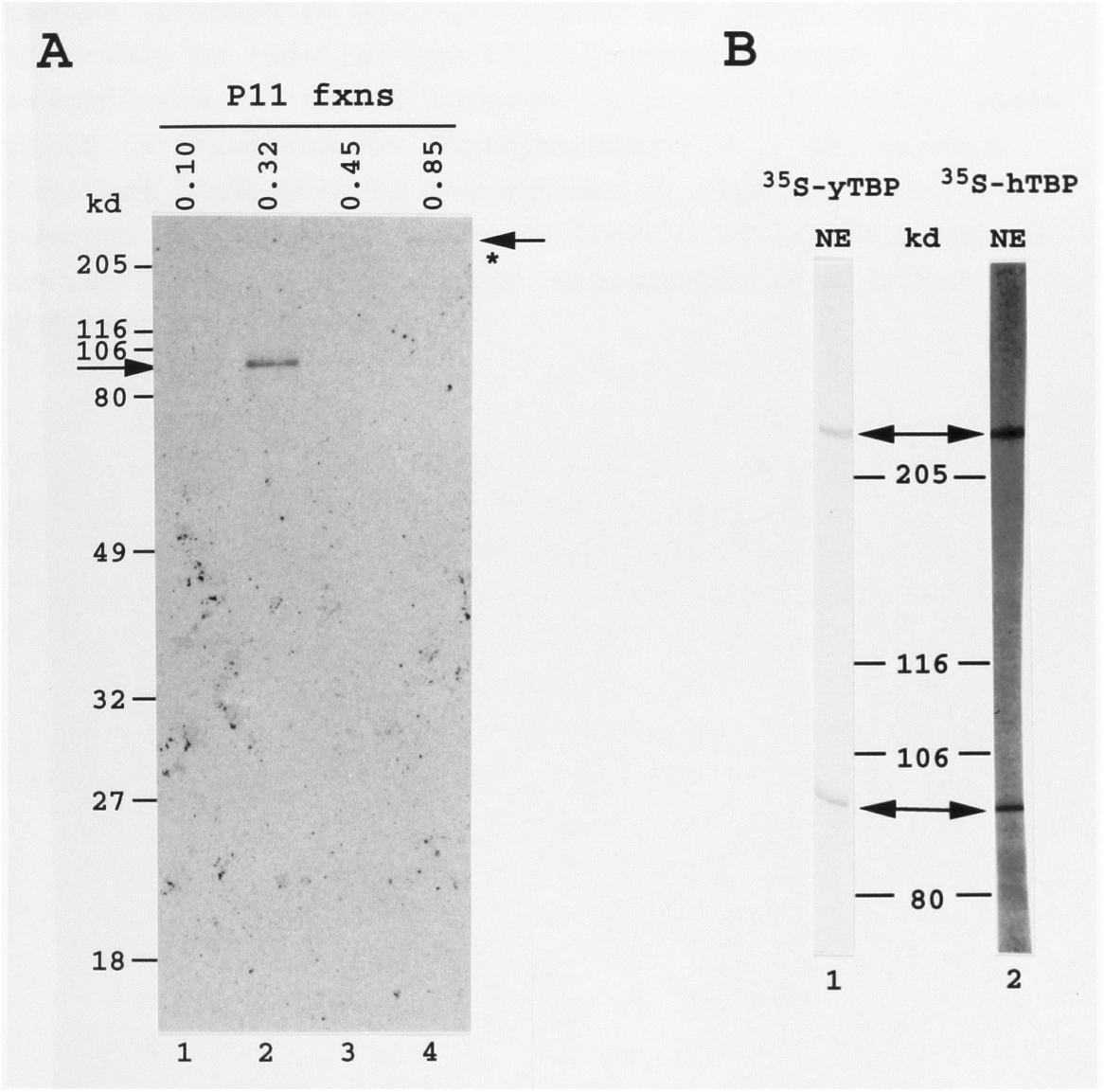
Identification of direct TBP-TAF interactions

Immunopurification of the TFIID complex cannot be used to distinguish between directly and indirectly TBP-interacting TAFs. One way of identifying direct interactions is the Far Western blot which involves probing partially renatured proteins immobilized on a nitrocellulose membrane following size separation via SDS-PAGE with a radio-labelled protein probe. When P11 fractions are probed with ^{35}S-labelled hTBP, two prominent polypeptides are detected: a very large protein of ≈250kd in the high salt fraction, and a 96kd polypeptide

Figure 35. Far Western of HeLa-derived fractions probed with TBP.

A. Far Western analysis with indicated HeLa-derived phosphocellulose (P11) fractions probed with *in vitro* produced and purified (^{35}S)labelled human TBP. Arrows indicate specific signals, the asterisk a weaker, possibly specific signal in the $P11_{0.85}$ fraction.

B. Far Western analysis with HeLa nuclear extract probed with *in vitro* produced and purified (^{35}S)labelled *S. cerevisiae* (lane 1) and human (lane 2) TBP. Arrows indicate specific TBP-binding polypeptides.


A
P11 fxns
0.10
0.32
0.45
0.85
kd
205
116
106
80
49
32
27
18
*
1
2
3
4
B
35S-yTBP
35S-hTBP
NE
kd
NE
205
116
106
80
1
2

in a 0.32M fraction (Fig. 35A). Further analyses with purified preparations of TFIID and $TFIIIB_2$ have confirmed that these correspond to $hTAF_{II}250$ (Takada et al., 1992) and $hTAF_{III}96$, a constituent of $TFIIIB_2$ (Fig. 34B, Z. Wang, personal communication). In light of previous reports of functional differences between the yeast and human TBP core domains (Cormack et al., 1991; Gill and Tjian, 1991), presumably to 'species-specific' protein-protein interactions, it was of interest to determine whether these direct interactions are 'species-specific'. The results presented in Figure 35B demonstrate convincingly that observed functional differences between human and yeast TBP are not based on differential interactions with either $hTAF_{II}250$ or $hTAF_{III}96$.

(iii) Purification of vaccinia-expressed His-TBP

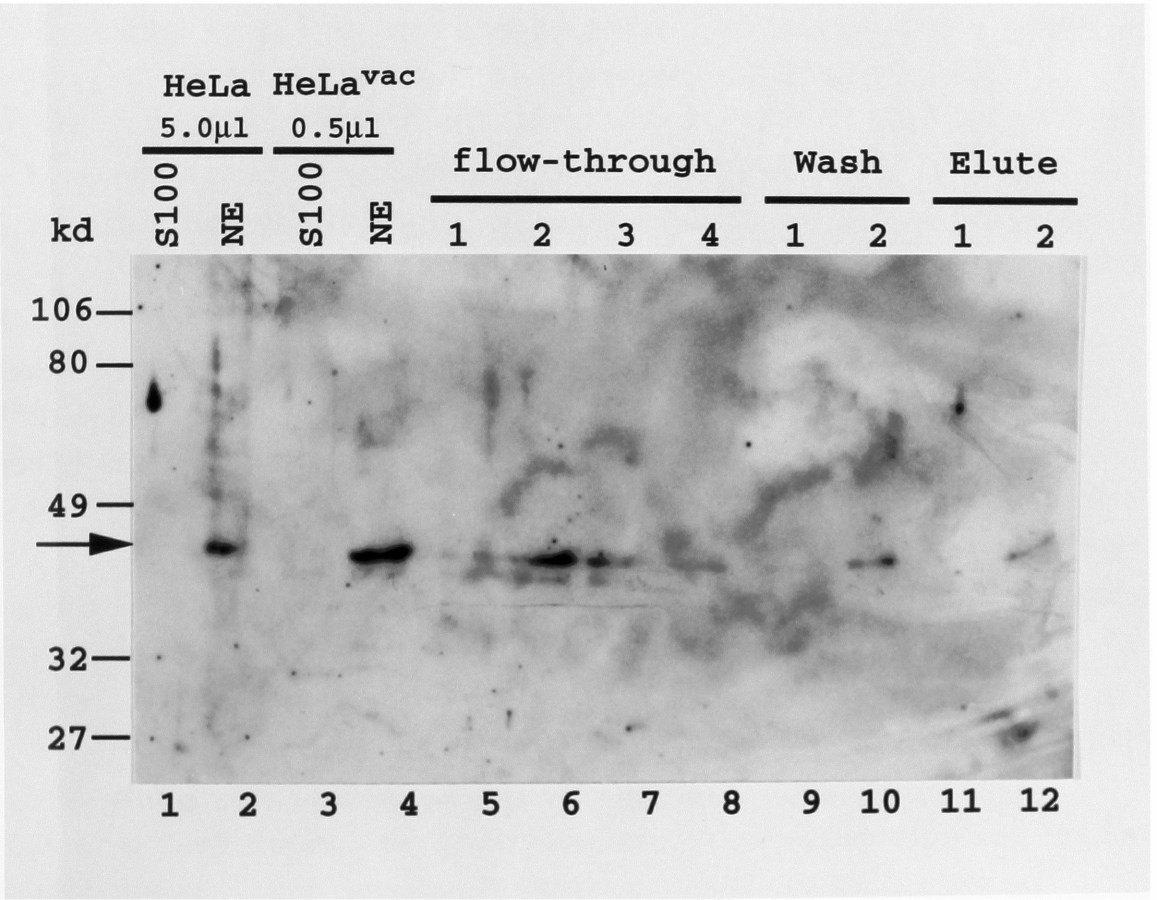
Rationale and purification strategies

Attempts to recover full functional activity characteristic of TFIID from eluted immuno-purified complexes were never fully successful, as the harsh conditions required to break antigen-antibody interactions or to elute the complete set of TAFs from the bound TBP seem to irreversibly denature TBP or inhibit reconstitution with exogenously supplied TBP (data not shown; cf.(Dynlacht et al., 1991; Tanese et al., 1991)). In some cases, though, activity can be assayed with purified complexes still immobilized on the antibody-sepharose beads, as shown for dTFIID (Kokubo et al., 1993b) and yTFIIIB (Poon and Weil, 1993), though such an approach seriously limits an in depth molecular characterization of TFIID functions. A reliable method for purifying native TBP-containing complexes in non-denaturing conditions was needed for physiologically relevant studies into the molecular functioning of TFIID.

Given the success of the non-denaturing purification method developed for His-tagged proteins expressed in *Escherichia coli* (chapter III, (i)), this tagging/ purification technology seemed suitable to provide a convenient step for the purification of native TFIID from mammalian nuclear extract. To this end, vaccinia virus capable of expressing hTBP with an N-terminal His-tag were produced, in a collaborative effort with the lab of Henk Stunnenberg (Janknecht et al., 1991), and used to infect HeLa cells. Western blotting of resulting nuclear extracts revealed a 20-fold overexpression of TBP with respect to uninfected cells (Fig. 36). Nuclear extracts were either applied to a metal-chelating column (Ni^{2+}-NTA) directly (Fig. 36) or fractionated conventionally on phosphocellulose and DE52 before loading this classical TFIID fraction onto Ni^{2+}-NTA. In both cases, however, the majority of the tagged protein failed to bind to the Ni^{2+}-resin or was eluted during extensive washing at moderate stringency (Fig. 36),

Figure 36. Overexpression of His-hTBP in HeLa cells using the recombinant vaccinia virus system.

Western blot with a -TBP antibody of fractions derived from plain HeLa cells (lanes 1 and 2) and HeLa cells infected with a His-hTBP expressing recombinant vaccinia virus (lanes 3 and 4) in indicated amounts. Fractions from a Ni^{2+}-NTA column loaded with nuclear extract derived from infected HeLa cells (lane 4) are shown as indicated in lanes 5-12. The arrow points out the mobility of human TBP on this SDS-gel.


HeLa
HeLa^vac
5.0µl
0.5µl
S100
NE
S100
NE
flow-through
Wash
Elute
kd
1 2 3 4 1 2 1 2
106
80
49
32
27
1 2 3 4 5 6 7 8 9 10 11 12

indicating at least a partial occlusion of the affinity tag, possibly by TBP-associated factors.

Characterization of His-TBP-containing eluates

Two assays were subsequently performed on the immidazole-eluted material to test for TFIID function. First, the gel shift assay demonstrated that both protocols had yielded protein complexes capable of binding specifically to a large TATA box-containing probe, resulting in a probe mobility shift equivalent to conventionally purified ω -amino-octyl TFIID (Fig. 37). As expected, the unselected NTA eluate also contained much larger amounts of a lower molecular weight species, presumably caused by uncomplexed TBP, than the NTA eluate derived from the second column TFIID fraction. Upon careful inspection of the competition behaviour of the complexes, it appeared that while native TFIID bound to the major late promoter fragment is inefficiently competed with short (30bp) TATA-containing competitor, the NTA eluates bind such short DNAs apparently very efficiently. These observations were reminiscent of my analysis of the gel-shift produced by retic-lysate translated TBP (chapter I, (iii)), indicating a possible involvement of TFIIA.

Consistent with this interpretation are the results from a transcription assay with the activator USF (Fig. 38). While native second or third column TFIID preparations allow for a significant level of USF-dependent stimulation of transcription, results with the unselected NTA eluate were indistinguishable from those with bacterially produced and purified TBP, indicative of the huge excess of uncomplexed TBP. Transcriptional activation assays with the NTA eluate from the second column TFIID fraction, however, exhibited some stimulation above backgound, though far short of the stimulation seen with native TFIID fractions.

Figure 37. Gel shift assay with His-hTBP^{vac}-containing fractions.

A radiolabelled TATA box containing, MLP-derived DNA fragment was incubated with amino-octyl fraction of native TFIID (lanes 1-3), vaccinia His-TBP extract-derived first column NTA fractions (lanes 4-6), or vaccinia His-TBP extract-derived third column fractions (lanes 7-9). Unlabelled oligonucleotides (MLP, positions -45 to -15) contained either a wild type (wt) or mutant (mu) TATA box. The arrow indicates a previously characterized TATA box-specific complex, the asterisk a nonspecific band.

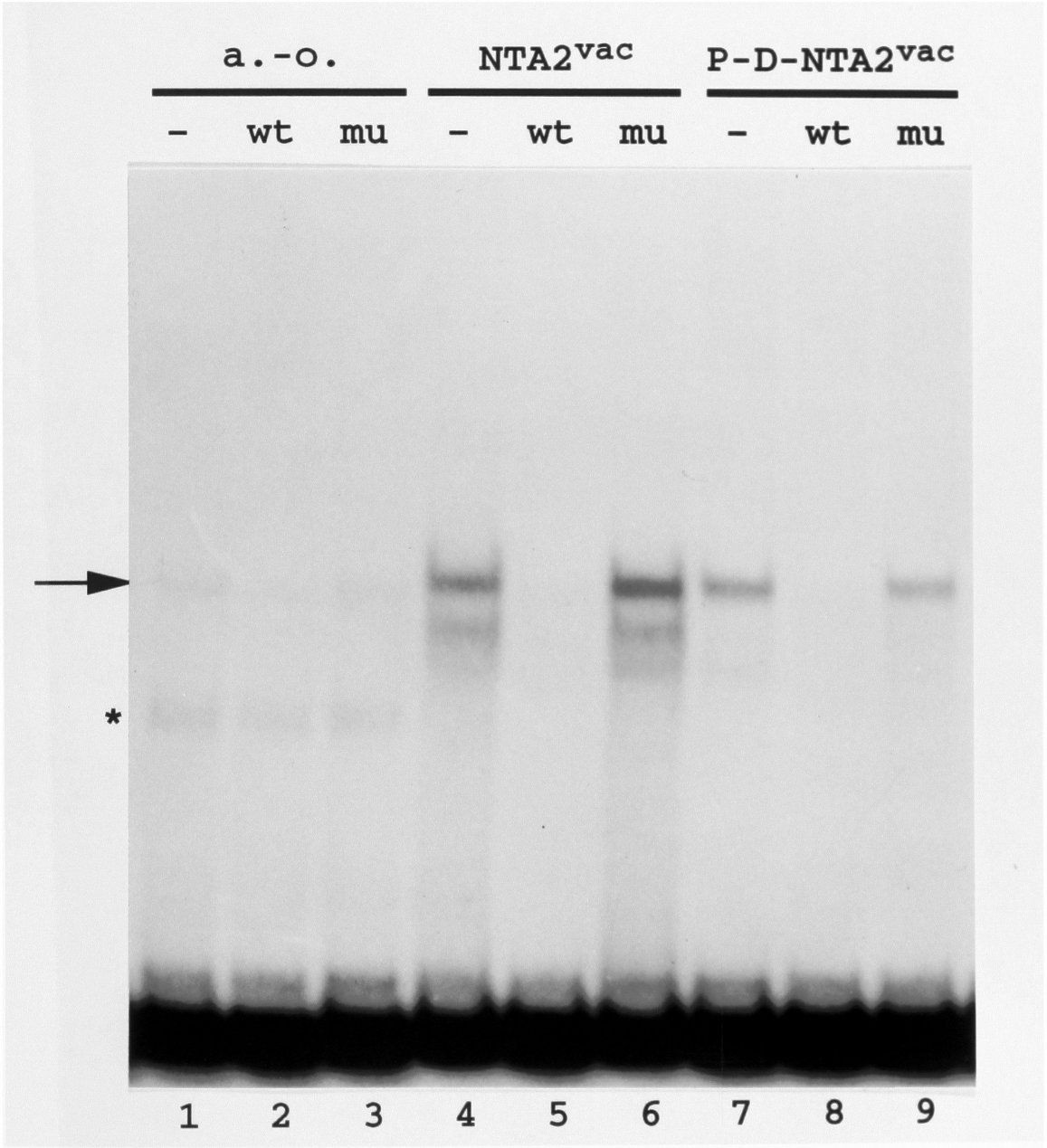
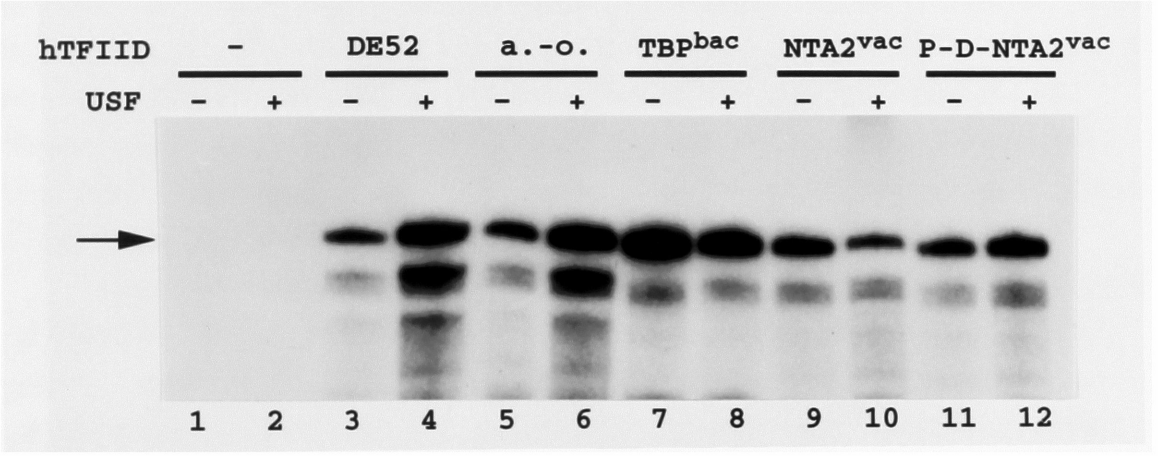

a.-o.
NTA2vac
P-D-NTA2vac
-
wt
mu
-
wt
mu
-
wt
mu
*
1
2
3
4
5
6
7
8
9

Figure 38. Testing for activation by USF with His-hTBP^{vac}-containing fractions.

Reactions containing an adenovirus MLP template with upstream sequences (-400 to +10) attached to the 380-bp G-less cassette pML(C_2AT) and HeLa factors TFIIB, TFIIE/F, and RNA polymerase were complemented with buffer (lanes 1 and 2), DE52 fraction of native human TFIID (lanes 3 and 4), amino-octyl fraction of bative human TFIID (lanes 5 and 6), purified bacterially expressed human TBP (lanes 7 and 8), first column NTA (lanes 9 and 10) and third column NTA (lanes 11 and 12) fractions deived from HeLa cells infected with His-TBP expressing recombinant vaccinia virus. A DE52 fraction of human activator USF was present in even-numbered lanes. The arrow indicates accurately initiated full length transcripts.

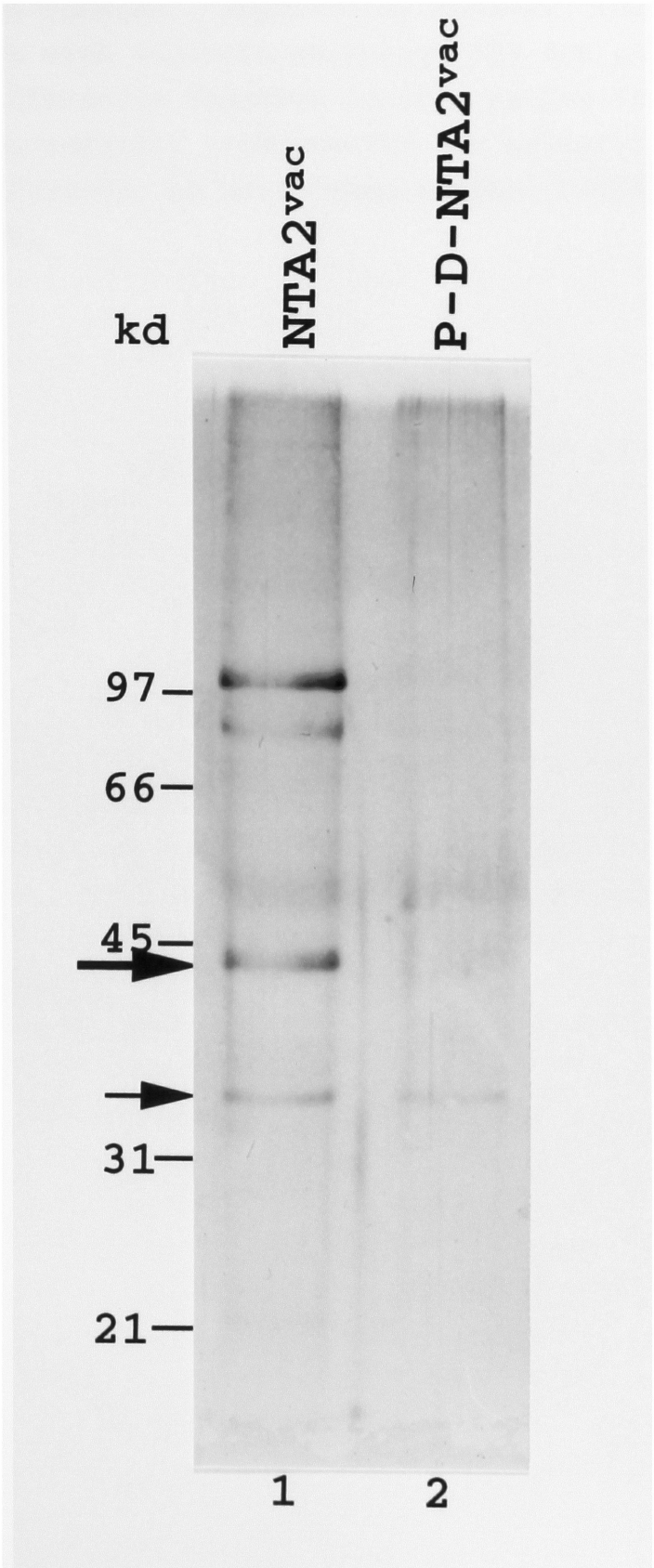

hTFIID
-
DE52
a.-o.
TBP^bac
NTA2^vac
P-D-NTA2^vac
USF
- + - + - + - + - + - +
1 2 3 4 5 6 7 8 9 10 11 12

Analyses of the prominent polypeptides present in the NTA eluate fractions by SDS-PAGE and silver staining (Fig. 39), show a lack of those bands subsequently identified as TAFs by immuno-precipitation techniques. With equivalent gel-shift units loaded, the cruder NTA eluate contained, as expected, a larger amount of TBP than the pre-fractionated NTA eluate, which showed only one co-purifying band above background of approximately 35kd. Due to limitations in amounts recovered from this purification, putative TAFs might be present at levels below the detection limit. It is intriguing to note that TFIIA, which partially co-purifies with TFIID and is a good candidate for yTBP-associated proteins (heDAP35, heDAP21, heDAP12) recently identified by protein affinity chromatography (Coulombe et al., 1992), was recently purified to homogeneity (DeJong and Roeder, 1993; Ma et al., 1993) and shown to consist of three subunits, of which the largest has an apparent molecular mass of 35,000 and binds metal-chelating resins due to endogenous histidine stretches in its sequence.

Several drawbacks of this strategy for the purification of native TFIID became apparent during the course of this work. In the first place, the histidine tag, despite its usefulness in bacterial, yeast or insect cell (baculovirus) expression systems, lacked the level of specificity of binding required for the efficient purification of histidine-tagged proteins from mammalian extracts. In addition, the low affinity of binding significantly reduced the recovery of the purification. Secondly, vaccinia virus mediated expression itself proved unsuitable for the particular purpose of purifying a multi-protein complex. Expression levels proved difficult to regulate, with a huge excess of the expressed protein in either a small or large portion of the infected cells (data not shown). Furthermore, the regulatory mechanisms intrinsic to the life-cycle of vaccinia virus dictate the shut-down of host cell protein synthesis, and thus presumably also of those proteins that are required

Figure 39. Analysis of polypeptides present in His-hTBP^{vac}-containing fractions.

Silver stained SDS-gel of first column (lane 1) and third column (lane 2) NTA fractions derived from HeLa cells infected with His-TBP expressing recombinant vaccinia virus. The thick arrow indicates the mobility of His-TBP, the thin arrow a repeatedly co-purifying polypeptide.


kd
NTA2vac
P-D-NTA2vac
97
66
45
31
21
1
2

in the desired complex. Importantly however, the results presented here with vaccinia-expressed TBP did rule out that functional differences observed between native TFIID and recombinant bacterially produced TBP are primarily due to important differences in post-translational modifications of these proteins.

(iv) Purification of epitope-tagged TFIID

The work described in this section is the result of a long standing commitment of mine, which was finally brought to full fruition in a collaboration with Cheng-Ming Chiang (Chiang et al., 1993).

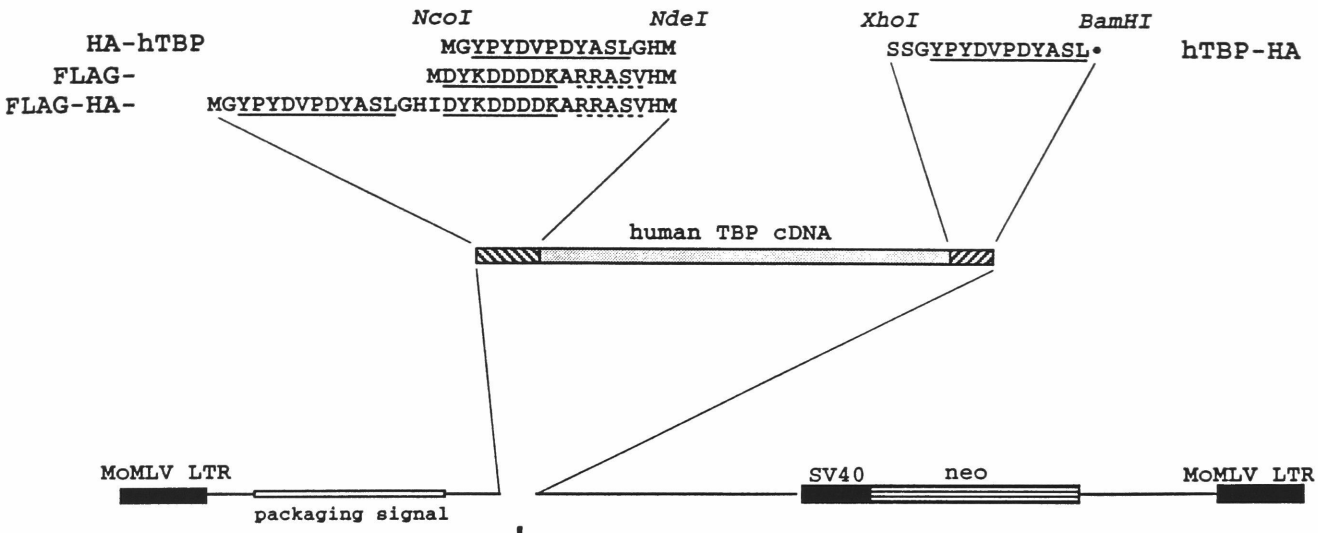
Strategy and methodology

The experience with techniques described in preceding sections had shown that for a convenient purification of native TBP-containing complexes suitable for functional characterization, a method was required that combined the specificity and high recovery of immuno-precipitation techniques (section ii) with the ability to elute bound proteins in non-denaturing conditions as described (section iii). Furthermore, in the interest of high yields and physiological relevance, the protein complexes should be extracted and purified from cells growing in log-phase, and the ligand that is supposed to bind to the affinity matrix should not be buried within the complex.

The strategy involving the establishment of cell lines expressing constitutively an N-terminally tagged TBP satisfied many of these criteria. Early attempts with histidine-tagged TBP cDNA in strong constitutive (CMV, ß-actin) or inducible (MMTV) promoters were frustrated due to the apparent toxicity of an over-expressed TBP in HeLa cells (data not shown) and NIH-3T3 cells (C. Marshall, personal communication), and the technical difficulty in distinguishing exogenous (histidine-tagged) from endogenous TBP by Western blotting. Two important improvements were introduced to facilitate the implementation of this approach. First, retroviral vectors in conjunction with retroviral amphotropic packaging cell lines increased the efficiency of complete integration into host genomic DNA. Second, the histidine-tag was replaced by epitope tags recognized with high specificity by commercially available monoclonal antibodies. This allowed not only for the efficient

Figure 40. Strategy for the native purification of epitope-tagged TBP-containing complexes.

Schematic of the transfection vector used to obtain recombinant retroviral particles directing the expression of N- or C-terminally tagged TBP in HeLa cells and summary of the experimental steps in this strategy leading to the purification of functional TBP-containing complexes. See text for further details.



DNA transfection into packaging cells Ψ-CRIP

harvest recombinant retrovirus particles from supernatant

viral infection of HeLa cells

neomycin selection

individual $G418^R$ colonies

expand into cell lines

identification of high expressors by Western blot

adapt to spinner culture

nuclear extract

fractionate, affinity purify

TBP-containing complexes

screening of cell lines for high expressors but also increased the maximum purification that can be achieved in a single affinity step. The resulting strategy thus combined the preferred aspects of the methods described in sections (ii) and (iii) : non-denaturing immunopurification of epitope-tagged TBP expressed at physiological levels.

Two different epitopes for commercially available monoclonal antibodies were used in these studies either by themselves or in combination. The influenza hemagglutinin (HA) epitope (Wilson et al., 1984) and the synthetic FLAG epitope (Hopp et al., 1988; Prickett et al., 1989) have both been used in to monitor expression of and to purify tagged proteins from bacteria and yeast (Field et al., 1988; Kolodziej and Young, 1989), but this methodology had not been extended to mammalian cells. While my collaborative work with Cheng-Ming Chiang concerning the establishment of FLAG-tagged TBP expressing cell lines and the functional analysis of purified FLAG-hTBP-containing complexes is described elsewhere (Chiang et al., 1993), I will describe in this section my work with the HA-epitope.

Plasmid constructs bearing N- and C-terminal tags encoding the HA-epitope were designed (Fig.40) with restriction enzyme sites analogous to bacterial expression vectors constructed previously (Hoffmann and Roeder, 1991) to permit convenient insert exchange. Retroviral vectors driving expression of the inserted cDNA with the Moloney murine leukemia virus LTR (Morgenstern and Land, 1990) contain a signal for amphotropic retroviral packaging in helper-free Ψ-CRIP cells (Danos and Mulligan, 1988), as well as an SV40 driven neo gene, conferring G418-resistant phenotype to transfected cells. Following the confirmation that added epitope tags did not affect the nuclear localization of TBP in transfected COS-7 cells as assayed by immunofluorescence (Chiang et al., 1993), HeLa cells were infected, as outlined in Figure 40, with recombinant retroviral particles harvested from transfected Ψ-CRIP

Figure 41. Identification of cell lines expressing HA-hTBP.
A. Western blot with α-HA-tag monoclonal antibody 12CA5 of nuclear extracts from cell lines derived from HeLa cells infected with HA-hTBP cDNA-containing retroviral particles in sense (lanes 2-8) or antisense (lanes 9 and 10) directions, or transfected with HA-hTBP-pBabe-neo vector (lanes 30-33). Lane 1 contains lysate from bacteria induced to express HA-hTBP. The arrow indicates the mobility of HA-hTBP.
B. Western blot with α-HA-tag monoclonal antibody 12CA5 of nuclear extracts from cell lines derived from NIH-3T3 cells infected with retroviral particles containing antisense HA-mTBP (lanes 1 and 2), sense HA-hTBP (lanes 3 and 4) and HA-mTBP (lanes 5 and 6). Lane 7 contains lysate from bacteria induced to express HA-hTBP. The arrows indicate the mobilities of HA-hTBP and HA-mTBP.

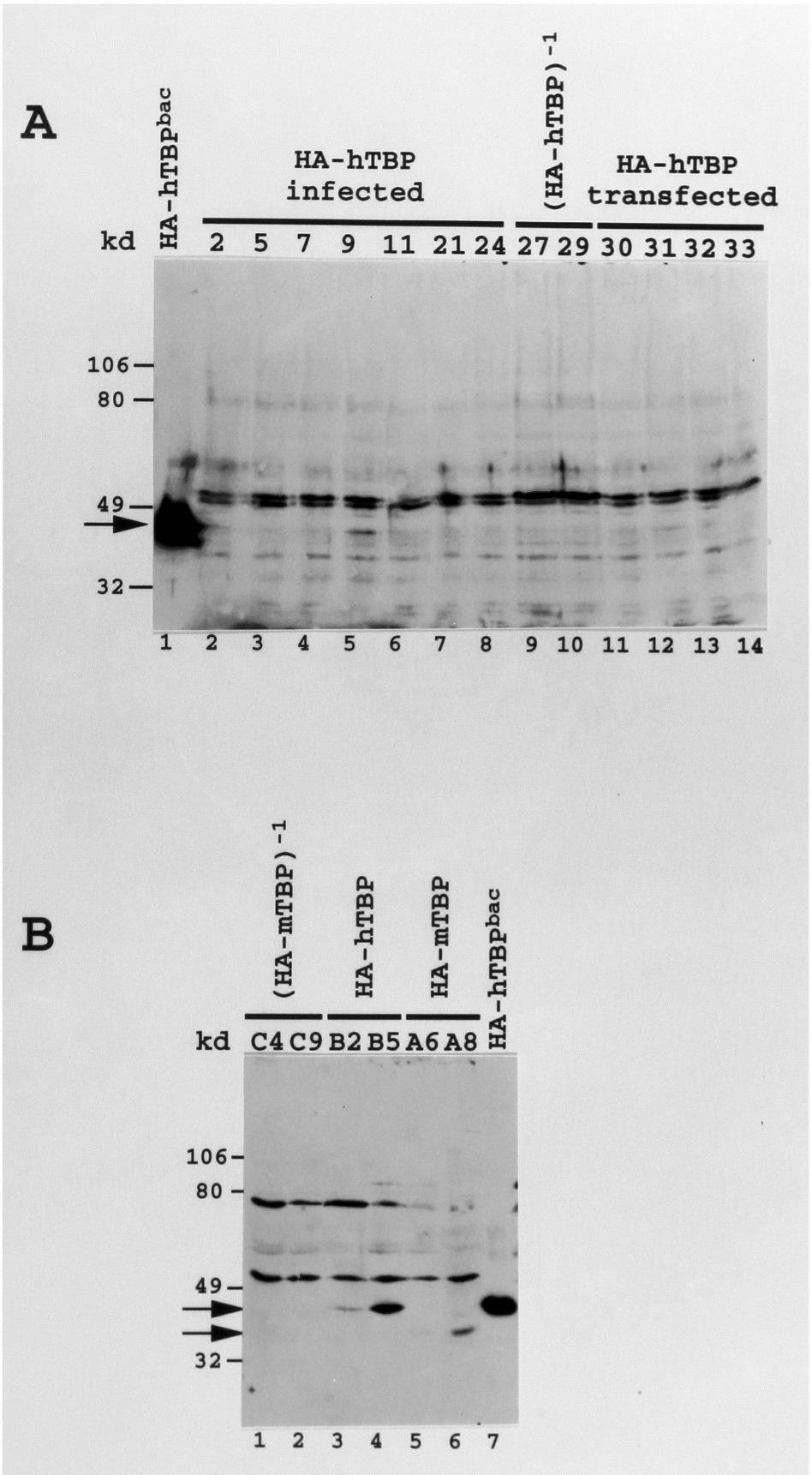

A
HA-hTBPbac
HA-hTBP infected
(HA-hTBP)-1
HA-hTBP transfected
kd
2 5 7 9 11 21 24 27 29 30 31 32 33
106
80
49
32
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
(HA-mTBP)-1
HA-hTBP
HA-mTBP
HA-hTBPbac
kd
C4 C9 B2 B5 A6 A8
106
80
49
32
1 2 3 4 5 6 7

Figure 42. Fractionation of HA-TBP across a P11 phosphocellulose column.

Distribution of HA-hTBP in P11 fractions revealed by Western blotting using the α-HA-tag monoclonal antibody 12CA5. Nuclear extract from N9 cells (lane 1) was loaded on a P11 column (flow-through fractions in lanes 2-8) and eluted at indicated KCl concentrations (lanes 9-11). The arrow indicates the mobility of HA-hTBP.

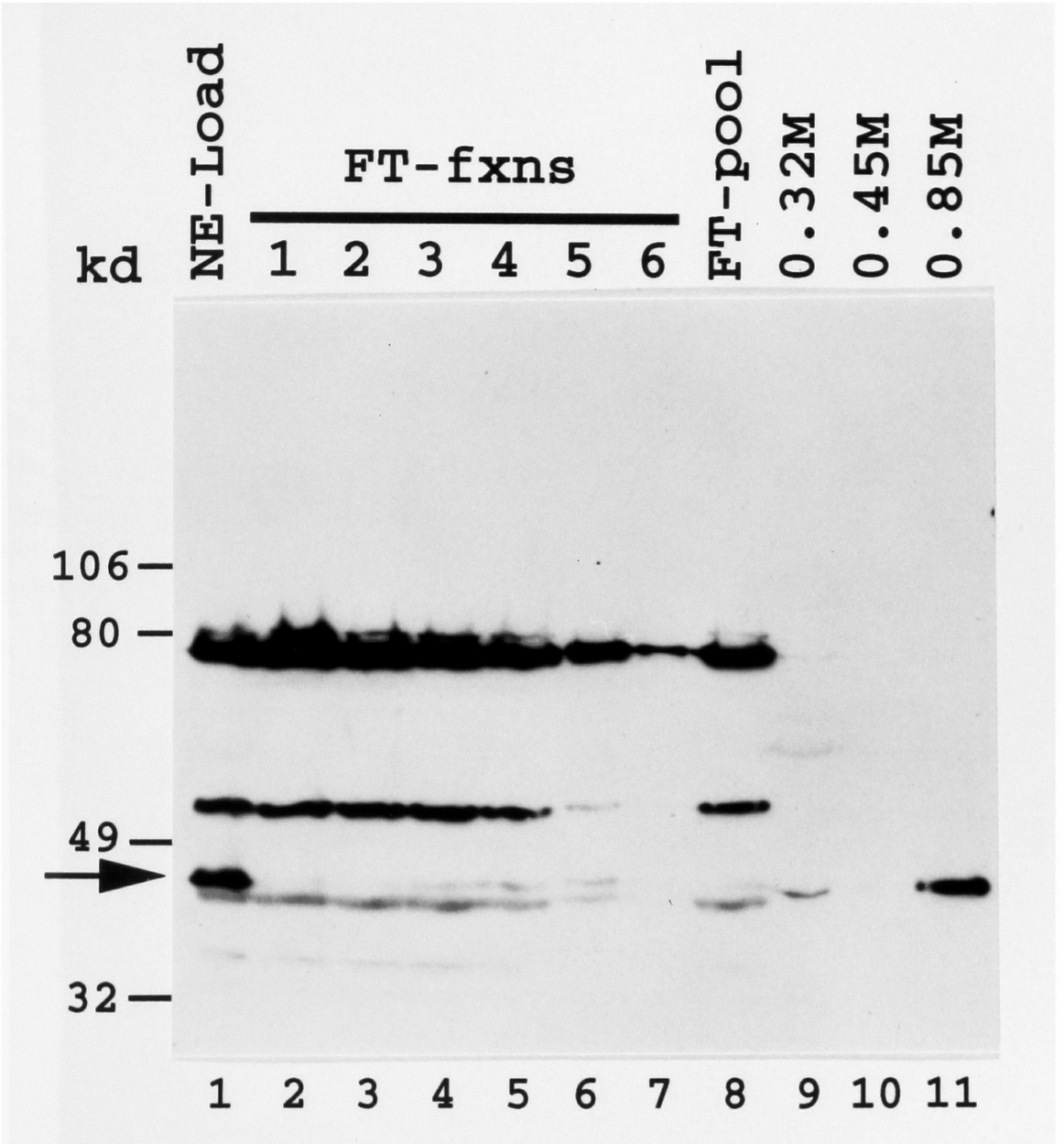
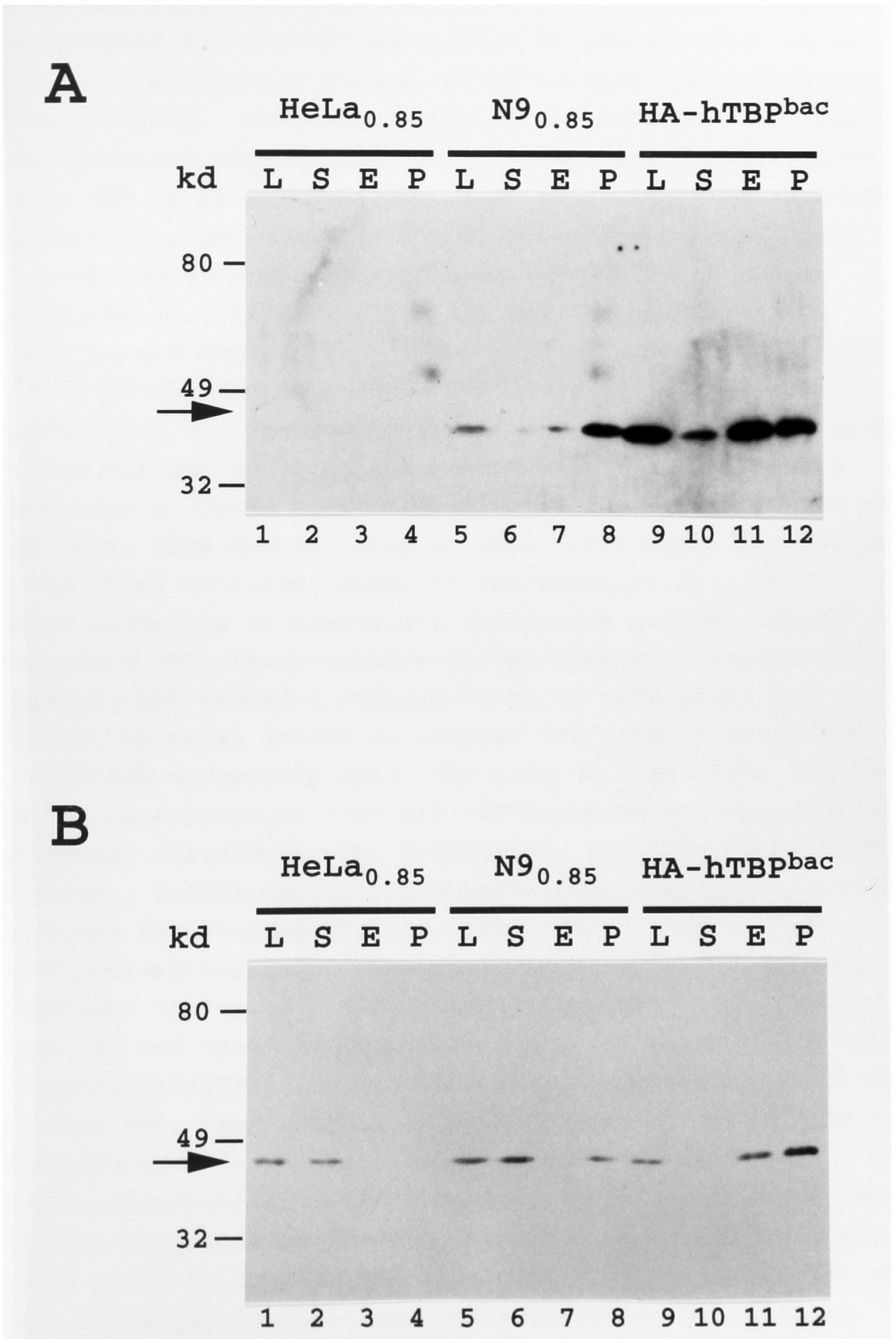

NE-Load
FT-fxns
1
2
3
4
5
6
FT-pool
0.32M
0.45M
0.85M
kd
106
80
49
32
1
2
3
4
5
6
7
8
9
10
11

Figure 43. Assessing the efficiency of immuno-affinity purification.

A. Western blot with a -HA-tag monoclonal antibody 12CA5 of samples derived from immuno-precipitation experiments with 12CA5-protein A-agarose and $P11_{0.85}$ fractions derived from normal Hela cells (lanes 1-4) or N9 HeLa cells (lanes 5-8), or extract derived from HA-hTBP expressing bacteria (lanes 9-12). Fractions are indicated as follows: Load (L), supernatant following incubation with antibody beads (S), eluate with HA-peptide containing buffer (E), and pellet of antibody beads following elution (P). The arrow indicates the position of HA-hTBP.

B. Western blot membrane described above reprobed with rabbit polyclonal α -TBP antiserum. The arrow indicates the position of hTBP and HA-hTBP.


A
HeLa0.85
N90.85
HA-hTBPbac
kd
L S E P L S E P L S E P
80
49
32
1 2 3 4 5 6 7 8 9 10 11 12
B
HeLa0.85
N90.85
HA-hTBPbac
kd
L S E P L S E P L S E P
80
49
32
1 2 3 4 5 6 7 8 9 10 11 12

cells, and resulting G418-resistant colonies were expanded and screened for HA-hTBP expression by Western blot of cell lysates or mini-scale nuclear extracts with the monoclonal antibody 12CA5. About one tenth of the more than 100 cell lines screened showed detectable, though low expression of tagged TBP in HeLa cells (Fig. 41A) reflecting the apparent selection for non-expressors presumably due to previously observed toxicity of overexpressed TBP (data not shown, (Chiang et al., 1993)). (NIH-3T3 cell lines expressing either tagged mouse TBP or human were established similarly with Ψ-CRE-derived retroviral particles (Fig. 41B) for the purification of the species-specific class I TBP-containing complex SL1/TIF-1B in collaboration with Ingrid Grummt's laboratory.) Other studies, published during the course of this work, also demonstrated an apparently tight regulation of the total cellular amount of TBP (Zhou et al., 1992): taking advantage of a mobility difference between tagged and endogenous TBP, these workers showed that high expressors of exogenous TBP showed a reduced level of endogenous TBP leaving the total amount of nuclear TBP roughly unchanged.

HA-TBP expressing HeLa cell line N9 (Fig. 41A, lane 5) and TBP-HA expressing line C17 (data not shown) were adapted to spinner culture for the preparation of large scale nuclear extracts. Fractionation phosphocellulose columns (Fig. 42) confirmed that tagged TBP was incorporated into distinct multi-protein complexes that elute in 0.32M and 0.85M KCl fractions, chosen to yield optimal separation and recovery of class III and class II TBP-containing complexes. (The class I complex SL1/TIF-1B, which constitutes only about 1% of the cellular TBP, requires higher salt in nuclear extraction for efficient protein solubilization and elution from phosphocellulose; I. Grummt, personal communication.) The relative abundance of these complexes varied somewhat from a 30-70% ratio (e.g. Fig. 42) to a 60-40% ratio (Chiang et al., 1993) depending on cell-line and preparation of nuclear

extract, though generally in good agreement with previous results (Timmers and Sharp, 1991).

To test the binding and elution characteristics of the immuno-affinity resin prepared by irreversibly crosslinking monoclonal antibody 12CA5 to protein A-sepharose, small amounts of the resin were incubated in parallel with phosphocellulose D fractions from HeLa and N9 cells, as well as bacterially expressed HA-TBP. Western blot analysis with tag-specific antibody 12CA5 (Fig. 43A) or α-TBP rabbit polyclonal antibody (Fig. 43B) confirmed that first, only tagged, not endogenous, TBP bound to the resin; second, binding of tagged TBP is highly efficient, a dramatic improvement over the His-tag binding to Ni^{2+}-NTA (Fig. 36); third, no leakage of cross-linked antibody was detectable. Elution in non-denaturing conditions by peptide competition, however, proved to be relatively inefficient (≈50% of bound HA-TBP eluable in three column volumes), compared to the elution efficiency observed with the α-FLAG "M2" resin (close to 100% eluable in three column volumes). Comparison of the respective peptide sequences recognized by the antibodies provides an explanation for this difference: whereas the FLAG sequence consists almost entirely of charged residues suggesting predominantly ionic interactions between antigen and antibody, the HA sequence features a high number of tyrosines allowing for more stable interactions. Finally, the Western analysis provided a way of estimating the amount of exogenous TBP expression relative to endogenous levels. Comparison of lane 6 in blots A and B (Fig. 43), for example, indicated a surprizingly low level of expression in this cell line, estimated at ≈10% of endogenous levels. Other cell lines expressing a FLAG or HA-FLAG tagged TBP were found to have somewhat higher relative expression levels (40% and 20% respectively) by similar assays (data not shown).

Identification of *bona fide* TFIID subunits

Immunopurification using described 12CA5 sepharose columns with phosphocellulose D fractions resulted in the specific purification of the TFIID complex, as comparison with mock purifications of non-expressing HeLa control fractions indicated (Fig. 44). At least 16 polypeptides were specifically and reproducibly retained with tagged TBP, some of which coincide in mobility with previously identified class II TAFs (section ii). The results from this purification method thus represent the most comprehensive identification of TFIID subunits to date. Additional, newly identified, TAFs were not detected previously because of poor resolution of SDS-PAGE due to the large amount of ethylene glycol in the elution buffer, and occlusion by leaking antibody light and heavy chains. Assuming a one-to-one stoichiometry the total molecular weight of all class II TAFs so far identified is well over 10^6 daltons, while Stokes radius estimates of native mammalian TFIID are consistent with an assymmetric molecule of ≈750kd (Conaway et al., 1991). Though parallel studies with Coomassie Blue, Ponceau S and Amido Black staining also noted the apparently equal stoichiometry of at least the higher molecular weight TAFs (Chiang et al., 1993), the idea of multiple complexes or structural heterogeneity in the TFIID fraction cannot be excluded.

No reproducible differences in polypeptide composition between TFIID preparations resulting from the N-terminally and C-terminally tagged TBP were detected. Antibody binding and elution was slightly more efficient with the N-terminal tag, consistent with the idea that the 154 residue N-terminus of human TBP is not involved in stable TAF interactions but rather is free to interact transiently with other components of the transcription machinery or factors involved in the regulation of initiation.

Figure 44. Analysis of polypeptides associated with HA-TBP. Immuno-purified complexes from $P11_{0.85}$ fractions derived from normal HeLa cells (lane 4), N9 HeLa cells expressing HA-hTBP(lane 2), or C17 HeLa cells expressing hTBP-HA (lane 3) were separated on a 10% polyacrylamide-SDS gel and visualized by silver staining. The estimated molecular weights of putative class II TAFs from this and other analyses (Takada et al., 1992; Chiang et al., 1993) are indicated on the right; the arrow marks the position of epitope-tagged hTBP.

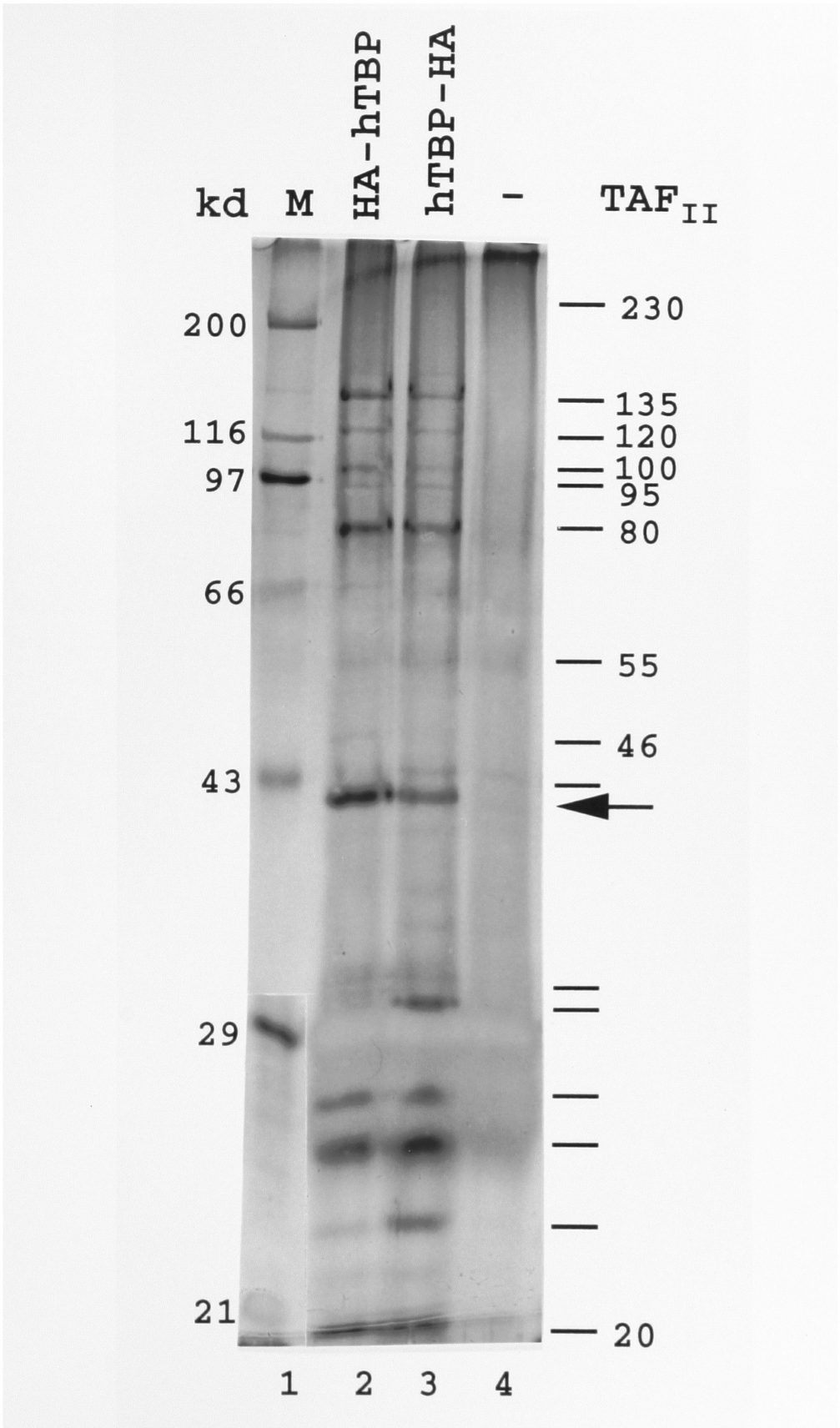
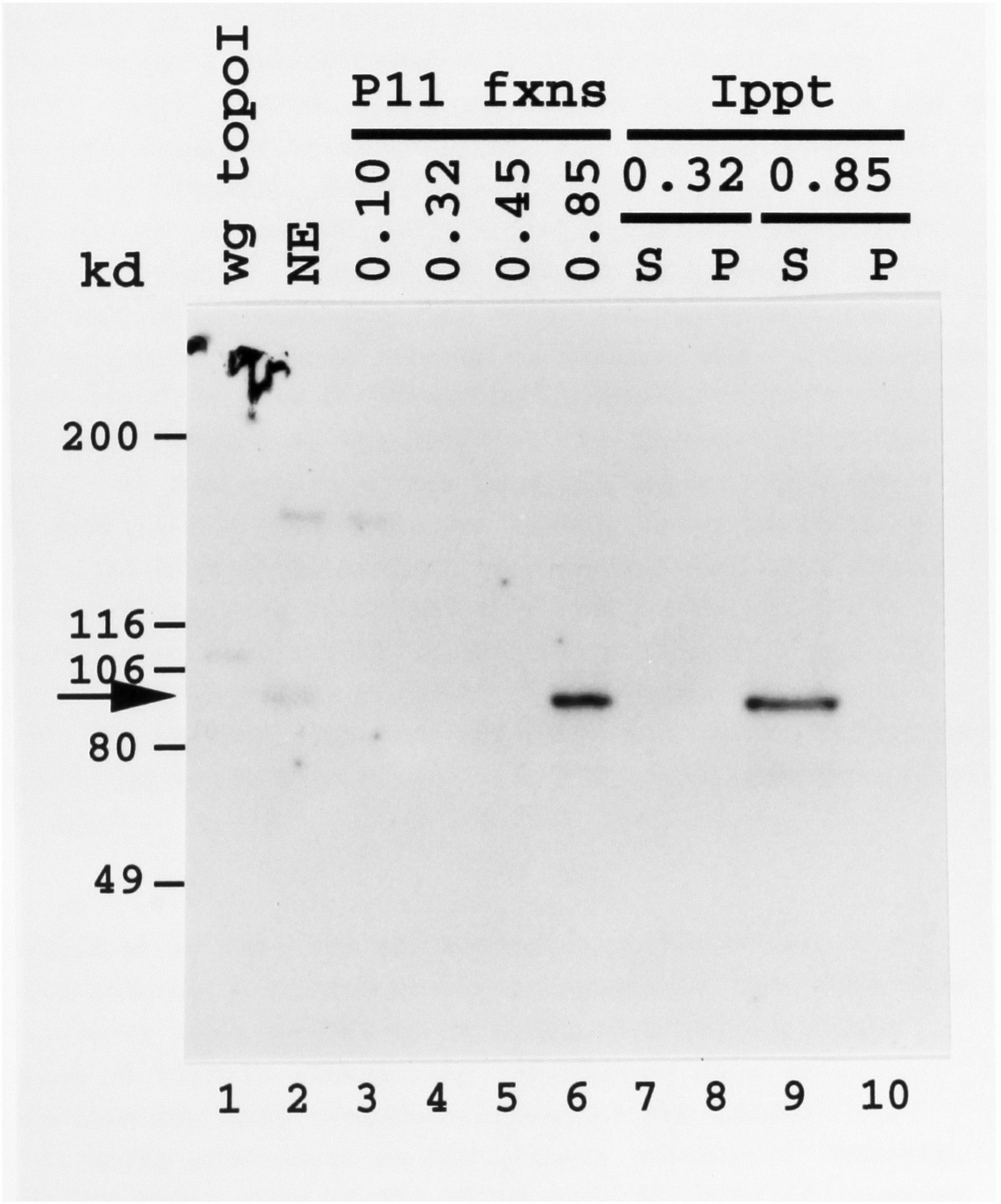

kd
M
HA-hTBP
hTBP-HA
-
TAF$_{II}$
200
116
97
66
43
29
21
230
135
120
100
95
80
55
46
20
1
2
3
4

Figure 45. Is TopoI a subunit of the TFIID complex ?
Western blot with human auto-immune serum against Topo I of HeLa N9-derived nuclear extract (lane 2) and P11 fractions (lanes 3-6), as well as HA-TBP immunoprecipitates from $P11_{0.32}$ (lanes 7 and 8) and $P11_{0.85}$ (lanes 9 and 10) fractions. Supernatant (S) and pellet (P) lanes contain approximately equal amounts of hTBP and HA-hTBP respectively. Lane 1 contains purified wheat germ Topo I. The arrow indicates the mobility of human Topo I.


wg topoI
NE
P11 fxns
0.10
0.32
0.45
0.85
Ippt
0.32
0.85
S
P
S
P
kd
200
116
106
80
49
1
2
3
4
5
6
7
8
9
10

Parallel studies proved that similarly purified TFIID preparations are capable of mediating high levels of activated transcription in conjunction with the soluble cofactor fraction USA (Chiang et al., 1993). As one of the components of the USA fraction had been identified as topoisomerase I (Kretzschmar et al., 1993) which under certain conditions was found associated with TFIID, it was of pertinent interest to find out whether observed TAF100 or TAF95, for instance, were in fact TopoI. Western analysis (Fig. 45) of phosphocellulose fractions with TopoI-specific antibodies indeed confirmed that TopoI fractionates in the TFIID-containing fraction, but TFIID immunoprecipitates at high stringency did not contain an immuno-reactive component, though controls with α-TBP-specific serum confirmed that a significant amount of TBP had been precipitated (data not shown). We can conclude that topoI is not an intrinsic component of the TFIID complex, but may be transiently associated with it to perform its co-activator function. Subsequent analysis confirmed this conclusion by demonstrating that substoichiometric amounts of TopoI are associated with highly purified TFIID, presumably via a direct interaction with TBP, but that this can be eluted from immobilized TFIID at moderate salt concentrations (Merino et al., 1993).

Attempts at mapping the TFIID-complex architecture

To distinguish TFIID-associated polypeptides from *bona fide* subunits of what appears to be a discrete multi-protein complex of tightly interacting components, high salt resistance has been employed throughout the immuno-purification procedures as the primary criterion. However, given the sheer size of the TFIID complex it is of interest to identify a potential "TFIID core complex" through the use of mild detergents and protein-protein interaction inhibitors. Such attempts might shed light on some of the

interactions within the complex and provide clues on the spacial co-ordination of the many subunits within it. Previously published work had apparently succeeded in splitting co-activator activities from the TFIID complex by the use of low urea concentrations in conventional chromatography (Dynlacht et al., 1991; Tanese et al., 1991). However, due to the crudeness of the functional complementation system, these observations cannot rule out that small concentrations of urea might, in fact, primarily facilitate the separation of USA-like (Meisterernst et al., 1991) co-activator activities from the TFIID complex. Furthermore, my own studies with purified TFIID attempting to characterize the molecular basis for those functional observations, indicated that increasing amounts of urea cause the complete disruption from TBP of some fraction of TFIID complexes, indistinguishable from TBP-specific heat-treatment (data not shown). This suggests that urea does not primarily disrupt specific protein-protein interactions within the TFIID complex, but has a denaturing effect on TBP leading to the disruption of the TFIID complex.

Other reagents were sought that have a more specific effect on protein-protein interaction within TFIID. Figure 46 presents an example of such a study, involving the immuno-purification of TBP-associated factors in the presence of molecular interaction inhibitors. While even high concentrations of heparin (lane 7) do not seem to have any effect, and guanidinium seems to primarily cause denaturation of TBP (lanes 1-4) similar to urea, variable concentrations of sarkosyl seem to have differential effect of the association of various TFIID subunits. Strongly staining TAF80 , and TAF135, for example, are easily stripped from the complex while TAF250, TAF120 and some lower molecular weight TAFs are more resistant to sarkosyl stripping.

Figure 46. Attempts at partially disrupting the TFIID complex.

Immuno-precipitations with 12CA5-protein A-agarose from $P11_{0.85}$ fractions derived from control HeLa cells (lanes 1, 3, and 9) or N9 HeLa cells (lanes 2 and 4-8) were performed in the presence of indicated concentrations of guanidinium-HCl (lanes 1-4), sarkosyl (lanes 5 and 6), and heparin (lane 7) in binding and wash buffers. Immuno-precipitaes were separated on a 12% polyacrylamide-SDS gel and visualized by silver staining. The arrow indicates the mobility of HA-hTBP.

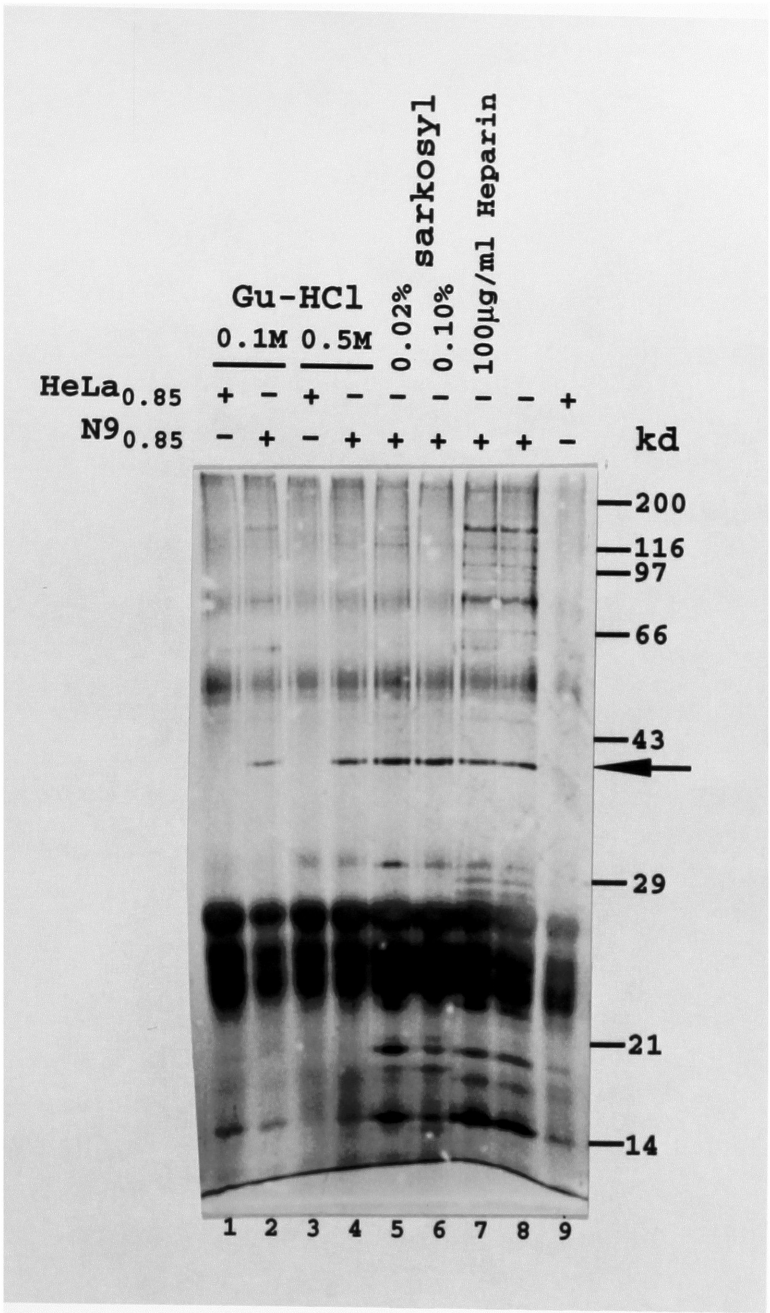

Gu-HCl
0.1M
0.5M
0.02% sarkosyl
0.10% sarkosyl
100µg/ml Heparin
HeLa0.85
N90.85
kd
200
116
97
66
43
29
21
14
1 2 3 4 5 6 7 8 9

Figure 47. Summary of partial disruption experiments. Polypeptides observed to co-immuno-purify with human TBP from HeLa cell-derived $P11_{0.85}$ fractions in this or other (Takada et al., 1992, Chiang et al., 1993) analyses are listed as putative class II TAFs and are denoted by their estimated molecular weight. Their reproducible presence or absence in 12CA5-protein A-agarose immuno-precipitation pellets following incubation and washes with high salt buffers containing indicated sarkosyl concentrations is noted by a "+" or "-", respectively. Question marks indicate difficulty in identifying the corresponding bands with certainty or lack of reproducibility.

sarkosyl /%	-	0.01	0.02	0.05	0.1
TAF_{II}					
230	+	+	+	?	?
170	+	?	-	-	-
135	+	+	+	?	-
120	+	+	+	?	?
105	+	-	-	-	-
100	+	+	-	-	-
95	+	+	-	-	-
80	+	+	-	-	-
70	+	-	-	-	-
60	+	-	-	-	-
55	?	?	?	?	?
46	+	+	+	?	-
43	?	?	?	?	?
TBP	+	+	+	+	+
31	+	+	+	+	+
30	+	+	-	-	-
28	+	+	-	-	-
20	+	+	+	+	+
19	+	?	?	?	?
18	+	?	?	?	?
15	+	+	+	+	+
14	+	+	?	?	?

The results from numerous such experiments (summarized in Fig. 47) allow for some tentative conclusions on the physical architecture of the TFIID complex. TAF80, TAF100, TAF95, TAF55 and to a lesser extent TAF135 are likely to be located in the outer part of the complex, not required for the association of any of the more inner core TAFs, as for example TAF250, TAF120, TAF31, TAF20, TAF15 that in turn are more likely to make direct contact with TBP.

The availability of recombinant TAF proteins allow for the mapping of direct interactions within the complex. Partial disruption experiments with the native complexes may complement these, particularly when a whole battery of anti-TAF antibodies can be used both for the confirmation of silver-stain identifications by Western blot as well as for the immuno-purification of partially disrupted TAF complexes that do not contain TBP (see chapter IV). Furthermore, functional assays with partial complexes may help in the functional characterization of individual TAFs, as for example co-activators (possibly activation surface-specific), mediators of pre-initiation complex assembly, or DNA binding components of TFIID, that may be important for template commitment of transcription initiation from TATA-less promoters.

Chapter III : Studies with recombinant TBP

(i) Bacterial His-tag expression system

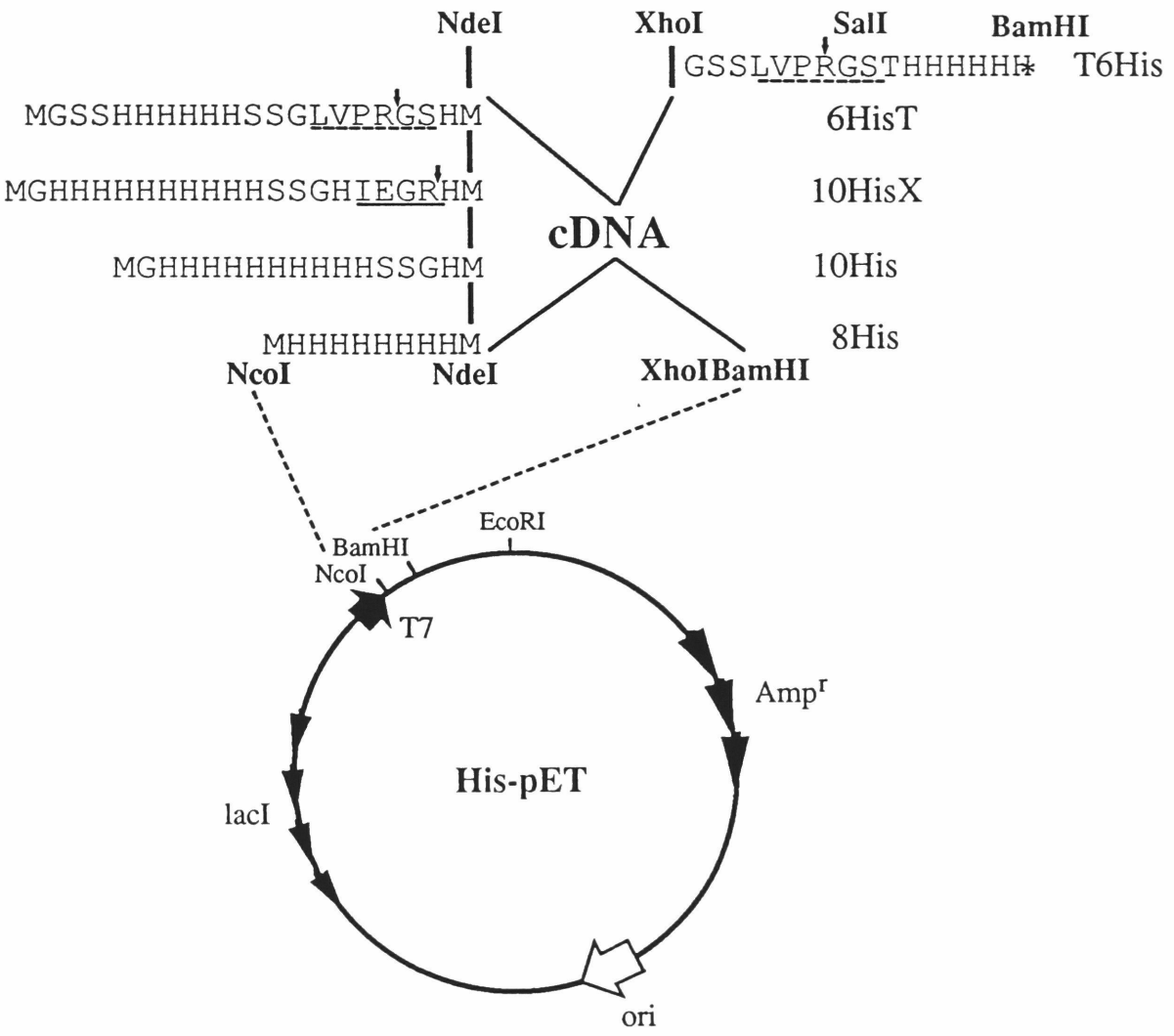
The work described in this section was previously published in Hoffmann and Roeder (1991); the His-pET bacterial expression vectors are available from Novagen Inc., Madison.

Construction of His-pET expression vectors

A detailed molecular characterization of biologically active polypeptides ideally involves the structural and functional analysis of recombinant wild type and mutant forms of the protein *in vitro*. Bacterial expression systems have proven to be convenient for the small or large scale production of many recombinant proteins in fully functional form and have therefore attracted considerable technological interest resulting in a number of important innovations regarding subcloning convenience, expression vector stability, host cell protease deficiency, inducibility of expression, and possible toxicity of recombinant gene products. A reliable expression system developed in the laboratory of Studier utilizes the efficient bacteriophage-encoded T7 polymerase and its highly specific cognate promoter element to produce high, inducible expression of exogenous cDNAs in *Escherichia coli*, with a minimum level of uninduced expression to allow high level production of even toxic proteins (Studier et al., 1990).

A number of approaches have also been pursued to facilitate the purification of the recombinant protein, based on a variety of well characterized high affinity interactions. The modular architecture of proteins has permitted the design of fusion proteins, in which the cDNA of the protein of choice is fused to the cDNA of a polypeptide that has a well characterized affinity to a particular ligand that can be immobilized on a support. Based on this principle protein purification systems have been designed

Figure 48. Construction of His-pET plasmids.
Design of the histidine tag encoding linkers cloned into the bacterial expression vector pET11d (Studier et al., 1990). N-terminal tags containing 6, 8,or 10 histidines and cleavage sites for the specific proteases thrombin (dashed line) and factor X_a (solid line) are inserted between the *Nco*I and *Bam*HI sites resulting in vectors 8His-pET11, 10His-pET11, 10HisX-pET11, 6HisT-pET11,and 6HisTX-pET11. cDNAs are inserted with their first ATG at the *Nde*I site and their 3'-end at the *Xho*I and *Bam*HI sites. A C-terminal tag is created by cloning an appropriate linker into the pET11a vector allowing open reading frames to be inserted between the *Nde*I and *Xho*I cloning sites.


NdeI
XhoI
SalI
BamHI
GSSLVPRGSTHHHHHH*
T6His
MGSSHHHHHHSSGLVPRGSHM
6HisT
MGHHHHHHHHHHSSGHIEGRHM
10HisX
cDNA
MGHHHHHHHHHHSSGHM
10His
MHHHHHHHHM
8His
NcoI
NdeI
XhoIBamHI
BamHI
NcoI
EcoRI
T7
Amp^r
His-pET
lacI
ori

utilizing the high affinity interaction between maltose binding protein (MBP) and amylose (Guan et al., 1987) as well as the popular glutathione S-transferase (GST) and glutathione (Smith and Johnson, 1988). Alternatively, epitopes to particular monoclonal antibodies (e.g. 12CA5, FLAG; chapter II(iv), (Chiang and Roeder, 1993)) can be fused to the protein under investigation, which can then be purified with the aid of the corresponding antibody resin. While the former method is relatively inexpensive and allows for high yields, the relatively large domain fused to the protein of interest might affect its functional properties and thus often needs to be removed by protease cleavage. The latter method is considerably more costly and unsuited for antibody production, but exhibits high specificity, only requires a short fusion peptide, and allows for protein monitoring by Western blot analysis.

Following the observation that stretches of histidine residues act as metal chelators and bind particularly specifically to nickel ions immobilized on agarose (Hochuli et al., 1987), it was of obvious interest to engineer such metal binding sites into the protein under study. For the purposes of *in vitro* structure function studies with TBP, I chose to construct a set of universally applicable expression vectors, based on the T7 polymerase mediated expression system, that allow for the production of recombinant protein fused (at the N- or C-terminus) with a variety of histidine stretch-containing tags (Fig. 48), that can be cleaved off, if necessary, with thrombin (Nagai and Thøgersen, 1984) or factor X_a (Eaton et al., 1985).

The resulting vectors have since proved popular both for the small scale production of recombinant protein for functional studies, as well as for large scale expression for the purposes of antibody production. As restriction sites used for the insertion of cDNAs are compatible with other bacterial expression vectors (no tag, pET system (Studier et al., 1990); FLAG-tag (Chiang and Roeder, 1993); GST fusion,

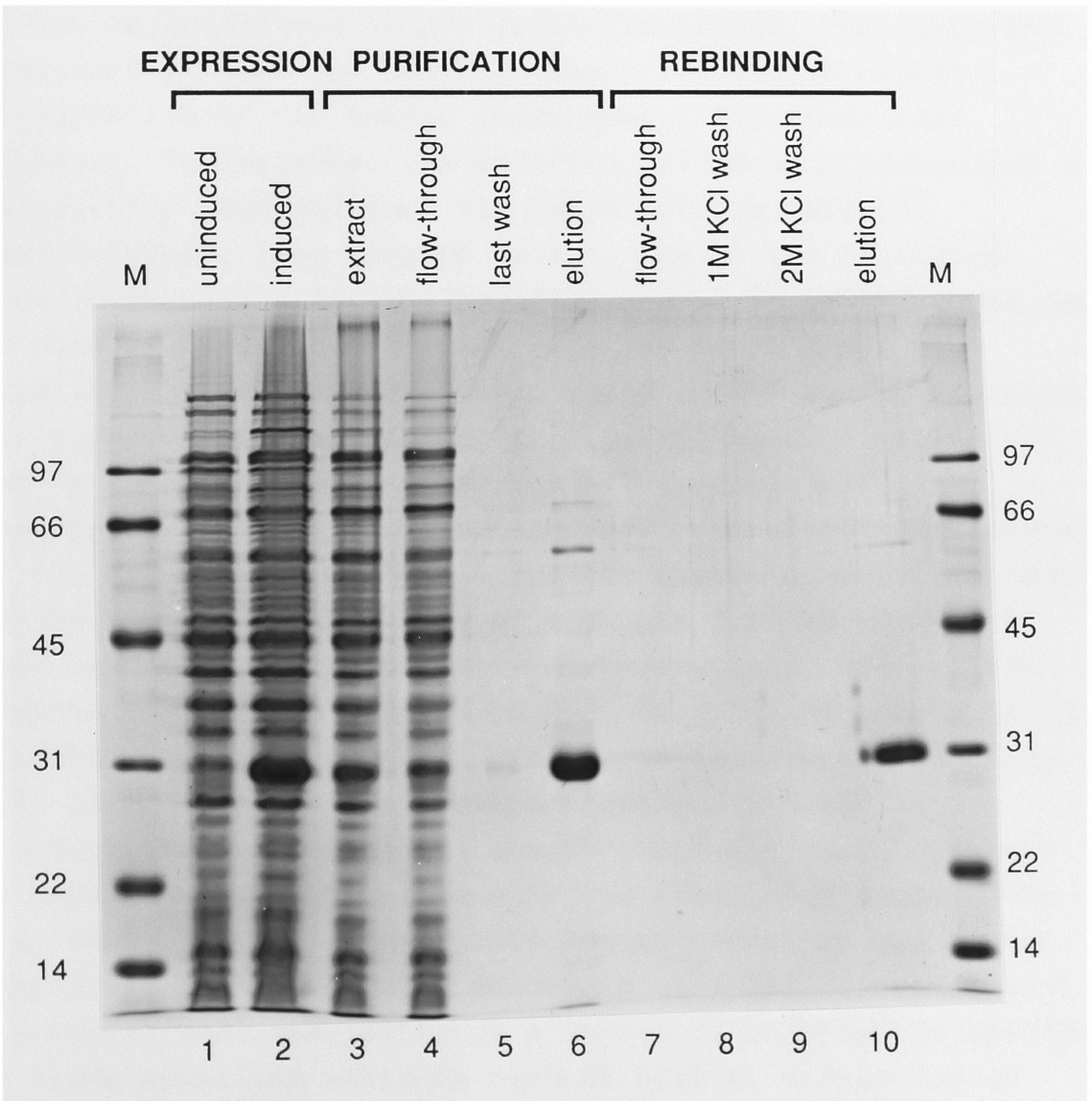
modified pGEX-2TL(+) and pGET vectors (data not shown)) as well as vectors for baculovirus-, vaccinia virus-, and retrovirus-mediated expression and transfection vectors (data not shown), cDNAs encoding the protein of interest can be conveniently transferred to the vector of choice in a single cloning step. Furthermore, none of the His-tagged proteins studied to date have shown any functional impairment by the addition of the tag at their N- or C-terminus. As the His-tag does not seem to affect the expression level or solubility of the protein in bacteria, purification of the expressed protein in denaturing conditions from insoluble inclusion bodies or in non-denaturing conditions from the soluble bacterial extract by metal-chelating chromatography can prove useful.

Non-denaturing purification methodology

While previous descriptions of immobilized metal-affinity chromatography (IMAC) had without exception used negative pH gradients for the elution of bound proteins, I began to purify bacterially-expressed proteins in non-denaturing conditions at physiological pH by 'chelator competition' using imidazole containing buffers, as shown for yeast TBP in Figure 49 (lanes 3-6). The purified protein was shown to be functional in DNA binding (gel shift and DNAseI footprint assays) as well as transcription assays with a specific activity equal or greater than conventionally purified, untagged recombinant yTBP (data not shown). Subsequent reduction of the imidazole concentration in the purified protein fraction by dialysis or dilution permits nearly quantitative rebinding to the resin (lane 7). The binding of the His-tagged protein to the resin is resistant to high salt concentrations (lanes 8, 9), permitting the washing off of contaminant DNA fragments non-specifically bound to recombinant transcription factors, to non-ionic detergents

Figure 49. Expression and affinity purification of His-yTBP.

Coomassie-blue-stained SDS-PAG demonstrating expression, one-step purification, and rebinding of His-tagged yeast TBP. Expression (lanes 1 and 2) of yTBP within the 6HisT-pET11 vector in BL21(DE3)plysS bacteria was accomplished according to published procedures (Studier et al., 1990). Bacterial extract was made by repeated sonication in a buffer containing 10mM Tris.HCl (pH7.9), 10% glycerol, 0.5M NaCl, 10mM 2-mercaptoethanol, and protease inhibitors and a subsequent 10min spin at 10,000g. Imidazole.HCl (pH7.9) was added to the extract to a concentration of 1mM before loading onto a Ni2+-NTA column (Qiagen, 1ml per mg of expressed protein) at a flowrate of 10 column volumnes per hour (lane 4). The column was then washed with about 20 column volumes of BC100 (20% glycerol, 20mM Tris.HCl (pH7.9), 300mM KCl, 10mM 2-mercaptoethanol, and 0.5mM PMSF) containing 20mM imidazole.HCl (last wash fraction in lane 5) before being eluted with BC300 containing 80mM imidazole.HCl (lane 6). The eluate was diluted 10-fold in BC300 and loaded onto an identical column (lane 7). Two high salt washes with BC1000 (lane 8) and BC2000 (lane 9) demonstrate binding is salt resistant, but the bound protein can be recovered in BC300 containing 80mM imidazole.HCl (lane 10).


EXPRESSION
PURIFICATION
REBINDING
M
uninduced
induced
extract
flow-through
last wash
elution
flow-through
1M KCl wash
2M KCl wash
elution
M
97
66
45
31
22
14
97
66
45
31
22
14
1
2
3
4
5
6
7
8
9
10

(data not shown), and 6M guanidine hydrochloride (Gentz et al., 1989), while it is easily eluted in the buffer of choice containing imidazole at physiological pH (lane 10; 80mM for 6His, 150mM for 10His).

Whereas the his-tag purification methodology was previously restricted to certain proteins that are unharmed at low pH conditions or are easily renatured, the described procedure permits its use for almost any cloned protein, irrespective of the buffer requirements for functional activity. Furthermore, the modified methodology represents a potentially powerful tool for rapid protein affinity chromatography (see chapter II(i)), due to its following characteristics : (i) The purification or re-purification can be done in native conditions. (ii) The resin has a high capacity (2-10mg tagged protein bound on 1ml resin) allowing for rapid working up of many samples in batch. (iii) As the tag is short (in contrast to the GST system), protease cleavage is in many cases not necessary for functional tests of the recombinant protein. (iv) The system allows anchorage of the protein to the resin at a single defined point. (v) When used for protein affinity chromatography, the protein complex can also be allowed to form in solution before being purified with nickel resin, similar to immunoprecipitations. (vi) In contrast to most antibody-antigen interactions, however, the His-tag-nickel interactions are such that that we have the choice to elute only the associated protein (by e.g. salt, non-ionic detergents, denaturants) or the entire complex in native form (by imidazole or EDTA) for subsequent functional analysis. Although a number of polypeptide species in crude mammalian extracts bind at similar stringency as tagged proteins to the chelating resin, these can be identified readily by appropriate controls and partially eliminated by passage of the extract through a control nickel column in the absence of the his-tagged protein. Given the ease with which high amounts of recombinant proteins can be expressed and purified, the described system permits a direct

mutational analysis of protein interaction domains that will improve our understanding of the molecular basis for the transient association of proteins involved in regulatory pathways.

(ii) Structural studies of TBP

Structural studies on TBP by biophysical methods were exclusively accomplished in the laboratories of Mitsu Ikura (Toronto) and Stephen Burley (RU). My role in these collaborations was limited to supplying purified protein (for NMR), the development of an efficient purification methodology, or the construction of various TBP expression vectors. The results from the stunning crystallographic work have been published in Nikolov et al. (1992), Kim et al. (1993a), and in more detail in Nikolov and Burley (1994) and Kim and Burley (1994).

Introduction

Detailed structural studies of biologically active molecules can have an enormous impact on our understanding of their functioning in the cell. Such studies can provide the most convincing evidence for or against postulated hypotheses regarding homology relationships with other known protein structures or structure-function relationships based on mutagenesis studies. Importantly, they may guide future experiments probing the function of domains and sub-domains of the protein under investigation. Given the fundamental role of TBP in all eukaryotic transcription initiation mechanisms, its central presence in at least three distinct multi-protein complexes, and its postulated function as a direct target for transcriptional activators, there has been enormous interest (and speculation) in its three-dimensional structure.

Complete three-dimensional solutions to protein structures are currently accomplished by two very different techniques, each with its own advantages and drawbacks : Nuclear Magnetic Resonance (NMR) spectroscopy and crystallography by X-ray diffraction analysis. While NMR can be used to study the solution structure of the protein, however, with only limited resolution, X-ray diffraction relies on an ordered three-dimensional array of the protein as present in a well-ordered crystal. This often allows

structural solutions of high resolution revealing amino-acid side chains in great detail, but possible folding artefacts resulting from packing forces within the crystal need to be born in mind. Thus structural solutions generated by both techniques can indeed be mutually complementary (e.g. SH2 domain (Overduin et al., 1992; Waksman et al., 1992; Waksman et al., 1993)), and I therefore committed my efforts to collaborations with the protein NMR laboratory of Dr. Mitsu Ikura and the X-ray crystallography group of Dr. Stephen Burley.

NMR analysis attempts

NMR spectroscopy is the method of choice for studying relatively small macro-molecules in solution. Recent advances in NMR technology (e.g. stronger magnets) and methodology (e.g. multi-dimensional NMR) have made it possible to apply this technique successfully to proteins as large as calmodulin (156 amino acids, (Ikura et al.,)). The TBP conserved core domain seems therefore with 180 residues in length just within reach of analysis by NMR spectroscopy. However, the success of multi-dimensional NMR techniques as applied to proteins heavily depends on a number of factors : the polypeptide under study must be highly soluble and stable room temperature at high concentrations ($\geq$ 1μM) in phosphate buffers (must be devoid of components containing carbon or nitrogen). As spectroscopy must also be done with proteins labelled with non-radioactive isotopes ^{15}N and ^{13}C, bacterial expression of the recombinant polypeptide must be accomplished in minimal (defined) media. Cost considerations, furthermore, dictate high expression levels and an efficient purification procedure.

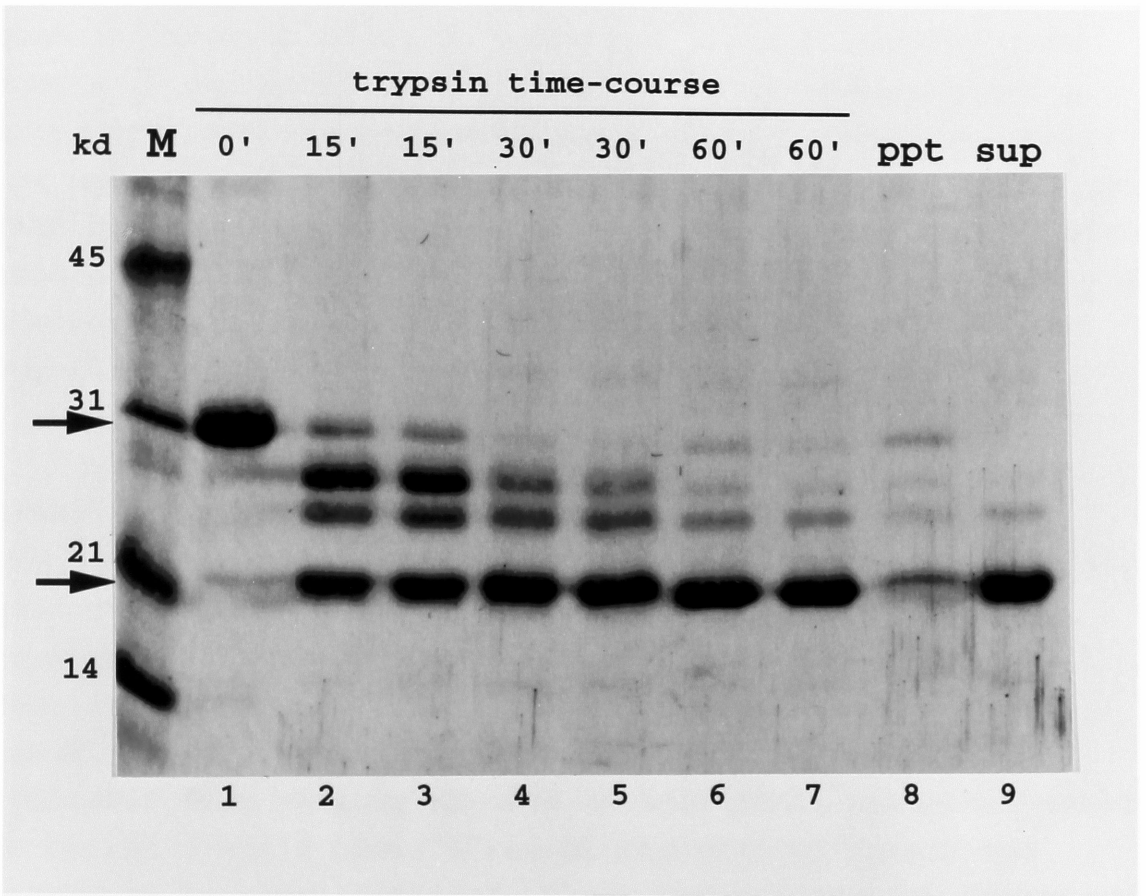
Latter requirements were fortunately met by the His-tag expression system : protein production proved to be as efficient in minimal media as in complex and cell yields per volume could be improved by a factor of 2.5 by the use of a bench top fermentor (data not shown). Given the availability

of TBP from a number of organisms, I compared their expression levels, extraction yields and solubilities in phosphate buffers (data not shown). The TBP from *Schizosaccaromyces pombe* was thus determined to have optimal properties (with a yield of 12-15mg of purified protein per liter of culture, data not shown), although extended storage at room temperature resulted in significant precipitation, so that spectroscopic analyses had to be limited to a 12hr period following buffer exchange (M. Ikura, personel communication). Mutation of the two endogenous cysteines to serines was hoped to alleviate solubility problems, but resulted in a mis-folded (by chromatographic behaviour) and non-functional polypeptide (data not shown).

To minimize the length of the polypeptide and thus the complexity of the spectroscopic analysis, several possible strategies were tested. First a short 7 amino acid His-tag was added to the 180 residue *S. pombe* core domain, which, however, reduced the purification yield somewhat, presumably due to steric interference. Subsequently, *S. pombe* core domain equipped with the standard 6HisT-tag was purified with satisfactory efficiency, but thrombin cleavage to remove the 20 amino acid tag proved highly inefficient. As early studies (Lieberman et al., 1991) with *Saccharomyces cerevisiae* TBP had shown that the conserved core domain is surprisingly resistant to trypsin digestion (Fig. 50), I attempted to use trypsin for the removal of the His-tag which contains an arginine at residue -4 relative to the endogenous first methionine. Limited trypsin digestion indeed resulted in the apparent removal of the 6HisT-tag, as confirmed by mass spectroscopy measurements (M. Ikura, personal communication).

An unexpected difficulty was encountered when initial spectroscopic results (M. Ikura, personal communication) as well as mass scattering measurements (S. Burley, personal communication) indicated that TBP forms a homodimer at high concentrations. While subsequent association constant

Figure 50. Limited trypsin digestion of yTBP.
6HisT-yTBP was incubated with trypsin (1µg for each 1mg of yTBP) at 25°C and aliquots were removed at indicated times (lanes 1-7) and boiled in SDS-sample buffer. After one hour samples were centrifuged for 20min at 10,000 g to remove precipitate (lane 8) from soluble protein (lane 9). The arrows indicate the positions of full length 6His-yTBP and the 185 amino acid long C-terminal protease resistant fragment containing the conserved core domain.


trypsin time-course
kd
M
0'
15'
15'
30'
30'
60'
60'
ppt
sup
45
31
21
14
1
2
3
4
5
6
7
8
9

measurements have indicated that this is most likely an artefactual complex (D. Nikolov and S. Burley, personal communication), it nevertheless presents a serious difficulty for further NMR studies as the size of the homodimer exceeded the capabilities of current NMR spectroscopy.

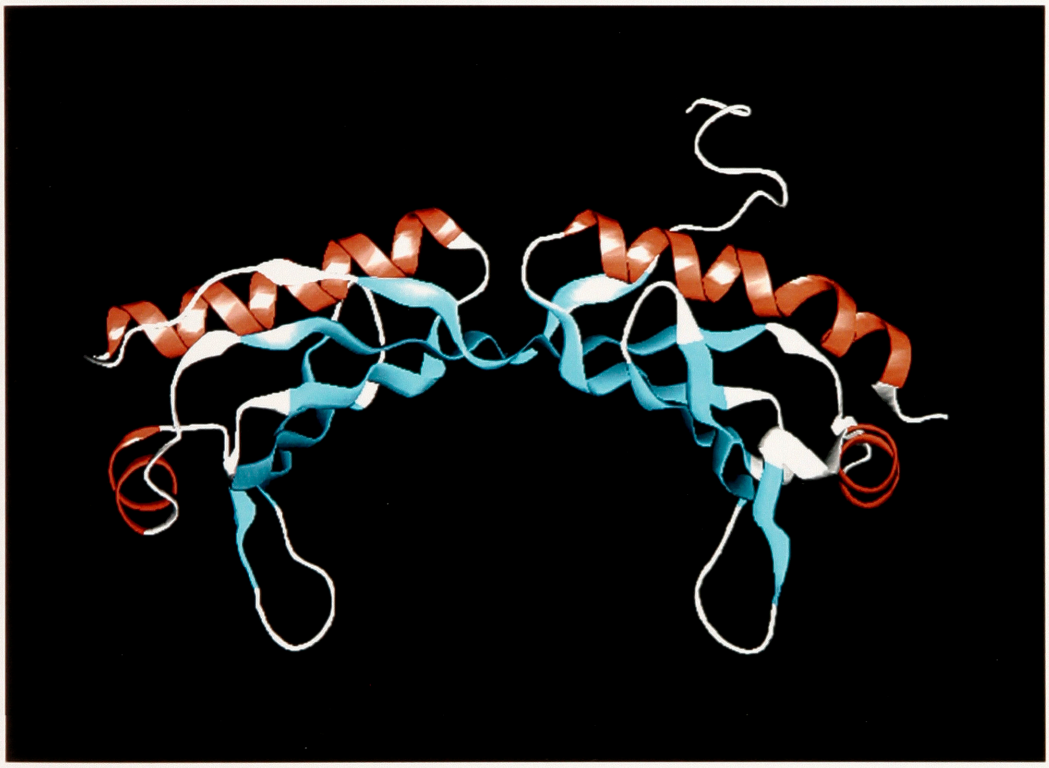
However, the TBP-TATA complex consists of a single polypeptide and thus might be amenable to further study. To this end TATA box-containing double-stranded oligonucleotides derived from the Ad2MLP were analyzed for their capability in quantitatively binding *S. pombe* and *S. cerevisae* TBP core domain by gel mobility shift analysis. Probes as short as 12bp were shown to be quantitavely shifted (data not shown) and have subsequently been used for NMR analysis of the TBP-TATA box complex. While solubility of this complex at high concentrations continues to be a problem (though stability of the complex is not), we anxiously await progress in the ongoing analysis.

X-ray crystallography

The limiting step in X-ray crystallographic studies of bio-molecules is the availability of large, well-ordered crystals. Large amounts of highly purified protein are required to screen a large number of possible buffer conditions for optimal crystal growth. Further, with TBP available from various species an additional parameter could be varied : while their three-dimensional structure was likely to be near-identical, their crystallization behaviour and quality of the final crystal need not be.

His-pET expression vectors were constructed for full-length and core domains of TBPs from *Saccharomyces cerevisiae* (Horikoshi et al., 1989a), *Schizosaccharomyces pombe* (Hoffmann et al., 1990), *Arabidopsis thaliana* 1 and 2 (Gasch et al., 1990), human (Hoffmann et al., 1990), *Drosophila melanogaster* (Muhich et al., 1990), and *Plasmodium falciparum*

Figure 51. Ribbon diagram of At2-TBP.
The three-dimensional structure of a TBP core domain as solved by X-ray crystallography presented here in diagrammatic form, emphasizing the two-fold symmetry of the protein. Beta-strands are coloured in blue, alpha helices in red. See text for details and references (courtecy of D.B. Nikolov and S.K. Burley).



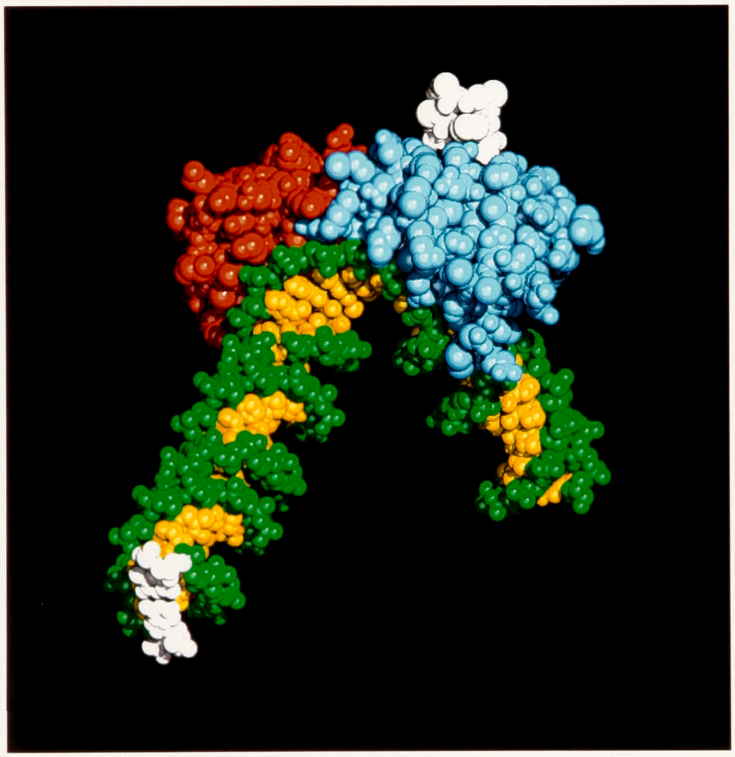
(McAndrew et al., 1993) as well as the *Drosophila* TBP-related factor (TRF) (Crowley et al., 1993). Proteins were purified based on the method described in section (i), but high levels of expression in *E. coli* was achieved only with yeast and plant TBP constructs. Crystallization trials with a number of these purified proteins (with intact tags or cleaved as shown in Fig. 50 for yTBP) resulted in promising crystals with the yTBPcore and At2 TBP which proved to give higher resolution in X-ray diffraction studies (Nikolov et al., 1992).

The solution of the three-dimensional structure of At2 TBP (for details, see Nikolov et al. 1992) represents an exciting step in the molecular characterization of this protein. The structure revealed a novel two-domain DNA-binding fold resulting in saddle shaped protein (Fig. 51) consisting of two domains with an almost identical polypeptide backbone tracing. While sequence analysis had predicted a direct repeat structure separated by an amphipathic alpha helix, the molecule in fact has no such spacer and instead has an approximate two-fold intramolecular symmetry : the basic repeat has its counterpart in the region previously erroneously hypothesized to contain a sigma homology. These two helices create a large convex surface, the seat of the saddle, available for a multiplicity of protein-protein interactions. The helices rest on top of a large curved anti-parallel β-sheet that forms a smooth concave surface whose dimensions are such that it may accomadate B-form DNA.

While mutagenesis studies of TBP have also implicated this concave underside of the molecular saddle, crystallographic studies with the TBP-TATA box complex (Kim et al., 1993) further revealed some highly unusual features. Binding of the saddle shaped molecule to DNA results in conformational changes in the protein fold, but much more dramatically in the structure of the DNA (Fig. 52). The 8bp TATA element is partially unwound exhibiting a dramatic opening of the minor groove, confined by an intercalating

Figure 52. Space filling model of At2-TBP complexed to the promoter.

The three-dimensional structure of a TBP core domain bound to a TATA box containing DNA fragment solved by X-ray crystallography is shown here as a space filling model with the DNA extended to the transcription start site (white). Each half of the symmetric molecule coded for by a previously identified direct repeat is coloured in blue and red. See text for details and references (courtecy of D.B. Nikolov, J.L. Kim, and S.K. Burley).



tyrosine 'clamp' at either end of the cognate sequence. As chemical probing of the TBP-DNA complex had predicted, side-chain/base interactions are restricted to the minor groove (Lee et al., 1991; Starr and Hawley, 1991) which forms a primarily hydrophobic interface with the entire underside of the molecular saddle. Severe bends and a positive writhe to compensate for the underwinding radically alter the trajectory of the flanking B-form DNA as observed gel shift-based DNA bending assays (Horikoshi et al., 1992). Similar studies with yeast TBP by itself (Chasman et al., 1993) or complexed with the yeast CYC1 TATA element (Kim et al., 1993) confirmed the structure and the high level of homology between evolutionary distant TBPs.

(iii) Designing a convenient mutagenesis of TBP

Introduction and rationale

Following the cloning of cDNAs from yeast, human and more than a dozen other organisms, a plethora of studies have not only demonstrated the TBP's central role in all eukaryotic transcription by virtue of being a subunit of each class-specific multi-protein general transcription initiation factor complex, but have also implicated it in activation mechanisms by direct protein interactions with activation surfaces on the one hand and co-activator proteins on the other (for a summary of published interactions see Fig. 53).

Specificity of direct protein-protein interactions are most convincingly demonstrated *in vivo* by classical genetic studies identifying second site suppressor mutations in the interaction partner of a protein whose function has been compromised through a specific mutation, as shown for example for TBP interacting protein BRF (Buratowski and Zhou, 1992; Colbert and Hahn, 1992; Lopez-de-Leon et al., 1992)). In the absence of an efficient genetic approach im mammalian systems, and in light of a wealth of biochemical data resulting from *in vitro* studies, specificity of proposed interactions are commonly confirmed by site-directed amino-acid mutations in either or preferably both interaction partners that specifically abolish the observed interaction. Furthermore, functional relevance of the proposed interaction can only be established by correlating interaction studies with specific mutants with results from functional assays (e.g. activation *in vitro* or *in vivo*). Finally, mutagenesis of surface residues allows the precise mapping of particular interactions that may shed light on functional mechanisms mediated by the proteins involved. Thus overlap of the binding sites of TFIIA and NC2 on the basic repeat of TBP provides the molecular basis for the observed negative or positive regulation of pre-initiation complex formation following TBP binding to the TATA box.

Figure 53. Proteins capable of interacting directly with TBP.

Summary of proteins reported to contact TBP directly and specifically as shown in a variety of *in vitro* and *in vivo* assays as demonstrated in given references.

TBP-interacting Proteins

Class II initiation factors	
TFIIA	Reinberg et al 1987, Lee et al 1992, Buratowski & Zhou 1992
TFIIB	van Dyke et al 1988, Buratowski et al 1989 Kim et al 1994
TFII-I	Roy et al 1993
TFIIE	Maxam & Tjian 1994
polII (CTD)	Dahmus & Kedinger 1983, Usheva et al 1992
SPT3	Winston & Minehart 1986, Eisenmann et al 1992
YY1	Usheva & Shenk 1994
Class II TAFs	
h250 (d230)	Takada et al 1991, Kokubo et al 1994b, Ruppert et al 1993, Weinzierl et al 1993
h80 (d60)	Kokubo et al 1994a, Weinzierl et al 1993
d40	Kokubo et al 1994a
h20/15 (d28/22)	Yokomori et al 1993b, Kokubo et al 1994, Hoffmann, unpublished observations
Class III TAFs	
h96	Hoffmann & Wang, unpublished observations
BRF1/TDS4/Pcf4	Buratowski & Zhou 1992, Colbert & Hahn 1992, Lopez de Leon et al 1992
Class I TAFs	Comai et al 1992
Viral activators	
Ad2 E1A	Horikoshi et al 1991
HTLV1 Tax1	Caron et al 1993
HIV1 Tat	Kashanichi et al 1994
EBV Zta	Lieberman & Berk 1991
HSV VP16	Stringer et al 1990, Ingles et al 1991
hCMV IE2	Hagemeier et al. 1992
BPV1 E2	Ham et al 1994
Cellular activators	
p53	Seto et al 1992, Truant et al 1993, Ragimov et al 1993, Liu et al 1993, Chen et al 1993 Mack et al 1993
c-myc	Maheswaran et al 1994, Hateboer et al 1993
Rel (NFkB)	Kerr et al 1993, Xu et al 1993
Sp1	Emili et al 1994
E2F	Hagemeier et al 1993a
PU-1	Hagemeier et al 1993b
USF	Sawadogo & Roeder 1985
Co-activators	
NC1 (DR2, HMG1)	Ge & Roeder, 1994
NC2 (DR1)	Inostroza et al 1992
ADI	Auble & Hahn 1993
PC3 (topoI)	Merino et al 1993
SRB2	Koleske et al 1992

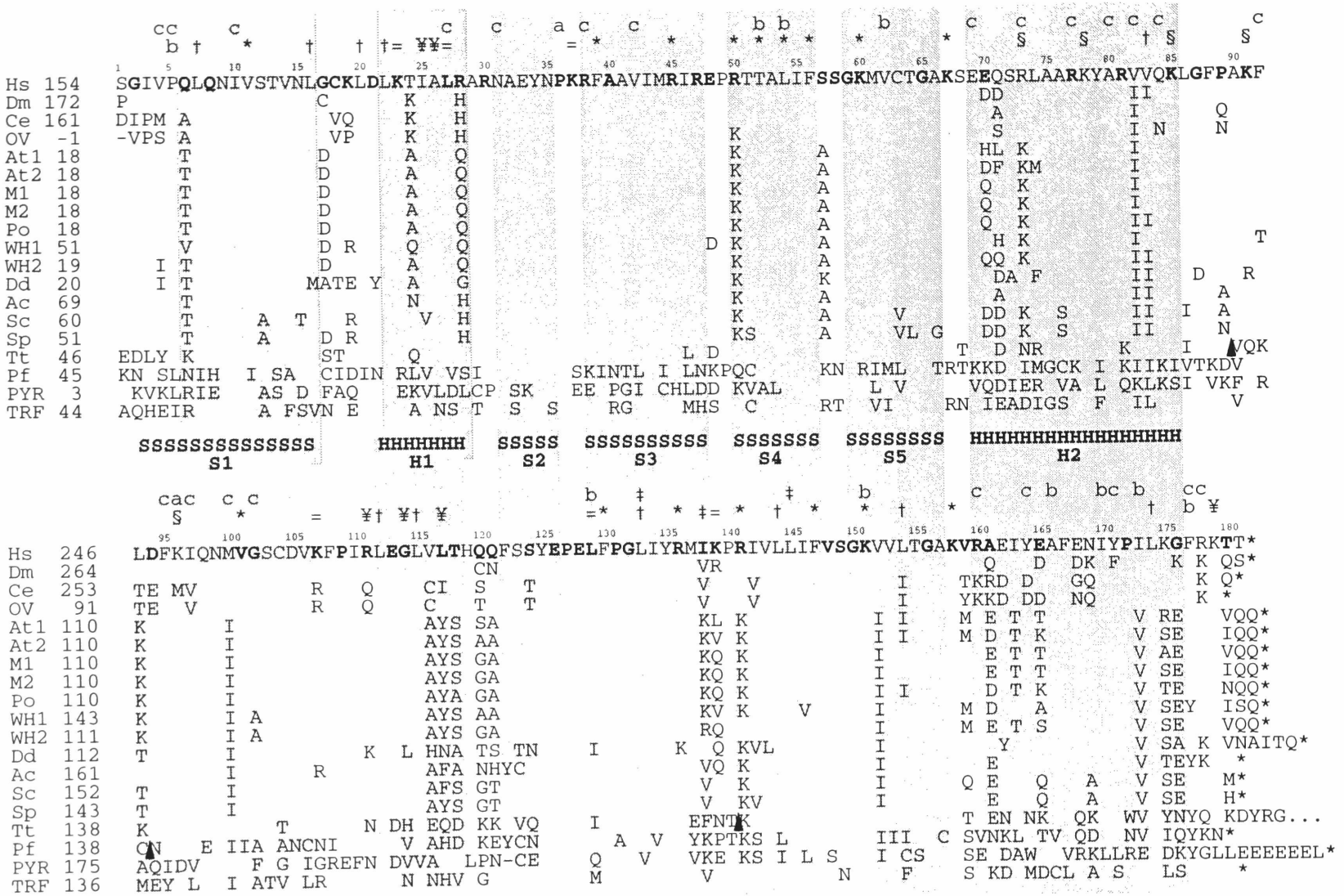
Early mutagenesis studies with TBP were severely hampered by the apparent extreme sensitivity of the protein fold to disrupting mutations. Of close to 200 single amino acid mutants engineered into the yeast TBP core domain by site directed mutagenesis, only about 20 resulted apparently correctly folded proteins (T. Yamamoto and M. Horikoshi, personal communication). Due to the fundamental roles of TBP in regulatory mechanisms central to cellular viability genetic screens in yeast on the other hand, have to be performed either in the background of the wild-type protein, limiting the scope of the investigation to a particular interaction (e.g. DNA binding specificity (Strubin and Struhl, 1992)), or are confined to the isolation of temperature sensitive mutants that generally exhibit aberrations in the hydrophobic core of the protein (Cormack and Struhl, 1993), limiting possible conclusions regarding interaction surfaces.

The three-dimensional structure of the evolutionary conserved TBP core domain has provided the stereochemical framework for a further extensive mutagenesis of this unique protein domain central to transcriptional regulation in eukaryotes. Close inspection of the structure combined with calculations of solvent accessibility allows likely surface amino acid side chains to be identified (bolded in Fig. 54) and targeted for mutagenesis. Furthermore, the availability of currently 19 sequences from across the eukaryotic spectrum that conform to the TBP fold (Fig. 54), can be used as guidance as to the tolerance of the protein fold to disruptive mutations. Finally, the results from numerous mutagenesis studies primarily with yeast TBP (indicated above the hTBP sequence in Fig. 54), in conjunction with molecular modelling should allow the design of mutant TBPs with highly specific functional deficiencies due to defects in identified interaction surfaces.

Figure 54. Summary of TBP sequence data.

Sequences of TBP core domains from TBP clones derived from indicated species. Amino acid residues are only given when different from the human sequence; shading corresponds to secondary structure elements elucidated by X-ray crystallography; bold letters indicate residues that are significantly surface exposed; symbols above the sequence data indicate effects of point mutations as laid out in the key.

Hs, human (Hoffmann et al., 1990b; Kao et al., 1990; Peterson et al., 1990)(sequences for mouse (Tamura et al., 1991) and *Xenopus laevis* (Hashimoto et al., 1992) TBP core are identical to human); Dm, *Drosophila melanogaster* (Hoey et al., 1990; Muhich et al., 1990); Ce, *Caenorhabditis elegans* (Lichtsteiner and Tjian, 1993); OV, *Onchocerca volvulus* (Li and Donelson, 1993); At1 and At2, *Arabidopsis thaliana* type 1 and 2 (Gasch et al., 1990); M1 and M2, maize type 1 and 2 (Haass and Feix, 1992); Po, potato (Holdsworth et al., 1992); WH1, wheat type 1 (Kawata et al., 1992); WH2, wheat type 2 (Apsit et al., 1993); Dd, *Dictyostelium discoideum* (J.E. Blume, D.R. Shaw, H.L. Ennis, Genbank Accession # M64861), Ac, *Acanthamoeba Castellanii* (Wong et al., 1992); Sc, *Saccharomyces cerevisiae* (Horikoshi et al., 1989a; Hahn et al., 1989; Schmidt et al., 1989; Cavallini et al., 1989); Sp, *Schizosaccharomyces pombe* (Hoffmann et al., 1990a; Fikes et al., 1990); Tt, *Tetrahymena thermophila* (Stargell and Gorovsky, 1994); Pf, *Plasmodium falciparum* (McAndrew et al., 1993); PYR, *Pyrococcus woesei* (S. Jackson, personal communication); TRF, *Drosophila* TBP-related factor (Crowley ret al., 1993).


mutants affecting
a pol I
b pol II
c pol III
† lethal
= no effect
‡ DNA recognition
* DNA binding
¥ SPT3 interaction
§ TFIIA interaction
insertion
S1
H1
S2
S3
S4
S5
H2

A cassette encoding the human TBP core domain

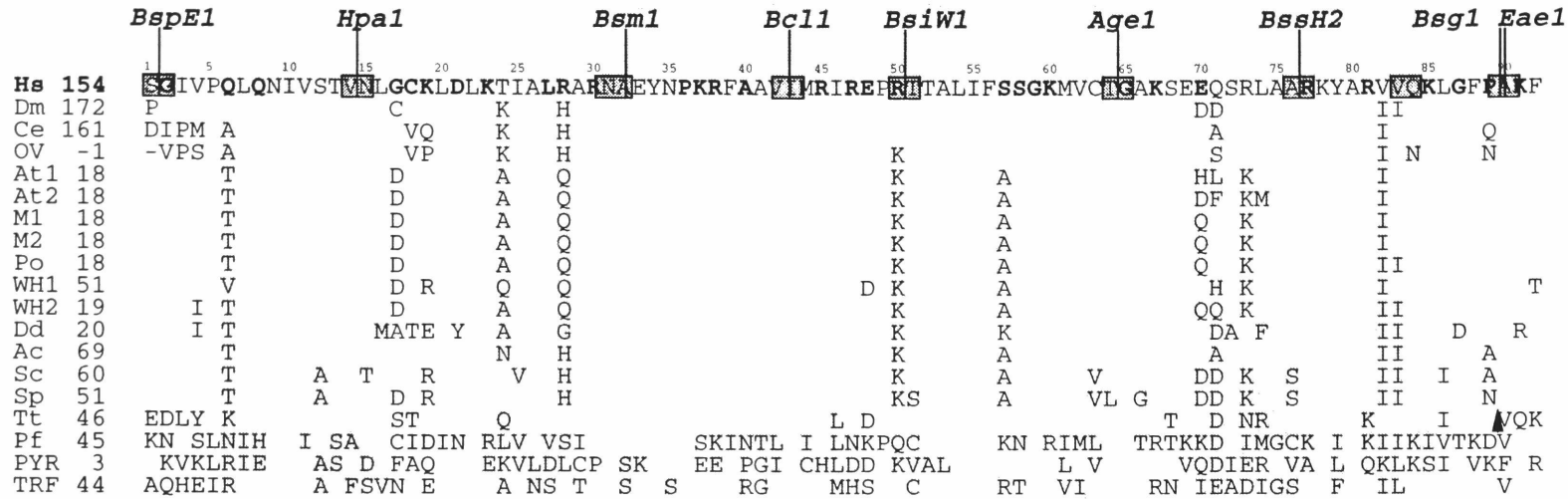
The structure of the conserved TBP core domain resembles a molecular saddle whose extended convex surface accomodates the majority - if not all - protein-protein interactions that this polypeptide engages in. Its repeated subdomain can be conveniently divided into six distinct hypothetical interaction surfaces, made up in part by different secondary structure elemets. Effective mutagenesis of each of the 12 hypothetical interaction surfaces requires the mutation of several surface-exposed residues, a task that requires laborious molecular biological manipulations with conventionally used recombinant DNA techniques (Kunkel's or PCR-mediated site directed mutagenesis methods). A more efficient approach is based on the construction of a cassette of unique restriction sites encoding the wild-type human TBP core domain, that allows subsequent mutagenesis of single or multiple residues within a cassette segment through simple fragment exchange. In addition, single interaction surface mutants can be conveniently combined to generate more severely defective mutants by virtue of common restriction sites and segment exchange within the cassette.

To this end, restriction sites, not present in commonly used cloning and expression vectors, or in the hTBP cDNA, were designed into convenient locations (particularly at secondary structure junctions) within the TBP core domain without affecting the wild-type amino acid sequence by computer-aided codon randomization (Fig. 55). In addition, other not yet determined codons were subsequently specified according to a weighted compromise of most frequently used *Escherichia coli* and human codons for each amino acid, in the hope of ensuring high expression levels in either of these organisms. The resulting sequence contained in the full length human TBP cDNA is shown in Figure 56.

The cassette was constructed using gel-purified oligonucleotides in PCR mediated cloning steps (Materials and Methods). Utilizing the restriction sites created at either

Figure 55. Designing a cassette construct encoding the human TBP core domain for convenient mutagenesis.

Restriction sites present neither in the natural human TBP cDNA nor in commonly used cloning and expression vectors can be engineered into the human TBP core domain without changing the amino acid sequence. Positions of unique restriction sites are shown superimposed on the comprehensive sequence comparison of Figure 54: positions of DNA recognition sequences are indicated by shading; positions of cut sites by vertical lines.


BspE1
Hpa1
Bsm1
Bcl1
BsiW1
Age1
BssH2
Bsg1
Eae1
Hs 154 SGIVPQLQNIVSTVNLGCKLDLKTIALRARNAEYNPKRFAAVIMRIREPRTTALIFSSGKMVCTGAKSEEQSRLAARKYARVVQKLGFPAKF
Dm 172 P C K H DD II
Ce 161 DIPM A VQ K H A I Q
OV -1 -VPS A VP K H K S I N N
At1 18 T D A Q K A HL K I
At2 18 T D A Q K A DF KM I
M1 18 T D A Q K A Q K I
M2 18 T D A Q K A Q K I
Po 18 T D A Q K A Q K II
WH1 51 V D R Q Q D K A H K I T
WH2 19 I T D A Q K A QQ K II
Dd 20 I T MATE Y A G K K DA F II D R
Ac 69 T N H K A A II A
Sc 60 T A T R V H K A V DD K S II I A
Sp 51 T A D R H KS A VL G DD K S II N
Tt 46 EDLY K ST Q L D T D NR K I VQK
Pf 45 KN SLNIH I SA CIDIN RLV VSI SKINTL I LNKPQC KN RIML TRTKKD IMGCK I KIIKIVTKDV
PYR 3 KVKLRIE AS D FAQ EKVLDLCP SK EE PGI CHLDD KVAL L V VQDIER VA L QKLKSI VKF R
TRF 44 AQHEIR A FSVN E A NS T S S RG MHS C RT VI RN IEADIGS F IL V

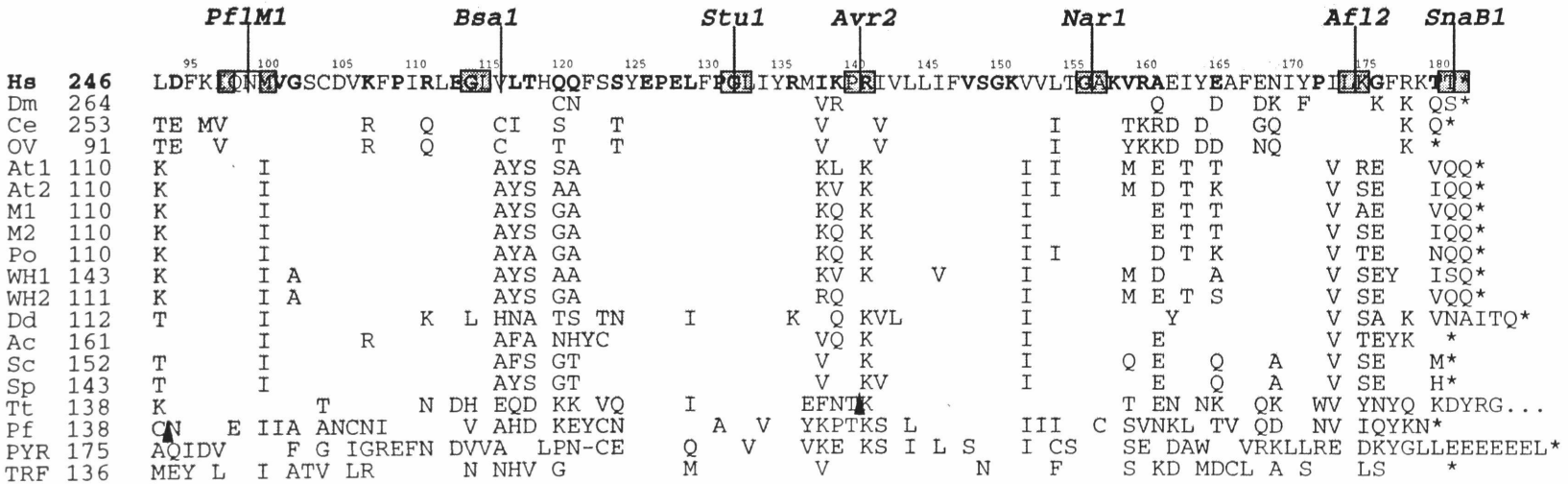
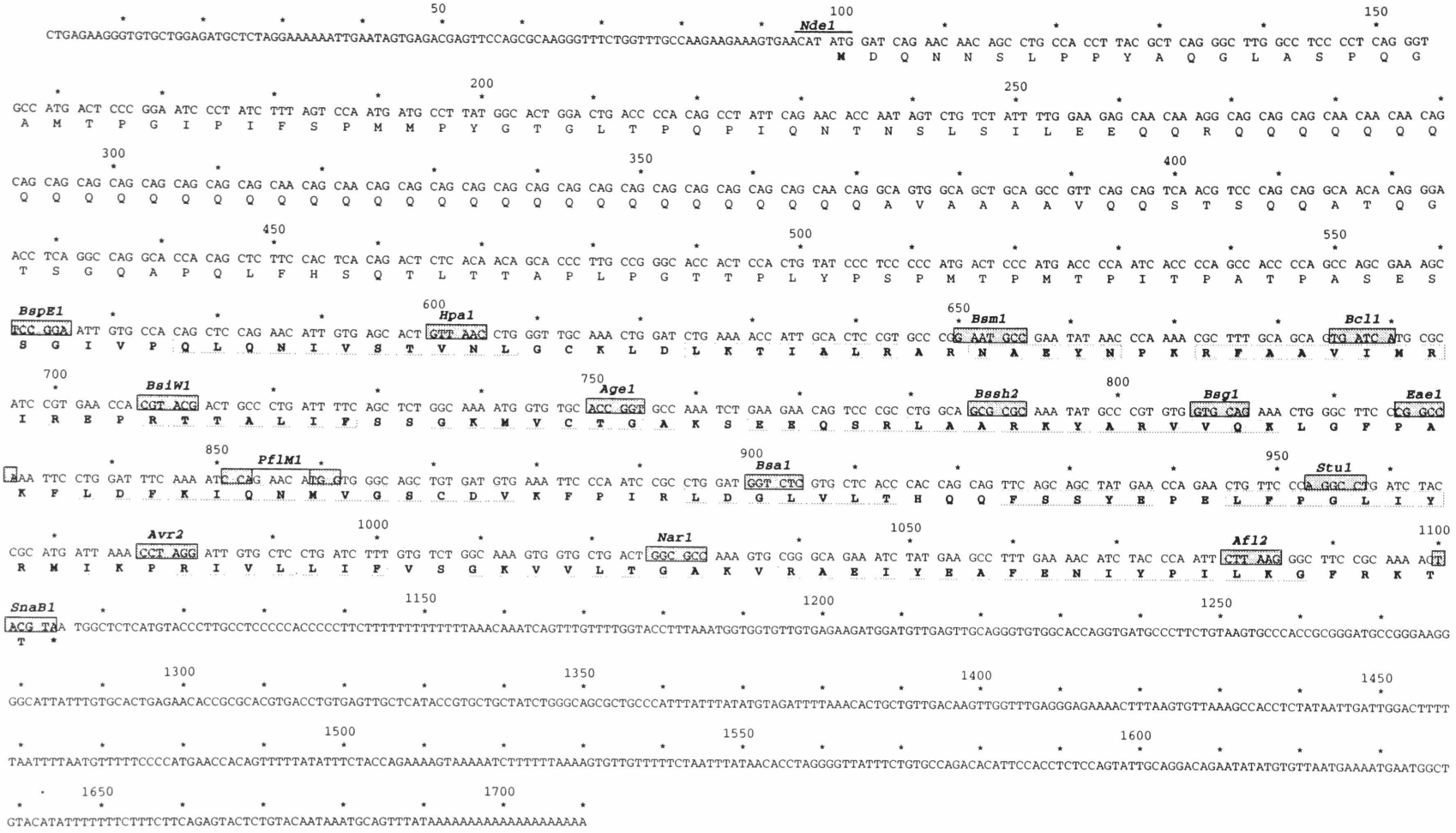

Pf1M1
Bsa1
Stu1
Avr2
Nar1
Af12
SnaB1
Hs 246 LDFKIQNMVGSCDVKFPIRLEGIVLTHQQFSSYEPELFPGLIYRMIKPRIVLLIFVSGKVVLTGAKVRAEIYEAFENIYPILKGFRKTF*
Dm 264 CN VR Q D DK F K K QS*
Ce 253 TE MV R Q CI S T V V I TKRD D GQ K Q*
OV 91 TE V R Q C T T V V I YKKD DD NQ K *
At1 110 K I AYS SA KL K I I M E T T V RE VQQ*
At2 110 K I AYS AA KV K I I M D T K V SE IQQ*
M1 110 K I AYS GA KQ K I E T T V AE VQQ*
M2 110 K I AYS GA KQ K I E T T V SE IQQ*
Po 110 K I AYA GA KQ K I I D T K V TE NQQ*
WH1 143 K I A AYS AA KV K V I M D A V SEY ISQ*
WH2 111 K I A AYS GA RQ I M E T S V SE VQQ*
Dd 112 T I K L HNA TS TN I K Q KVL I Y V SA K VNAITQ*
Ac 161 I R AFA NHYC VQ K I E V TEYK *
Sc 152 T I AFS GT V K I Q E Q A V SE M*
Sp 143 T I AYS GT V KV I E Q A V SE H*
Tt 138 K T N DH EQD KK VQ I EFNTK T EN NK QK WV YNYQ KDYRG...
Pf 138 CN E IIA ANCNI V AHD KEYCN A V YKPTKS L III C SVNKL TV QD NV IQYKN*
PYR 175 AQIDV F G IGREFN DVVA LPN-CE Q V VKE KS I L S I CS SE DAW VRKLLRE DKYGLLEEEEEEL*
TRF 136 MEY L I ATV LR N NHV G M V N F S KD MDCL A S LS *

Figure 56. Sequence of the resulting syn hTBP cDNA.
The DNA fragment encoding the human TBP core domain within the natural TBP cDNA was replaced by synthesized DNA fragment encoding the same amino acid sequence but containing restriction sites shown in Figure 55 and altered codons for all other amino acids to conform to maximum codon usage in *Escherichia coli* and human. The complete DNA sequence of the resulting cDNA as well as the unaltered amino acid sequence of the open reading frame are shown. Shading of amino acids indicate secondary structure elements determined by X-ray crystallography; shading of nucleotides indicate unique, newly engineered recognition sequences of restriction enzymes denoted above.


NdeI
BspE1
Hpa1
Bsm1
Bcl1
BsiW1
Age1
Bssh2
Bsg1
Eae1
PflM1
Bsa1
Stu1
Avr2
Nar1
Afl2
SnaB1

end of the TBP core domain (BstE1 and SnaB1) the cassette was cloned into the background of the wild-type hTBP cDNA which was subsequently inserted into commonly used bacterial or mammalian expression vectors to produce full length or N-terminally truncated hTBP with a variety of convenient N-terminal tags. Expression of this synthetic hTBP core domain in bacteria has resulted in a significant increase in yield : while only about 1mg of hTBP core protein can be recovered after purification from one liter of culture of wild-type construct expressing bacteria, close to 10mg of identical protein was purified with the synthetic hTBP core construct (data not shown).

A systematic mutagenesis of the postulated interaction surfaces is currently in progress. Correct three-dimensional folding of the resulting mutants will first be checked in a simple DNA binding assay (gel shift) that can be accomplished with reticulocyte lysate-expressed protein, before being used for a host of *in vitro* interaction analyses. Particularly interesting mutants, that are, for instance, specifically deficient in interaction with a single TAF, could subsequently be used in co-transfection assays and or to establish constitutively expressing cell lines that might allow for the purification of partial TBP-containing complexes, possibly leading to a more detailed functional characterization of TAFs and activation mechanisms.

Chapter IV : Characterization of TBP-associated TFIID subunits : TAF20/15

Introduction

The molecular characterization of TFIID experienced a significant step forward with the identification of the polypeptide subunits that make up this TBP-containing complex (chapter II). Furthermore, purification of TFIID in non-denaturing conditions allowed a more conclusive functional investigation of this multi-protein complex, and substantiated comparative analyses with the single polypeptide TATA binding protein (TBP). While these TBP-associated factors (TAFs) as a group have long been implicated in co-activator function of TFIID (Hoffmann et al., 1990; Pugh and Tjian, 1990) and are required for transcription from TATA-less promoters (Pugh and Tjian, 1991; Martinez et al., 1994), little is still known about the functional roles of individual TAFs in transcription. Molecular cloning and characterization of class II TAFs leading to the elucidation of their functions may also contribute to a more detailed characterization of TFIID, and is therefore of immediate interest.

(i) Primary structure analysis of human TAF20/15

Cloning of human cDNAs encoding TAF20 and TAF15

Purification of TFIID and separation of its polypeptide components by SDS-PAGE allowed for micro-sequencing of peptides derived from individual TAFs (Chiang et al., 1993). One peptide sequence derived from a TAF with SDS-PAGE mobility of 20kd was used to design oligo-nucleotides (best-guessmers) to screen a human cDNA library. Overlapping clones were sequenced and revealed an open reading frame of 161 amino acids (with a calculated molecular mass of 17,913 Daltons) that contained the peptide sequence near its C-terminus (Fig. 57). Curiously, only the second ATG within the open reading frame was set within a good Kozak consensus, possibly indicating multiple translation initiation sites. Translation in reticulocyte lysates of mRNAs derived from a cDNA clone spanning the complete open reading frame, resulted in the specific appearance of two polypeptides, one of 20kd the other of 15kd (data not shown), lending suport to the idea that multiple gene products are expressed from a single gene.

To verify that the polypeptide(s) encoded by cloned cDNAs is in fact a subunit of TFIID, antibodies were generated in rabbits against the full length protein or fragments thereof, expressed in, and purified from, bacteria as described (Hoffmann and Roeder, 1991). Western blot analysis with high titer sera or antigen purified antibodies (data not shown) confirmed the presence of immuno-reactive material in both a phosphocellulose D fraction as well as purified TFIID preparations (Fig. 58), but little in other phosphocellulose fractions (data not shown). Furthermore, such antibodies detected the presence of two polypeptides of SDS-PAGE mobility equivalent to 20kd and 15kd, and indistinguishable from mobilities of bacterially expressed open reading frames starting with the first or the second ATG of cloned cDNAs (Fig. 58). These results confirm that two

Figure 57. cDNA and predicted amino acid sequence of human TAF20.

Shown is the sequence of several independent cDNA clones. Peptide sequence obtained from micro-sequencing of proteolytic fragments derived from native human TAF20 is underlined. An ATG within a Kozak consensus context is underlined; predicted translation start and stop codons are indicated by boxes.

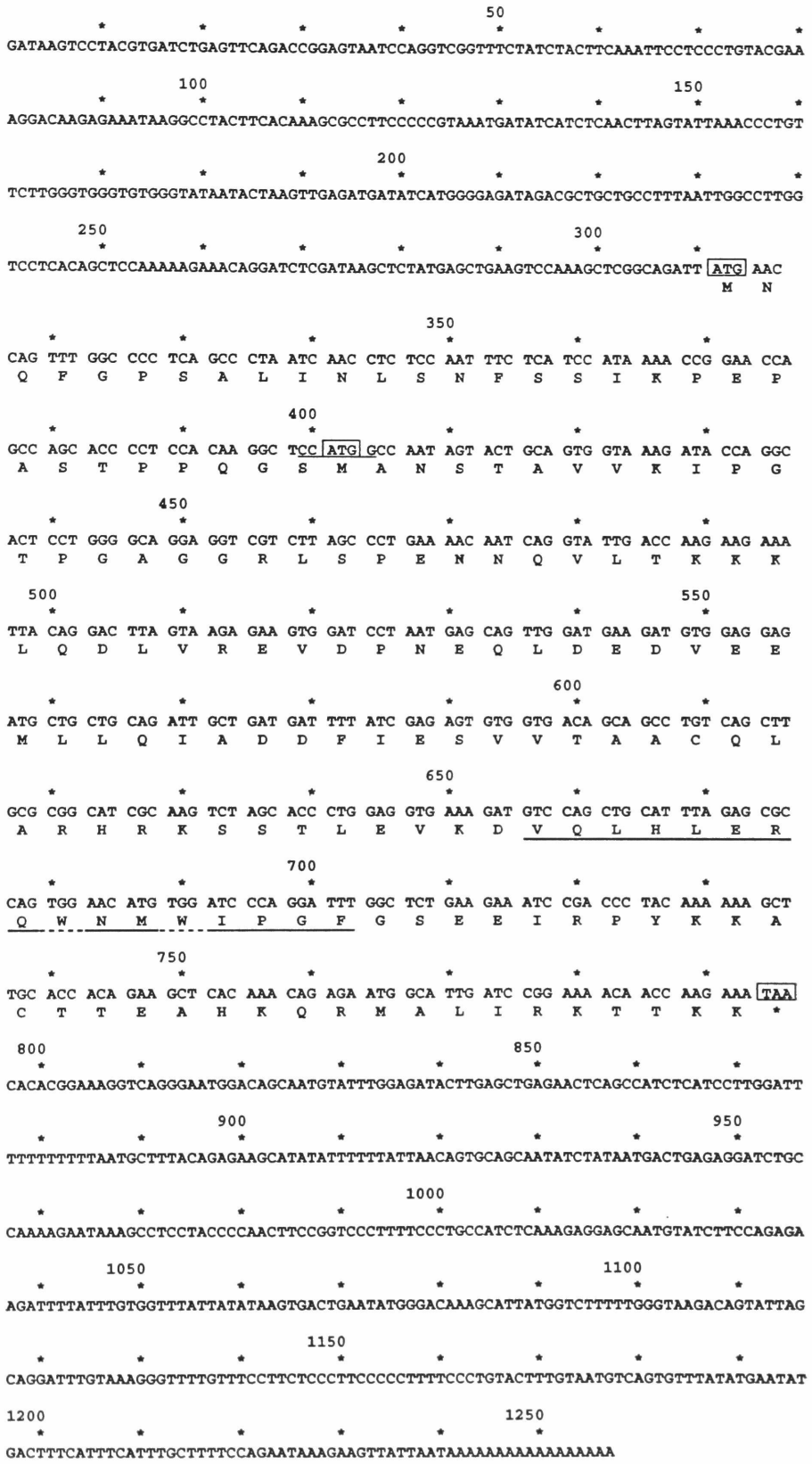

50
GATAAGTCCTACGTGATCTGAGTTCAGACCGGAGTAATCCAGGTCGGTTTCTATCTACTTCAAATTCCTCCCTGTACGAA
100
150
AGGACAAGAGAAATAAGGCCTACTTCACAAAGCGCCTTCCCCCGTAAATGATATCATCTCAACTTAGTATTAAACCCTGT
200
TCTTGGGTGGGTGTGGGTATAATACTAAGTTGAGATGATATCATGGGGAGATAGACGCTGCTGCCTTTAATTGGCCTTGG
250
300
TCCTCACAGCTCCAAAAAGAAACAGGATCTCGATAAGCTCTATGAGCTGAAGTCCAAAGCTCGGCAGATT ATG AAC
M N
350
CAG TTT GGC CCC TCA GCC CTA ATC AAC CTC TCC AAT TTC TCA TCC ATA AAA CCG GAA CCA
Q F G P S A L I N L S N F S S I K P E P
400
GCC AGC ACC CCT CCA CAA GGC TCC ATG GCC AAT AGT ACT GCA GTG GTA AAG ATA CCA GGC
A S T P P Q G S M A N S T A V V K I P G
450
ACT CCT GGG GCA GGA GGT CGT CTT AGC CCT GAA AAC AAT CAG GTA TTG ACC AAG AAG AAA
T P G A G G R L S P E N N Q V L T K K K
500
550
TTA CAG GAC TTA GTA AGA GAA GTG GAT CCT AAT GAG CAG TTG GAT GAA GAT GTG GAG GAG
L Q D L V R E V D P N E Q L D E D V E E
600
ATG CTG CTG CAG ATT GCT GAT GAT TTT ATC GAG AGT GTG GTG ACA GCA GCC TGT CAG CTT
M L L Q I A D D F I E S V V T A A C Q L
650
GCG CGG CAT CGC AAG TCT AGC ACC CTG GAG GTG AAA GAT GTC CAG CTG CAT TTA GAG CGC
A R H R K S S T L E V K D V Q L H L E R
700
CAG TGG AAC ATG TGG ATC CCA GGA TTT GGC TCT GAA GAA ATC CGA CCC TAC AAA AAA GCT
Q W N M W I P G F G S E E I R P Y K K A
750
TGC ACC ACA GAA GCT CAC AAA CAG AGA ATG GCA TTG ATC CGG AAA ACA ACC AAG AAA TAA
C T T E A H K Q R M A L I R K T T K K *
800
850
CACACGGAAAGGTCAGGGAATGGACAGCAATGTATTTGGAGATACTTGAGCTGAGAACTCAGCCATCTCATCCTTGGATT
900
950
TTTTTTTTTAATGCTTTACAGAGAAGCATATATTTTTTATTAACAGTGCAGCAATATCTATAATGACTGAGAGGATCTGC
1000
CAAAAGAATAAAGCCTCCTACCCCAACTTCCGGTCCCTTTTCCCTGCCATCTCAAAGAGGAGCAATGTATCTTCCAGAGA
1050
1100
AGATTTTATTTGTGGTTTATTATATAAGTGACTGAATATGGGACAAAGCATTATGGTCTTTTTGGGTAAGACAGTATTAG
1150
CAGGATTTGTAAAGGGTTTTGTTTCCTTCTCCCTTCCCCCTTTTCCCTGTACTTTGTAATGTCAGTGTTTATATGAATAT
1200
1250
GACTTTCATTTCATTTGCTTTTCCAGAATAAAGAAGTTATTAATAAAAAAAAAAAAAAAAA

Figure 58. Western blot analysis of TAF20/15 in human TFIID.

Antibodies against recombinant TAF20 (N-terminal 96 residues) are used to probe proteins present in a phosphocellulose $P11_{0.85}$ fraction (lane 2) and in an immuno-purified native TFIID complex (lane 3) following SDS-PAGE and electro-transfer onto a nitrocellulose membrane. Arrows indicate specific bands whose mobility is indistinguishable from bacterially expressed (lane 1) TAF20 (161 residues, upper arrow) and TAF15 (C-terminal 131 residues of the open reading frame, lower arrow); asterisk indicates the position of recombinant protein consisting of the C-terminal 109 residues of the TAF20 open reading frame (lane 1).

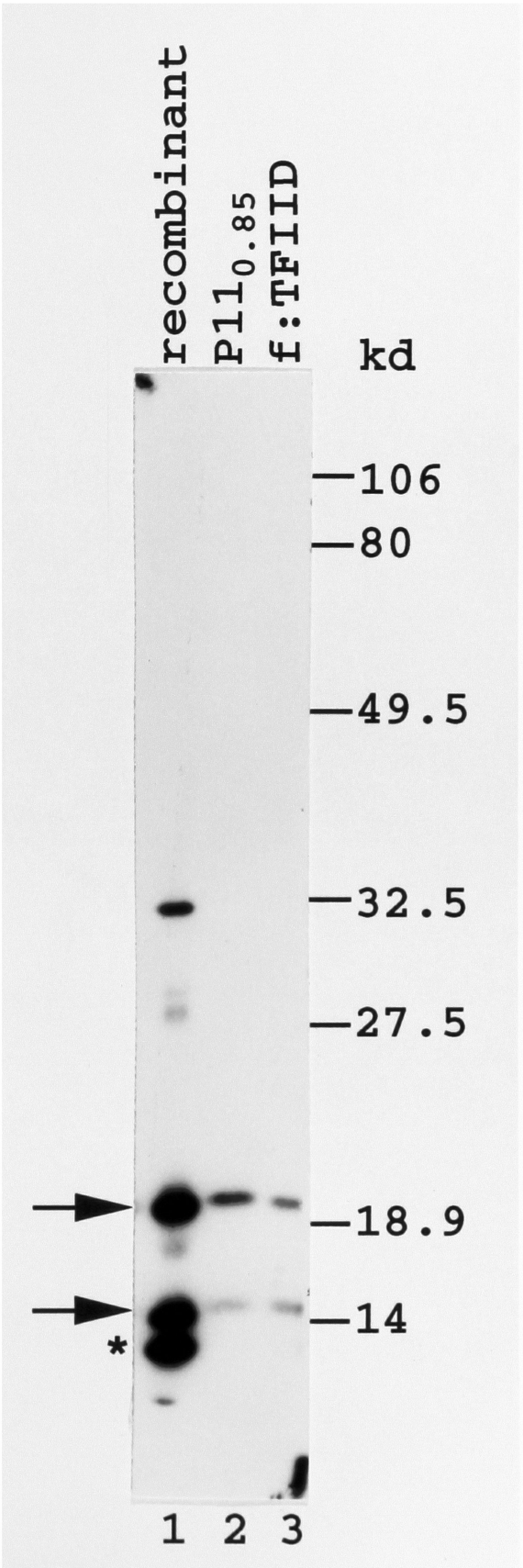
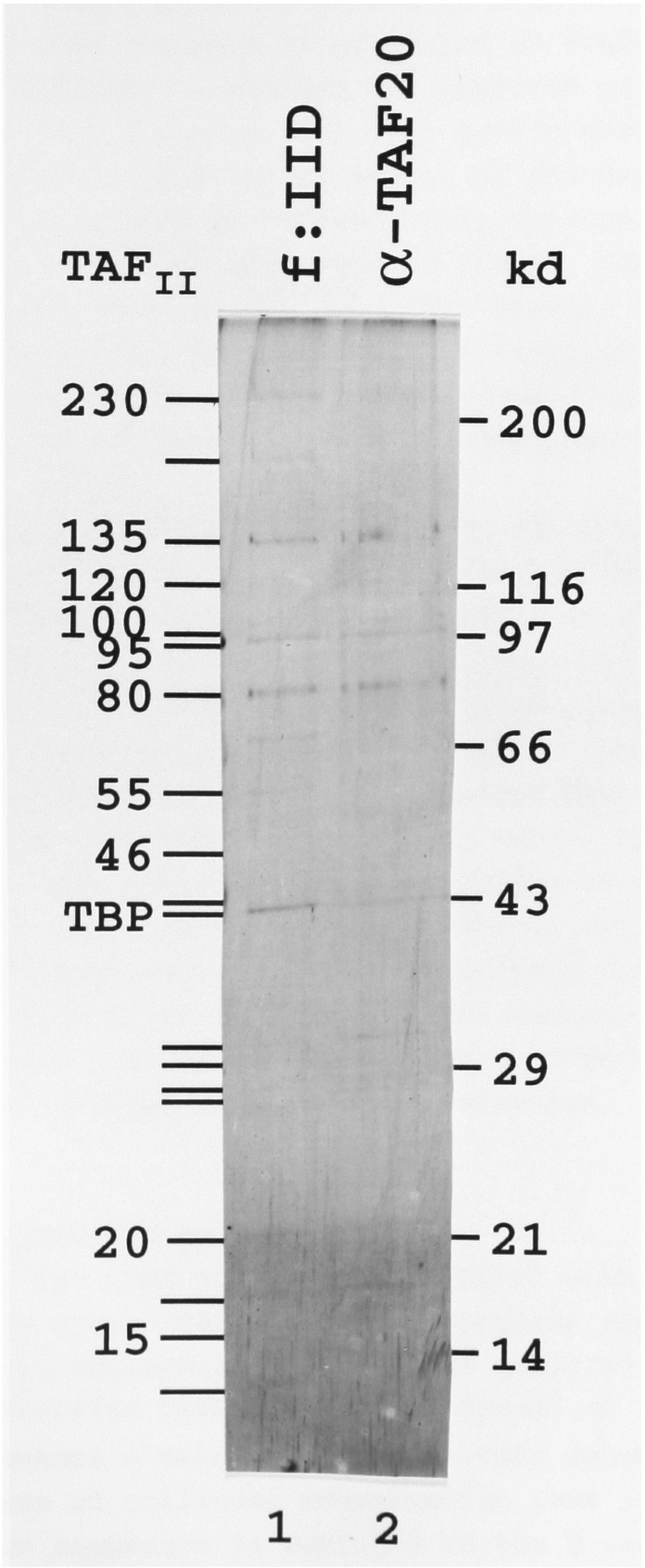

recombinant
P11$_{0.85}$
f:TFIID
kd
106
80
49.5
32.5
27.5
18.9
14
*
1
2
3

Figure 59. Immunoprecipitation of TFIID with α-TAF20 antibodies.

Antigen-affinity-purified antibodies against TAF20 (N-terminal 96 residues) were used to immuno-purify TAF20-containing complexes from the $P11_{0.85}$ fraction (lane 2) and were compared to polypeptides contained in an immuno-purified TFIID complex preparation (lane 2). A silver-stained SDS-PAG (7-17% gradient) is shown with molecular weight markers on the right and positions of previously identified class II TAFs indicated on the left.


f:IID
α-TAF20
TAF$_{II}$
kd
230
135
120
100
95
80
55
46
TBP
20
15
200
116
97
66
43
29
21
14
1
2

low molecular weight TAFs are generated from a single gene, whose derived cDNA sequence is presented in Figure 57. What is currently unknown, is whether the observed polypeptides are generated from a single cDNA in a rarely observed internal translation initiation event, or whether the shorter product is derived from an alternatively spliced message that encodes only the shorter open reading frame. Comparison of the relative abundance of the two forms of this TAF in various cell lines and tissues have not revealed dramatic differences; the longer TAF20 is between two- to ten-fold more abundant in all mammalian tissues examined so far (data not shown).

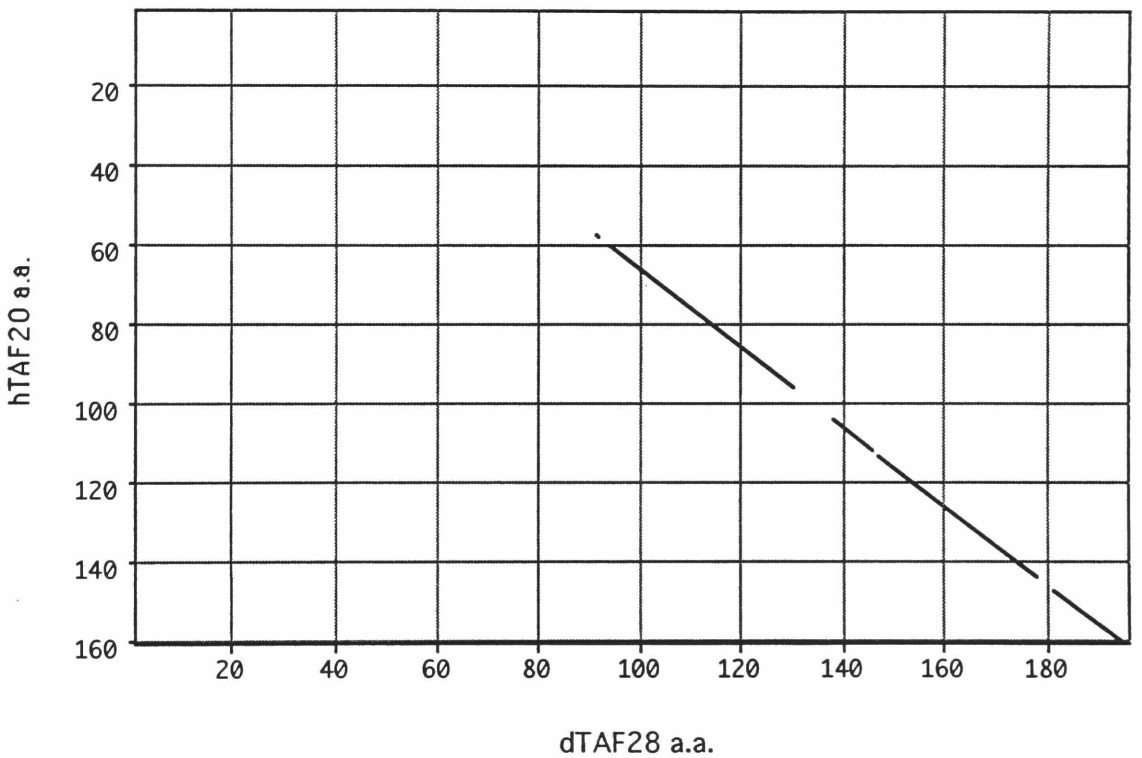
Antigen-purified antibodies against the hydrophilic N-terminal part of TAF20/15 were immobilized on sepharose resin and used for immuno-precipitation studies. As expected, immuno-precipitation of TAF20/15 from phosphocellulose D fractions led to the co-precipitation of an array of polypeptides, previously identified as TFIID subunits (Fig. 59), providing the most convincing evidence that cloned TAF20/15 are in fact *bona fide* class II TAFs. Careful inspection of the silver-stained gel reveals an apparently stoichiometric abundance of purified TAFs in TFIID by both TBP and TAF20 immuno-purification, consistent with the idea of a single large TFIID complex. Larger amounts of TBP are, however, apparent in the f:IID fraction indicating a portion of free TBP of unknown physiological relevance.

Sequence analysis and homologies

Comparison of the cDNA sequence of TAF20/15 with sequences of other recently cloned TBP-associated proteins revealed high similarities to *Drosophila melanogaster* TAF28/22 (Kokubo et al., 1994) otherwise called TAF30α (Yokomori et al., 1993b). Figure 60 presents a matrix comparison that demonstrates that the high degree of collinear conservation (80% identity) beween the two sequences is confined to the C-terminal 109

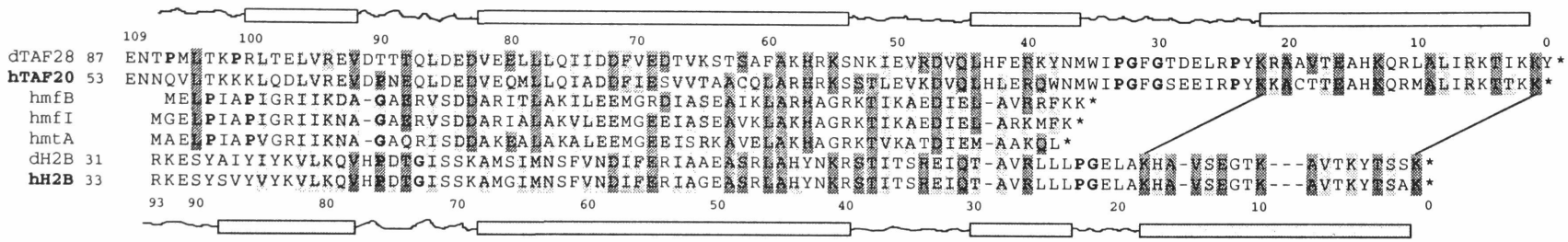
Figure 60 Similarity matrix between human TAF20 and *Dosophila* TAF28.

Human TAF20 amino acid sequence (vertical) is compared with *Drosophila* TAF28 amino acid sequence (horizontal) by the dot matrix method at the indicated stringency.



Window Size = 8
Min. % Score = 60
Hash Value = 1

Figure 61. TAF20 shows sequence homology to histone H2B. Amino acid sequences are aligned of portions of *Drosophila* TAF28, human TAF20, archaebacterial histone H2B like proteins hmfB, hmfI, and hmtA, as well as of *Drosophila* and human histone H2B. Residues identical between either TAFs and any of the shown histone or histone-like proteins at equivalent positions are shaded strongly; similar residues (in the following groupings: D,E,N; E,D,Q; K,R,H; L,T; L,V,I,A,M; P,G; S,T; C,S; F,Y) are weakly shaded; potential helix breakers (amino acids P and G) are bolded. Above, empty boxes indicate alpha-helices predicted by secondary structure analysis of the hTAF20 and dTAF28 sequences; below, shaded boxes indicate alpha-helices predicted by an X-ray crystallographic analysis of the nucleosome core octamer (Arents et al., 1991).


109 100 90 80 70 60 50 40 30 20 10 0
dTAF28 87 ENTPMLTKPRLTELVREVDTTTQLDEDVEELLLQIIDDFVEDTVKSTSAFAKHRKSNKIEVRDVQLHFERKYNMWIPGFGTDELRPYKRAAVTEAHKQRLALIRKTIKKY*
hTAF20 53 ENNQVLTKKKLQDLVREVDENEQLDEDVEQMLLQIADDFIESVVTAACQLARHRKSSTLEVKDVQLHLERQWNMWIPGFGSEEIRPYKKACTTEAHKQRMALIRKTTKK*
hmfB ME LPIAPIGRIIKDA-GAERVSDDARITLAKILEEMGRDIASEAIKLARHAGRKTIKAEDIEL-AVRRFKK*
hmfI MGELPIAPIGRIIKNA-GAERVSDDARIALAKVLEEMGEEIASEAVKLAKHAGRKTVKATDIEM-ARKMFK*
hmtA MAELPIAPVGRIIKNA-GAQRISDDAKEALAKALEEMGEEISRKAVELAKHAGRKTVKATDIEM-AAKQL*
dH2B 31 RKESYAIYIYKVLKQVHPDTGISSKAMSIMNSFVNDIFERIAAEASRLAHYNKRSTITSREIQT-AVRLLLPGELAKHA-VSEGTK---AVTKYTSSK*
hH2B 33 RKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRLAHYNKRSTITSREIQT-AVRLLLPGELAKHA-VSEGTK---AVTKYTSAK*
93 90 80 70 60 50 40 30 20 10 0

residues. The N-termini though similarly unstructured (by secondary structure predictions) and hydrophilic, are of different length and show no sequence relationship. Interestingly, though, in both human and *Drosophila* this TAF is produced in a short and a long form (TAF20/15 and TAF28/22), suggesting a possible functional significance for the N-terminal extension.

Detailed database searches did not identify any previously characterized protein sequences with high degrees of homology. More careful inspection of the search data, however, indicated weak sequence similarities with an abundance of archaebacterial histone-like proteins as well as histone H2B sequences. These indications led me to attempt an alignment between such histone-like sequences with the C-terminal polypeptide segement conserved between *Drosophila* TAF28/22 and human TAF20/15 (Fig. 61). As is now evident, a low level of sequence similarity is present along the whole length of the globular domain of histone H2B and the conserved domain of TAF20, allowing for a few amino acid insertions. Furthermore, independent analyses by computer-aided methods led to very similar secondary structure predictions for dTAF28, hTAF20 and histone H2B within the domain in question (data not shown). While a sequence similarity between TAF20 and H2B of 15% (counting only amino acid identities) or 40% (allowing for conservative exchanges) will not seem convincing, the number of residues in hTAF20 conserved with respect to any one of the family members indicated in Figure 61 may indicate significance and, possibly, a true homology relationship, that may be tested by interaction studies with mutant proteins.

(ii) Mapping protein-protein interactions

Protein-protein interactions within the TFIID complex

A first level of analysis of this novel TFIID subunit may be to identify direct contacts with other subunits that may shed light onto the spatial configuration of the complex and the location of the TAF20/15 within it, and which might possibly give clues to the functional role of TAF20/15 in transcription.

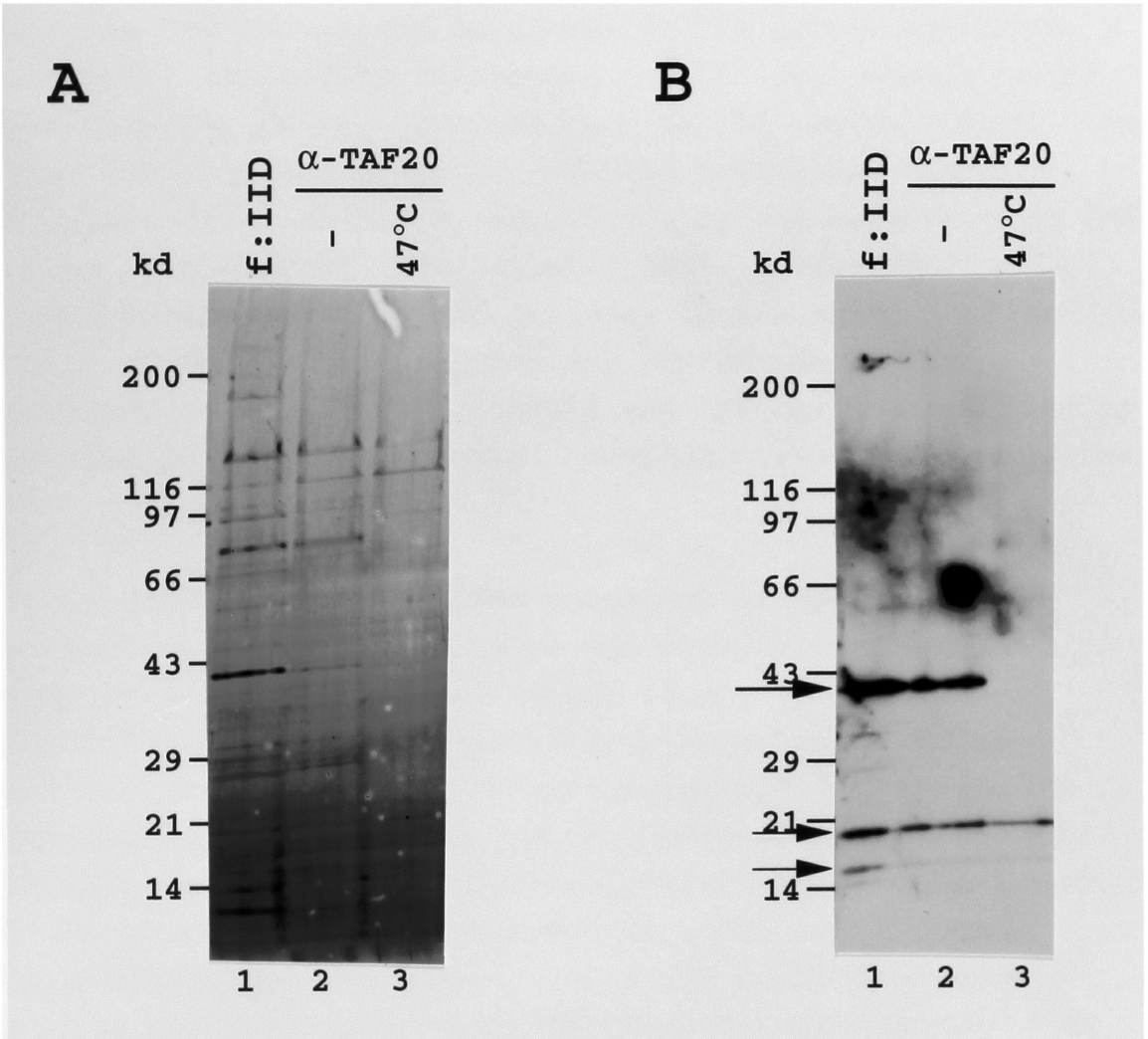
Previous studies utilizing the detergent sarkosyl to partially disrupt the native TFIID complex had revealed that TAF20 is one of the more tightly TBP-associated proteins (chapter II(iv)): such titrations suggested a direct interaction between TBP and TAF20 that is resistant to 0.5% sarkosyl. Other potential protein-protein interaction inhibitors, however, had a largely denaturing effect on TBP and therefore were not useful in TBP-targeted immuno-precipitations. With the availability of a new handle onto the TFIID complex it was now possible to ask what class II TAFs make stable contacts with TAF20/15, independent of TBP's presence in the complex.

A relatively mild but efficient method for TBP denaturation is a 15 minute heat treatment at 47°C (Nakajima et al., 1988). Immuno-precipitations with described α -TAF20 resin were carried out following such heat treatment and compared to intact immuno-precipitated native TFIID complexes. SDS-PAGE and silver staining (Fig. 62A) revealed the loss of a number of polypeptides in heat-treated complexes, notably TAF100, TAF95, TAF80, but also the retention of TAF135 and TAF120. Western blot analysis with available anti-sera confirmed the absence of TBP in this partial complex, which was confirmed to contain similar amounts of TAF20/15 as the native complex (Fig. 62B). These data suggest a possible direct TBP-independent interaction of TAF20/15 with TAF150 and TAF120, in addition to the previously described TBP contact.

Figure 62. Identifying TBP-independent protein interactions within native TFIID.

A. Silver-stained SDS-PAG (7-17% gradient) of immuno-purified TBP-containing (lane 1) and TAF20-containing complexes from untreated (lane 2) or heat-treated (47°C for 15min, lane 3) phosphocellulose $P11_{0.85}$ fractions using anti-FLAG M2 monoclonal antibodies directed against FLAG-hTBP (lane 1) or epitope-affinity-purified antibodies against TAF20 (N-terminal 96 residues, lanes 2 and 3).

B. Western blot of a duplicate SDS-PAG to the one shown in A. probed with α -TAF20, α -TBP, and α -TAF230 polyclonal rabbit serum. The thick arrow indicates the position of TBP, thin arrows those of TAF20 and TAF15.


A
f:IID
α-TAF20
-
47°C
kd
200
116
97
66
43
29
21
14
1
2
3
B
f:IID
α-TAF20
-
47°C
kd
200
116
97
66
43
29
21
14
1
2
3

While TAF250 stains weakly with silver, precluding definitive conclusions about its presence, specific α-TAF250 antibodies did not detect this polypeptide in heat treated or native TFIID complexes immuno-precipitated with α-TAF20 resin (Fig. 62B). This observations may be the result of a disruptive effect by α-TAF20 antibodies on the association of TAF250, or may also indicate the presence of multiple TBP-containing complexes in the phosphocellulose D fraction. While this intriguing possibility remains to be investigated, we may conclude that TAF120 and/or TAF150, the human homologue of *Drosophila* TAF110 previously described to be tethered to the TFIID complex via an interaction with TBP-interacting dTAF230 (Hoey et al., 1993), makes additional stable contacts with TFIID subunits in the absence of TAF250. One of these may be TAF20/15; its postulated direct interactions with TAF120, TAF150 and TBP may be confirmed in interaction assays utilizing individual recombinant proteins.

Construction and interaction potential of TAF20 mutants

Two main objectives were to be addressed with the construction of mutants of TAF20: first, it should be possible to demonstrate specificity in proposed direct protein-interactions and those interactions may be mapped to more defined regions of TAF20; and second, mutants designed according to secondary structure predictions may lend support to the proposed homology between TAF20/15 and H2B whose three-dimensional structure within the nucleosome core complex has been under study by X-ray crystallography. As these proteins, postulated to be structurally related in their C-terminal domains, contain N-termini of very different lengths, amino acids are counted from the common C-terminus of TAF20/15 and denoted as such in the following discussion.

Based on computer-aided sequence analysis, the secondary structure of the conserved 109 residue C-terminal domain of TAF20 was predicted to be largely alpha-helical. Presumably

structurally more flexible loops that connect defined secondary structure elements are more likely the site of insertions or deletions in structurally related proteins and are therefore generally seen as appropriate breakpoints in the design of deletion mutants. A set of deletion mutants of TAF20 was designed with such breakpoints set at C20, C36, C53, C89 and C109 (Fig. 61). An additional breakpoint at C65 was chosen because of a notable sequence divergence in this region between dTAF28/22 and hTAF20/15 (Fig. 60) that may be indicative of structural flexibility. Using a PCR-aided cloning strategy, all combinations of fragments were generated and cloned into vectors for the bacterial expression of resulting protein fragments as GST-fusion proteins that allow for a convenient analysis of protein-protein interactions *in vitro*.

Interaction studies were set up to define the shortest fragment of TAF20 capable of interacting efficiently with recombinant proteins to be tested. To allow for a semi-quantitative analysis, amounts of TAF20 fragment-GST fusion proteins were approximately equalized by estimation of protein amounts from Coomassie Blue-stained SDS-gels (Fig. 63A). Interactions were probed *in vitro* in reactions containing a 1000-fold excess of bacterial proteins as non-specific inhibitors. Precipitates were washed extensively in buffers containing high salt as well as the detergents NP-40 (0.1%) and sarkosyl (0.02%), before probing them, following SDS-PAGE, by Western analysis for the presence of co-precipitating proteins.

Initially, full length human TBP and the N-terminally truncated TBP core domain were shown, as expected, to interact directly with TAF20/15 at high stringency in preliminary experiments using the His-tag interaction method (chapter III(i), data not shown). These were now similarly analyzed for their ability to interact with TAF20 fragments (Fig. 63B). The N-terminus of human TBP, however, caused considerable background and was thus a less reliable

Figure 63. Mapping interaction domains of TAF20.

A. Coomassie-blue-stained SDS-PAG of TAF20-derived GST-fusion proteins (lanes 3-30) after their expression in *E. coli* and employment in interaction assays ('GST pull-down'). Lane 1 contains the protein mixture before 'GST pull-down', showing primarily bacterial proteins used as non-specific competitor in these reactions; lane 2 contains a molecular weight marker. See text and Figure 64 for mutant protein nomenclature.

B. Western blot of SDS-PAG loaded identically to the one shown in A. following a 'GST pull-down' assay with bacterially expressed His-hTBP and His-hcore, probed with α -hcore polyclonal rabbit antiserum. Arrows indicate the mobilities of His-hTBP and His-hcore.

C. Western blot of SDS-PAG loaded identically to the one shown in A. following a 'GST pull-down' assay with FLAG-hTAF80 expressed in the baculovirus system, probed with anti-FLAG M2 monoclonal antibody. Arrow indicate the mobility of FLAG-hTAF80.

D. Western blot of SDS-PAG loaded identically to the one shown in A. following a 'GST pull-down' assay with bacterially expressed FLAG-hTAF20, probed with anti-FLAG M2 monoclonal antibody. Arrow indicate the mobility of FLAG-hTAF20.

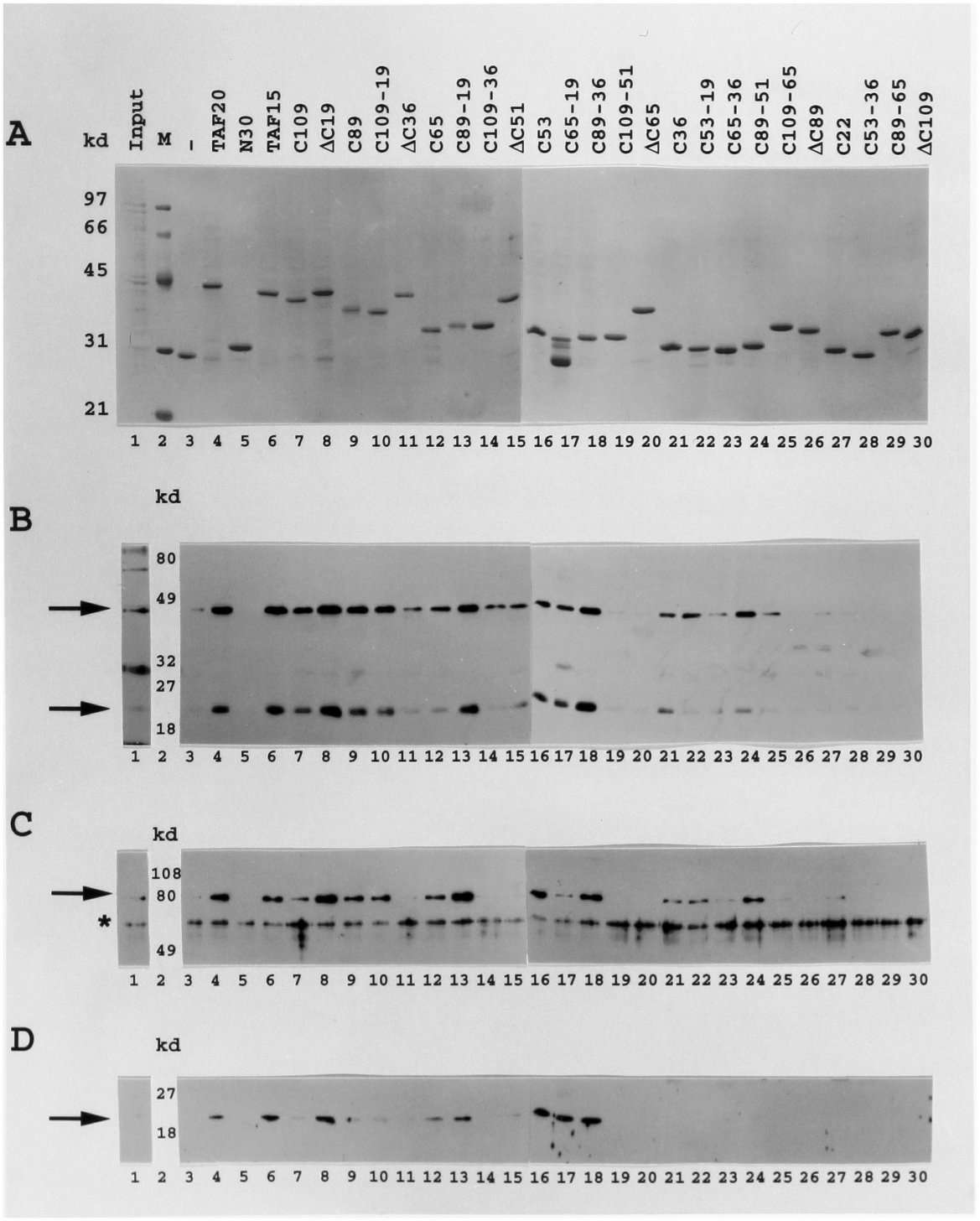
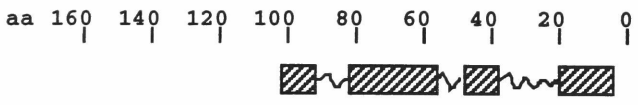

A
kd
Input
M
-
TAF20
N30
TAF15
C109
ΔC19
C89
C109-19
ΔC36
C65
C89-19
C109-36
ΔC51
C53
C65-19
C89-36
C109-51
ΔC65
C36
C53-19
C65-36
C89-51
C109-65
ΔC89
C22
C53-36
C89-65
ΔC109
97
66
45
31
21
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
B
kd
80
49
32
27
18
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
C
kd
108
80
49
*
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
D
kd
27
18
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Figure 64. Summary of interaction studies.
GST fusion proteins employed in interaction assays (Fig. 63) contain TAF20-derived fragments listed on the left with their relative lengths and positions indicated by shaded boxes. Above, alpha-helices predicted by secondary structure analyses of the TAF20 amino acid sequence are indicated. On the right, the results of interaction studies with hTBP/hcore, hTAF80 and hTAF20 are summarized.



	hTBP	TAF80	TAF20
TAF20	+	+	+
N30	–	–	–
TAF15	+	+	+
C109	+	+	+/–
ΔC19	++	++	+
C89	+	+	+/–
C109-19	+	+	+/–
ΔC36	+/–	–	–
C65	+/–	+	+/–
C89-19	++	++	+/–
C109-36	–	–	–
ΔC51	+/–	–	–
C53	+	+	+
C65-19	+	+/–	+
C89-36	++	+	+
C109-51	–	–	–
ΔC65	–	–	–
C36	+/–	+	–
C53-19	–	+	–
C65-36	–	+/–	–
C89-51	+	+	–
C109-65	–	–	–
ΔC89	–	–	–
C22	–	+/–	–
C53-36	–	–	–
C89-65	–	–	–
ΔC109	–	–	–

Figure 65. Human TAF20-TBP interaction is not species-specific.

Interaction ('GST pull-down') study with indicated TAF20-derived fragments fused to GST and human His-TBP core domain (lanes 1-3), *Saccharomyces cerevisiae* His-TBP (lanes 4-6), or *Schizosaccharomyces pombe* His-TBP core domain (lanes 7-9). All Western blots were probed with rabbit polyclonal antiserum raised against His-hTBP core domain, but required different exposure times.

h core
Sc TBP
Sp core
Input
ΔC109
C89-36
Input
ΔC109
C89-36
Input
ΔC109
C89-36
1
2
3
4
5
6
7
8
9

indicator for the specific interaction under investigation. Two very small fragments were each found to be sufficient for a relatively efficient interaction with the conserved TBP core domain: C36 and C89-51 (summarized in Fig. 64). However, portions of the protein containing the most N-terminal predicted helix (helix A), but lacking the most C-terminal (Δ36 or C109-36) were negative in this interaction, possibly due to occlusion of the interaction surface in C89-51 by helix A. The fact that larger pieces derived from other regions of the protein are negative in such an assay is generally taken as further indication for (Arents et al., 1991) the specificity of the interaction.

Previous *in vivo* studies had shown that certain molecular interactions of yTBP essential for yeast viability are not mediated by human TBP (Cormack et al., 1991; Gill and Tjian, 1991). This so-called 'species-specific' interaction was mapped to the C-terminal 70 residues of the conserved TBP core domain (Cormack et al., 1991). In this context it was of interest to show that TBPs from a variety of species including *Saccharomyces cerevisae* and *Schizosaccharomyces pombe* were capable in interacting specifically with human TAF20 (Fig. 65). While it is not known whether human TAF20 is functional in yeast or whether its yeast homologue is capable of interacting with human TBP, we may conclude that the TBP interaction with human TAF20, which is mediated by the core domain, is not 'species-specific'.

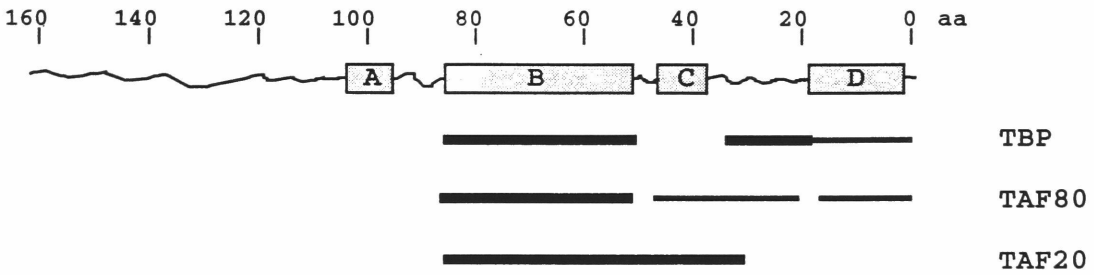
Two additional protein-interactions were probed with the available mutants. Given an interaction in the nucleosome between H2B and H4 (Arents et al., 1991), and the recent identification of a tentative homology relationship between H4 and dTAF60 (Kokubo et al., 1994) and its human homologue hTAF80, it was not unreasonable to expect a direct interaction between hTAF80 and hTAF20. As shown in Figure 63C, such an interaction can be demonstrated lending support to the proposed homology relationships between TAF20/15, TAF80 and H2B, H4 respectively. Analogy between H2B

Figure 66. Proposed model of TAF20 tertiary structure and protein interaction surfaces.

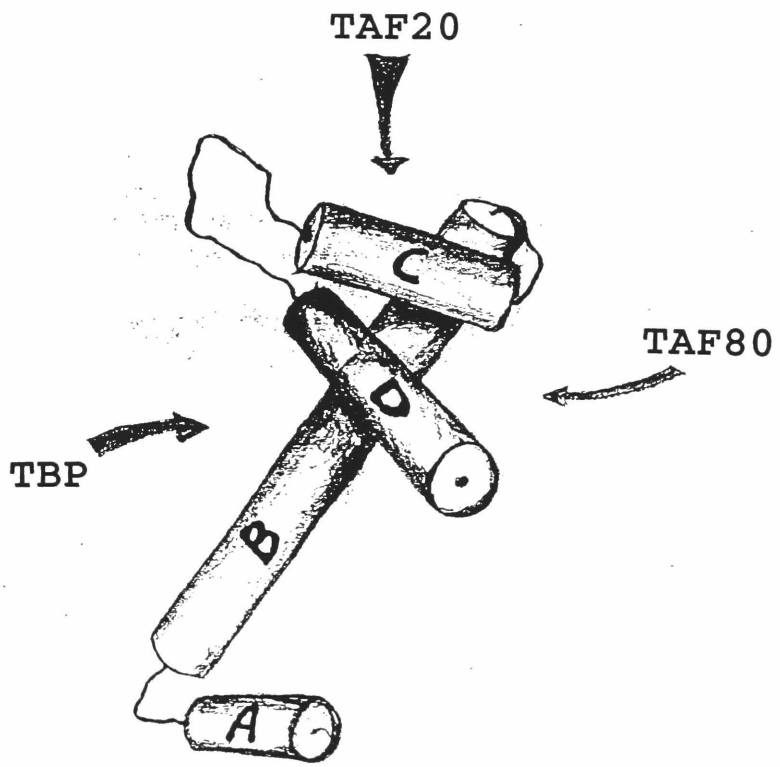
A. Summary of interaction studies indicating the portions of TAF20 protein involved in contacting TBP or TAF80, or required for dimerization.

B. Model of the proposed three-dimensional structure of human TAF20 based on X-ray crystallography studies of histone H2B within the histone octamer (Arents et al., 1991). Proposed interaction surfaces based on 'GST pull-down' assays with fragments of TAF20 are indicated.

A


160
140
120
100
80
60
40
20
0 aa
A
B
C
D
TBP
TAF80
TAF20

B


TAF20
TAF80
TBP
A
B
C
D

interacting stably with H2A and TAF20/15's presence in two forms in the cell, prompted me to investigate a possible self-interaction of this TAF (Fig. 63D). While stable dimerization of TAF20 has subsequently been demonstrated by glycerol gradient sedimentation analysis of de- and renatured bacterially expressed TAF20 protein (data not shown), this interaction, as well as the one with TAF80 requires a large portion of the intact TAF20 polypeptide implicating extensive, though clearly distinguishable, interaction surfaces for the two interaction partners.

Discussion

Important clues for the characterization of TAF20 may be drawn from the proposed sequence similarity to histone H2B (Fig. 61). The sequence similarity recently reported for portions of *Drosophila* dTAF40 and dTAF60 proteins with histone H3 and H4 respectively (Kokubo et al., 1994) may be seen to add some weight to the proposal that these proteins are evolutionarily related. Furthermore, interaction studies presented here have demonstrated a tight interaction between TAF20 and TAF80, as well as homo-dimerization of TAF20/15, that appear to be analogous to characterized interactions between H2B-H4 and H2A-H2B respectively (Arents et al., 1991). In addition, related experiments have demonstrated an interaction of hTAF20 with H2A and H4 (data not shown). These data suggest that a certain subset of TFIID subunits exhibit structural homology to nucleosomal components, and may form protein complexes with each other in a way that is analogous to interactions found in the nucleosome core. In support of such a hypothesis complementary lines of work have demonstrated that isolated histone similarity domains from dTAF40 and dTAF80 form a highly stable heterodimer (Y. Nakatani, personal communication) and that they interact, as expected, with the TAF20/15 homodimer (data not shown).

Interaction studies presented here identify the regions in TAF20 involved in stable interactions with TBP, TAF80, or itself (schematized in Fig. 66A). Given a structural homology between the C-terminal portions of TAF20 and histone H2B, I may attempt to model mapped interaction regions onto the structure of H2B as solved to 3.1Å resolution by X-ray crystallography (Arents et al., 1991). Amazingly, regions apparently disconnected in a linear map of TAF20 make up a contiguous interaction surface, such that all proteins so far identified as interacting with TAF20 can be assigned to distinguishable interaction surfaces (Fig. 66B). Further, more detailed studies, involving point mutants in each proposed interaction surface, are required to confirm the hypotheses derived from the preliminary molecular analysis presented here.

It might be of interest to note that an alternative secondary structure proposal for the Drosophila homologue dTAF28/22 (Yokomori et al., 1993b) is incompatible with some of the data presented here. The authors of that study also argue for a largely alpha-helical protein, but predict a flexible loop region at around C65. While results from my interaction studies with TAF20 mutants are entirely consistent with the proposed three-dimensional structure homology to H2B, I also note that the only deletion mutant of TAF20 whose expression in bacteria indicates an extreme protease susceptibility is C65-19 (Fig. 63A, lane 17). Such an observation is generally interpreted to signify misfolding due to disrupted secondary structure elements.

What might structural homology of hTAF20 to H2B mean with regard to its function within the TFIID complex ? It is tempting to speculate that the hTAF20/15 homodimer is a component of a 'nucleosome-like' complex together with hTAF80 and as yet unidentified human homologue of dTAF40. While there are precedents for histone variants active in DNA packaging (Vanfleteren et al., 1986; Brandt et al., 1988; de Andrade Rodrigues et al., 1988) due to sheer size differences

between TAF and histone proteins, these seemingly analogous complexes would differ significantly in their architecture and presumably in function. Although it is of immediate interest to detect DNA binding capabilities in TAF20 and the other TAFs, it seems highly unlikely from a structural point of view as well as with reference to characterized DNA binding properties of TFIID that the 'nucleosome-like' complex has DNA wrapped around itself. It is intriguing to note, however, that the extended TFIID footprint (Nakajima et al., 1988) is reminiscent (with its 10bp repeating pattern of hypersensitive sites) of DNase I footprints generated after nucleosome assembly. Such a repeating pattern is indicative of proteins protecting only one surface of the DNA helix over an extended region, and interestingly can be generated also by single histone proteins interacting with DNA (Kerrigan and Kadonaga, 1992). Thus it is conceivable that TAF20, along with other histone-related TAFs, contributes to the DNA binding properties of TFIID observed by DNase I footprinting. It may be noteworthy that an early UV cross-linking analysis with native *Drosophila* TFIID preparations identified proteins of 42 and 26kd that bind TATA-dependently to the hsp70 promoter (Gilmour et al., 1990), sizes that match exactly the apparent mobilities of *Drosophila* histone-related TAFs Functionally, such DNA interactions might be particularly important in template commitment, in the context of a more physiological chromatin template, and/or in transcription initiation from a TATA-less promoter.

Histone-related TAFs are clearly not just involved in nucleosome-like interactions. A recent report on the functional characterization of dTAF40 has proposed that this TAF is involved in mediating the transcriptional activation effect of acidic activators by stabilizing the interaction between TFIIB and, for example, VP16 (Goodrich et al., 1993). Furthermore, results presented here and published elsewhere (Yokomori et al., 1993b) have indicated interactions between hTAF20 and hTAF120/135 (or their *Drosophila* homologues

dTAF28/22 and dTAF110), which itself has been proposed to play a co-activator role in the activation process by glutamine-rich activators such as SP1 (Hoey et al., 1993). Such results might seem surprising at first in light of the histone-like structure of one domain of the respective TAF. However, classic studies implicating TFIID in the mechanism of activation, noted that TFIID's footprinting pattern on weaker promoters changes dramatically with activator binding to its cognate site (Horikoshi et al., 1988a; Horikoshi et al., 1988b). Thus the functional characterization of TAFs exhibiting histone homologies may shed light upon the apparently intrinsic link between mediation of activation by TFIID and its DNA interactions, and thus contribute to our understanding of the molecular basis for transcriptional activation.

Summary, Conclusion and Perspective

Introduction

The work presented in this thesis describes a beginning molecular characterization of the general transcription factor IID. In 1988, important aspects of its activity had been functionally described which attributed to this factor a central role in the regulation of transcriptional initiation. As the only DNA-bound general transcription factor it was shown to nucleate pre-initiation complex formation. Further, activator-dependent changes in its DNA interaction properties had implicated it as a direct target for activation surfaces and thus as a mediator for *trans*-activation. The beginning molecular characterization of this transcription factor has not only provided some insight into the molecular basis for these described activities, but more importantly has yielded some clues that will inform further functional analyses, for example, on the mechanism of initiation of transcription from the great variety of class II promoters, on transcriptional initiation and regulation within the context of chromatin templates, and on the mechanisms of activation by upstream regulatory factors. The availability of purified preparations of TFIID, or its subunit TBP, has allowed the characterization of other transcription factors (e.g. TFIIA, or *trans*-activators), and has even led to the identification of novel general factors (e.g. IIG, IIH, IIJ) and of a class of soluble co-activator proteins (USA) that have crucial functions in initiation and the activation of transcription, respectively.

Molecular cloning of TFIID subunits, allowing for their molecular and functional characterization, has shed some light on the functioning of the whole TFIID complex and thus on the mechanisms of transcriptional initiation by RNA polymerase II and its regulation. Just as the molecular characterization of TBP, for example, has revealed unexpected parallels between class I, class II and class III

transcription, so do we expect particular cues on the function of TAFs from sequence analysis and simple interaction studies. These may ultimately inform our understanding of control and regulation of transcription as a whole.

Contrary to widely-held notions regarding yeast TFIID, the recent demonstration that yeast also possesses a number of TAFs complexed with TBP will allow for a concerted effort in tackling the questions of transcriptional control in three different model organisms. The power of convenient reverse genetics available in yeast has already been important in confirming the importance of TAFs in transcriptional regulation: four TAFs (yTAF150, yTAF130, yTAF90, yTAF60) have been shown to be essential for cell viability (D. Poon and P.A. Weil, unpublished), while overexpression of dTAF60 and dTAF40 in yeast cause a very sick phenotype (Y. Nakatani, personal communication).

In this final chapter I have attempted to summarize all currently available data on the molecular characterization of TFIID and its subunits. This summary will serve as the basis for my more general, highly speculative discussions in the final sections of this chapter about initiation pathways, TFIID functions within it and the regulation of transcription.

Figure 67. Comparison of TFIID subunits in yeast, *Drosophila*, and human.

TFIID subunits identified in yeast, *Drosophila*, and human are listed according to their apparent molecular weight. Cloned TAFs (to date) are boxed and homology relationships are indicated by stippled lines. Distinguishing structural or functional characteristics are noted on the right of the table.

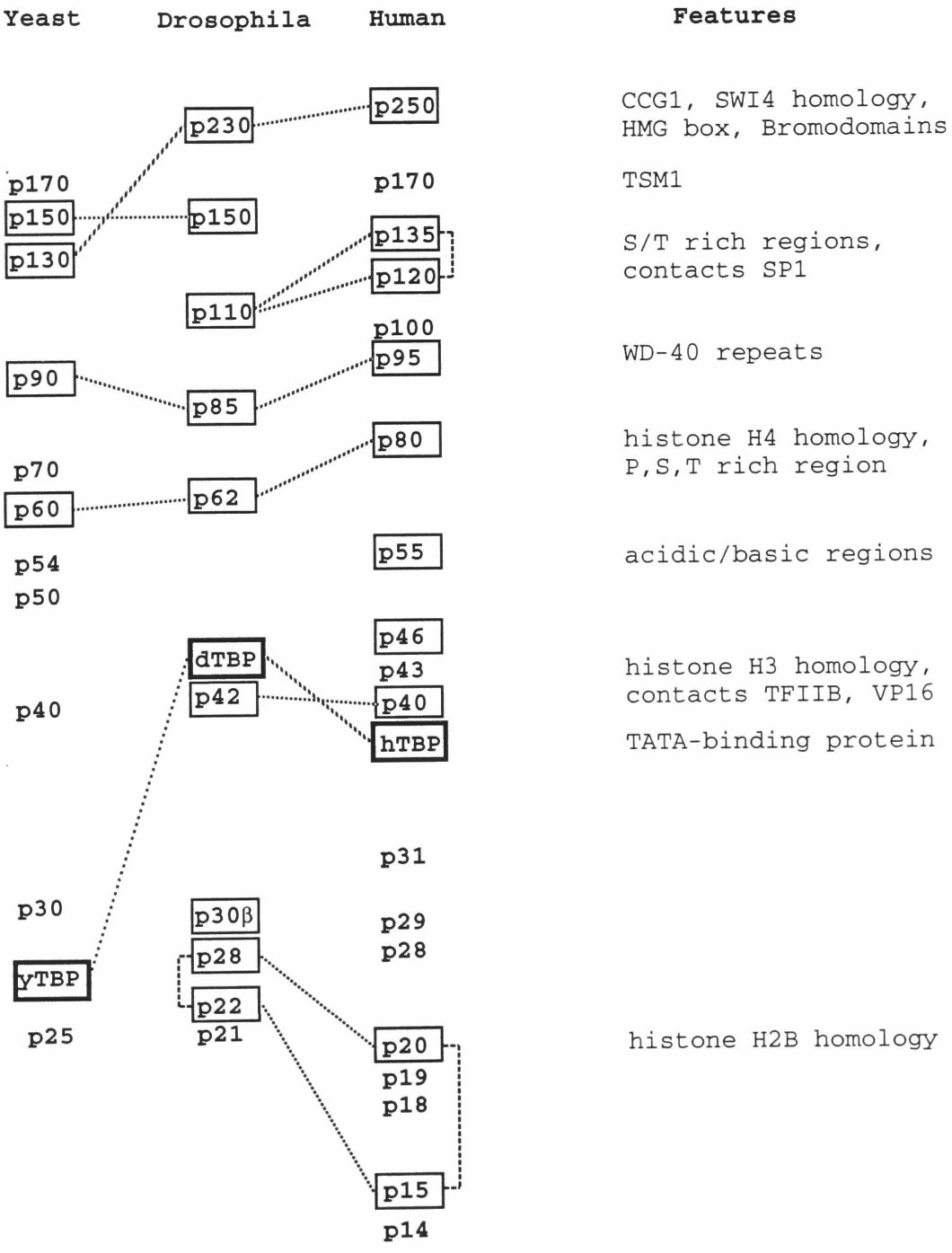

Yeast
Drosophila
Human
Features
p250
p230
CCG1, SWI4 homology,
HMG box, Bromodomains
p170
p170
p150
p150
TSM1
p135
p130
p120
S/T rich regions,
contacts SP1
p110
p100
p95
p90
WD-40 repeats
p85
p80
histone H4 homology,
P,S,T rich region
p70
p60
p62
p55
acidic/basic regions
p54
p50
p46
dTBP
p43
histone H3 homology,
contacts TFIIB, VP16
p42
p40
p40
hTBP
TATA-binding protein
p31
p30
p30β
p29
p28
p28
yTBP
p22
p20
histone H2B homology
p25
p21
p19
p18
p15
p14

Figure 68. Human TAFs: protein interactions
Positive results from a variety of protein-protein interaction studies are summarized in the table. For details and references see the text.

250		+	?	+	+		+	+
135/120	+		+					+
95	?	+					+	
80	+					+	+	+
55	+							
40				+				
TBP	+		+	+				+
20/15	+	+		+			+	+
$hTAF_{II}$	250	135/120	95	80	55	40	TBP	20/15
DNA	?		?	?	?	?	+	?
TFIIB						+		
acidic						+		
Q-rich		+			+			
P-rich					+			

(i) The TFIID complex

Characterizing TFIID subunits

As TFIID has been studied in a variety of experimental systems I have attempted to summarize its components identified so far in the three most intensively studied of these. Figure 67 is intended as a reference guide, emphasizing homology relationships between TFIID subunits from *S. cerevisae*, *Drosophila*, and human to avoid confusion of the TAF nomenclature. Interaction studies with recombinant proteins, as described below and summarized in Figure 68, are the basis for much of our current ideas concerning the architecture of the TFIID complex. In the following I will attempt to give an overview of the current state of the physical and functional characterization of each subunit of the TFIID complex studied to date. As this is a research topic currently under intensive scrutiny, the following can necessarily only function as a temporary progress report, soon to be superseded by future reviews of the subject.

TBP As the first TFIID subunit to be identified, the TATA binding protein is currently the best characterized of all general transcription factors. As described in chapter III, crystal structures of At2 TBP (Nikolov et al., 1992) and *S. cerevisiae* TBP (Chasman et al., 1993; Geiger et al., 1994) as well as recent refinements (Nikolov and Burley, 1994) has given us a precise molecular description of this symmetrical, saddle-shaped molecule. Co-crystal structures of At2 TBP with a major late TATA element (Kim et al., 1993a; J.S. Kim and S.K. Burley, in press) and *S. cerevisiae* TBP with the TATA box from the yeast *CYC1* promoter (Kim et al., 1993) have described in detail the remarkable DNA interaction by TBP via the minor groove and the gross distortion of the DNA helix, though these have failed to explain TBP's sequence specificity. Specific TATA box binding is dramatically

increased by TFIIA whose interaction has been mapped to basic residues in the long amphipathic helix of the first repeat (Buratowski and Zhow, 1992; Lee et al., 1992). TBP binding to DNA by itself has long been known to be resistant to nucleosome competition (Meisterernst et al., 1990) and to be sufficient for nucleating the assembly of the pre-initiation complex via direct interactions with TFIIB (Buratowski et al., 1989; Kim et al., 1994) and possibly TFIIE (Maxam and Tjian, 1994).

Direct interactions of TBP with a vast number of proteins, including viral and cellular activators and co-activators have been reported (see chapter III(iii) for a summary), and a few of these interactions have been mapped to particular residues on the surface of TBP (Eisenmann et al., 1992; Hagemeier et al., 1992; Inostroza et al., 1992). However, it is a currently unresolved question how this relatively small molecule is able to engage in such a large number of functional interactions at the same time. Further mutagenesis and physical studies might shed light on this question, while complementary functional *in vitro* or *in vivo* studies may be directed to elucidate the functional/ physiological significance of the reported interactions. Other issues to be addressed regard TBP's DNA interactions on TATA-less class II promoters - where competition with TATA-containing oligonucleotides were found to have no effect on transcriptional activity (Martinez et al., 1994) - as well as its function in general on class I and class III promoters as part of SL1 and TFIIIB$_2$ complexes. Finally, the TBP N-terminal structure, which is sequence conserved in higher eukaryotes, has not directly been assigned any function, though recent experiments with monoclonal antibodies seem to implicate it in transcriptional initiation from TATA-less promoters.

hTAF250/230/CCG1;dTAF230;yTAF130 The largest TBP-associated factor in higher eukaryotes (Hisatake et al., 1993; Kokubo et

al., 1993a; Ruppert et al., 1993; Weinzierl et al., 1993b) was identified upon peptide sequencing as a previously cloned protein essential for the progression of the cell cycle from G1 phase (Sekiguchi et al., 1991). Using the originally identified temperature mutant cell line (ts13) in transfection assays, it was observed that this mutant CCG1/TAF230 (unlike other cell cycle mutant cell lines) causes a transcriptional defect only on some promoters (Hirschhorn et al., 1984; Liu et al., 1985) that can be rescued with wild type TAF230 (Wang and Tjian, 1994). However, why this mutant causes a G1 arrest and what the molecular function is of TAF230 within the TFIID complex remains to be investigated.

Far Western analysis with TBP had previously identified this polypeptide as directly interacting with TBP (Fig. 35; Takada et al., 1992; Ruppert et al., 1993). Further interactions of dTAF230 with dTAF110 substantiated the view that this large TAF directs the assembly of the TFIID complex on TBP (Weinzierl et al., 1993b), but more recent data indicates that while this largest TAF engages in a number of interactions, other TAFs (e.g. dTAF110/hTAF135/120) are tethered to the TBP-containing complex by additional protein interactions, and that some immuno-purified TFIID preparations contain substoichiometric amounts of this TAF (Fig. 62).

Detailed analysis of the TBP-TAF230 interaction has demonstrated that recombinant TAF230 has an inhibitory effect on basal transcription when added to a reconstituted system containing TBP (Kokubo et al., 1993a). This functional inhibition is correlated with an inhibition of TBP binding to the TATA box, suggesting a modulatory function for TAF230 as a possible mechanism for *trans*-activation if it can be shown that this mechanism is targetted by an activator. Further mutational analyses have mapped this inhibitory function to an N-terminal 60 residue fragment of TAF230 (Kokubo et al., 1994) and implicated a phenylalanine within this peptide in

the direct interaction with TBP (Y. Nakatani, personal communication). The recent isolation of the yeast homologue yTAF130 (D. Poon and P.A. Weil, unpublished) should facilitate functional testing of models with appropriately constructed mutations in this protein.

dTAF150;yTAF150/TSM1 Little is known about this TAF whose human homologue has not yet been identified with certainty. Recent cloning of the yeast (D. Poon and P.A. Weil, unpublished) and *Drosophila* (C.P. Verrijzer, L.D. Attardi, K. Yokomori, J.L. Chen, R. Tjian, unpublished) protein has revealed its identity as previously described yeast essential gene TSM1 (Ray et al., 1991). Recent cross-linking studies with Drosophila TFIID on the downstream region of the hsp70 promoter (Sypes and Gilmour, 1994) have implicated this TAF as a prime candidate (along with dTAF230) for the observed downstream DNase I protection on this promoter.

dTAF110;hTAF135/120 Drosophila TAF110 was the first TAF characterized that seemed to have some of the functions postulated for TAFs as factors required for mediating the activator effect (Hoey et al., 1993). dTAF110 was shown to contain domains similar to Sp1's glutamine-rich activation domain that is capable of effecting activation when targeted to an upstream binding site via a fused DNA binding domain. Furthermore, the same domain was shown to interact with Sp1 *in vitro* and in the yeast dihybrid system (Fields and Song, 1989) suggesting that dTAF110 may function as a co-activator by interacting with and amplifying thereby the effect of the glutamine-rich activation domain of Sp1 in a phenomenon termed "superactivation" originally observed with multimerized Sp1 activation domains (Courey et al., 1989; Pascal and Tjian, 1991). Subsequently, a glutamine-rich hydrohobic patch within transcription factor Sp1 was identified to be responsible for the interaction with TAF110 as well as for Sp1's activation effect (Gill et al., 1994).

A similar correlation between activation function and observed interaction was observed with deletion mutants of the cAMP-regulated transcriptional activator protein CREB whose 120 amino acid glutamine-rich activation domain (Q2) can also be shown to interact with dTAF110 (Ferreri et al., 1994).

The human homologue for this TAF seems to be present in at least two isoforms generated by alternative splicing as TAF135/120 (S. Hasegawa, personal communication). However, it is currently unclear whether these proteins have similar activities as dTAF110; no yeast homologue has been reported yet.

hTAF95;dTAF85;yTAF90 Although this TAF has been cloned from all three organisms, remarkably little is known about it; sequence analysis has revealed the presence of the universally occurring b -transducin (WD40) repeats (Dalrymple et al., 1989) in this TAF. These presumed protein-protein interaction domains were previously implicated in the functioning of the general repressor Ssn6-Tup1 in yeast (Keleher et al., 1992) and may in this case be involved in either TAF/TAF or TAF/activator/repressor interactions (Dynlacht et al., 1993). Genetic studies in yeast will no doubt elucidate the role of these conserved repeats and the possible function of this TAF.

hTAF80/70;dTAF62/60;yTAF60 Cloning of this TAF from human (K. Hisatake, personal communication) and *Drosophila* (Weinzierl et al., 1993a; Kokubo et al., 1994) has led to the mapping of protein-protein interactions with dTAF42 and dTAF230 within the complex. More imortant, however, is the proposed homology of the most N-terminal 100 residues of this TAF to histone H4 and of dTAF42 to histone H3 (Kokubo et al., 1994). A recent mutagenesis study has mapped the dTAF42-dTAF62 interaction to the histone homology domains in each protein. Furthermore, the corresponding fragments from each

TAF form stable heterodimers analogous to the dimerization of H3 and H4 (Y. Nakatani, personal communication). Possible functional implications are discussed below and will no doubt be addressed experimentally in the near future.

hTAF55 This highly charged TAF was recently shown to interact with hTAF230 and a number of transcriptional activation domains, including CTF (C.-M. Chiang, personal communication). While homologues in yeast and *Drosophila* are yet to be identified, one exciting possibility is that this TAF is involved in the mediation of *trans*-activation by so-called proline-rich activators, few of which have been identified in non-mammalian species. The results from further investigations are eagerly awaited.

hTAF40/dTAF42/40 Aside from its proposed histone H3 homology (Kokubo et al., 1994), the *Drosophila* TAF has been implicated in the activation mechanism of the acidic activator VP16 (Goodrich et al., 1993) by demonstrating interactions of this TAF with the C-terminal half of the activation domain of VP16 as well as with general transcription factor TFIIB whose recruitment to the pre-initiation complex has been proposed to be stimulated this acidic activation domain. Furthermore, antibodies against dTAF40 seem to have an inhibitory effect on the activation effect by VP16, an observation that lends some support to the model that dTAF42 mediates VP16 activation by promoting the acidic activation domain-dependent recruitment of TFIIB.

dTAF30β While neither homologues human and yeast have been identified nor potential functions have been proposed for this TAF, primary interaction partners within the TFIID complex appear to be dTAF230 and dTAF110 (Yokomori et al., 1993b).

hTAF20/15;dTAF28/22;TAF30α This TAF is present in two isoforms in both *Drosophila* (Yokomori et al., 1993b; Kokubo et al., 1994) and human (chapter IV) and consists in its conserved C-terminal half of a histone H2B-like domain that mediates dimerization, interactions with H4-like hTAF80 (chapter IV), TBP (chapter IV; Yokomori et al., 1993b; Kokubo et al., 1994) and hTAF135/120/dTAF110 (chapter IV; Yokomori et al., 1993b; Kokubo et al. 1994). Recent studies seem to indicate that the stable hTAF20/15 dimer interacts efficiently with the stable dTAF42-dTAF62 heterodimer suggesting the intriguing possibility that a nucleosome-like complex is part of TFIID and may be involved in mediating activator response (see next section).

Promoter interactions by TFIID

The intense interest in promoter DNA interactions by TFIID are primarily two-fold : First, eukaryotic proximal promoter regions exhibit a great variety of dicernible sequence elements or just plain sequence divergence within the vicinity of the initiation region. Based on our knowledge of prokaryotic promoters it may be reasonable to expect that such sequences provide the code for promoter strength, activity or inducibility in response to a particular activator. Although the TATA binding protein does not bind DNA with great sequence specificity it has become apparent that on true TATA-less promoters, TBP's DNA binding activity is not required at all (Martinez et al., 1994). Instead, the initiator (INR), a weak sequence consensus around the transcription initiation site, plays an important role in determining initiation level and start site. Whether, this initiator consensus is recognized by a component of the TFIID complex or a chromatographically separable factor (e.g. TFII-I, YY1) is currently the focus of intense research. While binding of such a factor plays an important role in promoter activity, start site selection probably involves other

factors, as systematic swapping of basal factors between *S. cerevisae* and *S. pombe* has demonstrated (Li et al., 1994). It should also be born in mind that there is some evidence that RNA polymerase II itself has sequence preference for the INR (Carcamo et al., 1991).

Second, DNase I footprinting patterns of TFIID on a variety of promoters vary from 'restricted' protection around the TATA element to a much broader footprint indicating interactions with DNA as far as +40. Furthermore, footprinting by TFIID on certain promoters was shown to be affected by presence of activators bound at upstream cognate sites. Thus the DNase I protection assay can be used to probe the conformation of TFIID on the promoter; the conformation of the TFIID-DNA complex furthermore appears to be correlated with the activity of the general factor in catalyzing the steps required for transcriptional initiation. Subunits of TFIID responsible for the observed extension of the footprint may therefore be directly or indirectly involved in the mechanism of activation.

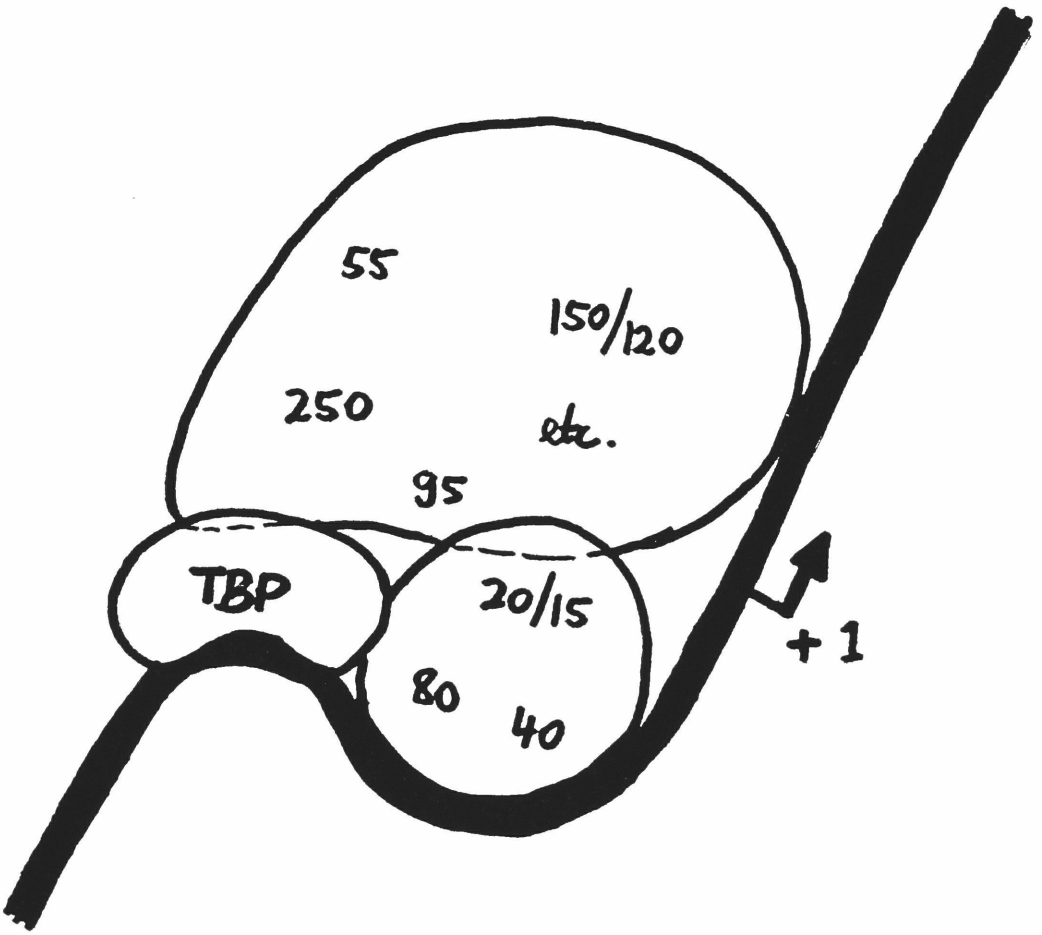
Following the demonstration of a large footprint by TFIID fractions from *Drosophila* (Parker and Topol, 1984b) and human extract (Sawadogo and Roeder, 1985b; Nakajima et al., 1988) that covers the entire TATA box and initiation region, there has been an active interest in identifying the protein(s) associated with this activity. With the identification of an intiator (INR) consensus sequence (Smale and Baltimore, 1989) and proteins that seem to bind to and function through the INR (Roy et al., 1991; Seto et al., 1991; Du et al., 1993) to promote initiation complex assembly, an exciting debate has developed about the proteins that bind in the initiation region of functional, basal or activated promoters.

First, downstream sequences have been shown to play a role in TFIID binding to at least some promoters (Nakatani et al., 1990; Emanuel and Glimour, 1993; Purnell and Gilmour, 1993) although competition with short fragments containing

INR sequences have not indicated specificity. DNA binding specificities of identified INR binding proteins on the other hand, match or at least overlap with the established sequence consensus (Smale and Baltimore, 1989; Roy et al., 1991; Seto et al., 1991). However, careful footprinting analysis with purified TFIID also demonstrates that TFIID-DNA interactions around the intiator region depend on the precise initiator sequence, which, importantly, is correlated with functional activity. Furthermore, an elegant PCR-aided selection scheme utilizing a gelshift assay with *Drosophila* TFIID resulted recently in the identification of an initiator binding specificity for TFIID that match a number of independently established initiator consensus sequences (Purnell et al., 1994).

Second, attempts to crosslink proteins in nuclear extract to the downstream region have identified *Drosophila* polypeptides of 250kd, 150kd, 60kd, 26kd by UV irradiation whose close DNA contacts are sensitive to salt and competition with a TATA-containing oligo, as is the downstream protection by TFIID seen in DNase I and Exonuclease III footprinting assays (Gilmour et al., 1990; Sypes and Gilmour, 1994). Similar crosslinking to the TATA box has yielded a strong 42kd band whose SDS-PAGE mobility is indistinguishable from *Drosophila* TBP. Furthermore, many components of pure human TFIID can be cross-linked to precise regions of a variety of promoters with a long range cross-linking agent such that we can derive a description of the architecture of the TFIID-DNA complex (T. Oelgeschläger, personal communication). The recent realization that four of the TFIID subunits have significant homology to histones (chapter IV) and are capable of forming a stable tetrameric complex (Y. Nakatani, personal communication) may indicate that such a nucleosome-like complex may be responsible for observed downstream DNA interactions. As nucleosome components have only limited sequence or DNA structure specificity and bend the DNA in the opposite direction from

Figure 69. Model of TFIID on a class II promoter.
DNA in the vicinity of a class II initiation site is proposed to be bent in one way by TBP and by a nucleosome-like complex in the other way, which consists of a heterodimer of TAF80, TAF40 and a dimer of TAF20/15. Other TAFs are present in the TFIID complex through multiple interactions with TBP and this nucleosome-like complex, and might engage in activation co-factor functions.


55
150/120
250
etc.
95
TBP
20/15
+1
80
40

TBP at the TATA box, such a model would appear plausible and will undoubtedly be tested in the near future.

Several lines of circumstantial experimental evidence indicate that TFIID-DNA interactions are functionally relevant and furthermore that there may be multiple binding modes by TFIID on the promoter. As an important pre-amble to this discussion it should be noted that there are evidently a number of factors that modulate TFIID-DNA interactions. First of all, TFIIA, initially identified as a general transcription factor and then implicated in activation phenomena (Meisterernst and Roeder, 1991), seems to act by enhancing TFIID's interactions with the TATA element. Interestingly, the activation domain of VP16 also qualitatively alters TBP binding to the TATA box, an effect that is reduced when trans-activation-defective mutants of VP16 are used, and can be suppressed by addition of TFIIA (Liljelund et al., 1993). The N-terminus of TAF230/250 on the other hand, has an inhibitory effect that is conceivably modulated by other regulatory proteins (Kokubo et al., 1994). In addition, a biochemically characterized yeast protein, ADI, binds to TBP and is then capable of removing it from the DNA in an ATP-dependent manner (Auble and Hahn, 1993). Importantly, this protein was recently found to be the same as Mot1, a genetically identified global repressor of class II transcription (Auble et al., 1994), thus demonstrating the link between TFIID-DNA interactions and transcriptional regulation. Finally, PC3, a component of the HeLa-derived USA fraction that was identified as Topo I, has also an inhibitory effect on TATA box recognition but, interestingly, functions to increase the level of activator-dependent transcription (Kretzschmar et al., 1993; Merino et al., 1993).

Detailed studies currently undertaken on the human hsp70 promoter demonstrate that significant downstream cross-linking is dependent on the presence of the upstream activation surface CTF (T. Oelgeschläger, personal

communication). These are consistent with the earlier footprinting studies on the E4 promoter with a partially purified TFIID fraction and the activator ATF (Horikoshi et al., 1988b), and, importantly, support the interpretation that TFIID subunits undergo functionally relevant conformational changes that can be observed with respect to their DNA interactions. The fact that these conformational changes on the hsp70 promoter are triggered only by the CTF interaction surface (T. Oelgeschläger), and not VP16 for example, which does not function on the wild-type promoter, may lead to an intriguing idea regarding the activation potential of initiation regions. While either ATF or gal4-AH were shown to ellicit TFIID downstream interactions on the E4 promoter, corresponding sequences of the human hsp70 promoter may in some way lead to distinguishable responses by TFIID (regarding its DNA interactions) to CTF (proline-rich) or VP16 (acidic) activation domains. Such results seem to relate to observations made much earlier that the precise sequence of the TATA element and surrouding region not only affect the in vitro basal activity of a promoter, but also influence its inducibility by E1a or upstream bound activators.

In summarizing the available data it is becoming increasingly clear that the coding potential of the sequences and/or the DNA structure surrounding transcription initiation site, as well as, the multiple DNA binding modes of the TFIID complex resulting from quaternary conformational changes of its subunits need to be understood much more detail in order to study the mechanisms of complex assembly, initiation and activation in molecular terms.

Function in mediating *trans*-activation

The existence of TBP-associated factors was originally hypothesized as an *in vitro* reconstituted transcription system supplemented with recombinant TBP was found not to be

responsive to upstream activators as when supplemented with a partially purified native TFIID preparation (chapter I(iii); (Hoey et al., 1990; Hoffmann et al., 1990; Peterson et al., 1990)). *Qua definitionem*, TAFs were thereby linked to mediating *trans*-activation, although this need not be their only functional role and further, not all of the many class II TAFs identified to date necessarily participitate in this process. (TAFs are, for example, required for basal transcription from TATA-less promoters; (Martinez et al., 1994)) Initial functional characterization of purified TFIID (Tanese et al., 1991; Zhou et al., 1992) led to the assumption that TAFs are the only factors required in *trans*-activation, but it is now apparent that components of the soluble co-activator fraction USA, identified and purified from the phosphocellulose D fraction (Meisterernst et al., 1991) but also present as a contaminant in basal factor preparations (TFIIE/F/H), are required to observe high levels of activated transcription *in vitro* (Chiang et al., 1993). Thus TAFs are required but not sufficient for an activator-responsive *in vitro* transcription system.

When studying the molecular basis for activated transcription, it may be useful to conceptualize and distinguish different mechanisms for how experimentally observed activation of transcription may be achieved. A first mechanism is "anti-repression", a process in which inhibitory factors (e.g. histone H1) are overcome (e.g. by competition) in direct response to activator binding or via positive co-factors. A second mechanism underlying increased levels of transcription can be described as the enhanced recruitment of TFIID to the promoter by upstream bound activators and is therefore termed "elevated basal" transcription. Both of the above mechanisms may or may not be true activation phenomena: an operative definition for true activation may be that the observed increase in transcription levels is not dependent on limiting amounts of a basal factor. In certain conditions activator-dependent

increases in transcription can thus be observed *in vivo* and *in vitro* without the participation of TAFs or co-factors (Meisterernst et al., 1990; Colgan and Manley, 1992). By this definition, true activation involves a qualitative change in the initiation complex that is activator-dependent. Interestingly, activation-deficient mutants of yeast TBP may fail to recruit TFIIB efficiently in response to activator interactions (Kim et al., 1994). Similarly, we may speculate that conformational changes in TFIID triggered by upstream bound activators lead to an enhanced recruitment or increased initiation rate of a holoenzyme complex.

Co-factors for activated transcription can thus be imagined to play several roles in the initiation process. A 'co-activator', for example, functions by enhancing the effect of the activation surface by co-operative interactions with the activator. Proteins that seem to function in this way have been identified in yeast (e.g. gal11, (Himmelfarb et al., 1990; Nishizawa et al., 1990)). Other activation co-factors might function as originally envisioned 'bridging factors' or 'mediators' that provide the link or stabilize interactions between activation surface and basal machinery. One yeast protein thought to function in this way is ada2, isolated by its ability to suppress the squelching effect of overexpressed VP16 activation domain in yeast (Berger et al., 1990; Berger et al., 1992), although its precise molecular mechanism remains to be analyzed *in vitro*.

A third, evidently large class of activation co-factors point towards a link between chromatin packaging and transcriptional regulation. While chromatin should clearly be taken into account when considering general transcriptional mechanisms (Felsenfeld, 1992), in yeast, nucleosome components have also been shown to have specific effects on transcriptional regulation on particular genes (Durrin et al., 1991; Mann and Grunstein, 1992), while *in vitro* studies have identified other chromatin components as activation co-factors (e.g. NC1/DR2/HMG1 (Ge and Roeder,

1994); Topo I (Kretzschmar et al., 1993; Merino et al., 1993)). Genetic screens in yeast have led to the isolation switch and SNF genes that are required for the expression of a broad spectrum of diversely regulated genes (Laurent and Carlson, 1992; Yoshinaga et al., 1992), and have been shown to form at least one biochemically described protein complex (Cairns et al., 1994) which is thought to play a role in the rearrangment of chromatin components as a regulatory mechanism for gene expression (Winston and Carlson, 1992). Biochemical analyses with human and *Drosophila* nuclear extracts have not only identified a homologous complex (brahma (Tamkun et al., 1992), BRG1 (Khavari et al., 1993), and human SWI2 complex (Muchardt and Yaniv, 1993; T. Ohta, personal communication)) but also possibly distinct nucleosome disruption factors (Kretzschmar et al., 1994; Tsukiyama et al., 1994).

What is beginning to emerge from the molecular characterization of TFIID subunits , as summarized in section (i), is that class II TAFs may function in all three ways to mediate activation effects. An analysis of dTAF110 (Hoey et al., 1993) has demonstrated that a glutamine-rich domain of this TAF can act act cooperatively with SP1 in a superactivation assay when targeted to the promoter via the gal4 DNA binding domain. Evidence of direct interactions between SP1 and dTAF110 may then suggest that this TAF acts as a 'co-activator' for at least some glutamine rich activation domains, such as those of Sp1 (Hoey et al., 1993) and CREB (Ferreri et al., 1994). Interaction studies with dTAF40 on the other hand, lead to the hypothesis that this TAF acts as a 'bridging factor' by stabilizing interactions between the activation surface of VP16 and TFIIB (Yokomori et al., 1993b) effecting enhanced recruitment of this general transcription factor in response to bound activator. While interactions between hTAF55 and a number of activation domains have been demonstrated (e.g. CTF), it is not yet

clear how this TAF might be involved in mediating *trans*-activation (C.-M. Chiang, personal communication).

Finally, structurally significant homologies between some TAFs and histones (as discussed in chapter IV; (Kokubo et al., 1994)), furthermore, may point towards an involvement of these TAFs in DNA packaging. We can speculate that as specialized chromatin components these may effect changes in the DNA conformation that can be regulated by activators, activation co-factors, or activator-responsive DNA chaperones (Travers et al., 1994). It may be relevant to note that the severe distortions introduced into the DNA by TBP bound to the TATA box (Kim et al., 1993; Kim et al., 1993) are in the opposite direction to DNA wrapping around the nucleosome core particle or its predicted bending around histone-like TAFs. Such an arrangement might provide for multiple potential energy minima that are the basis for multiple yet discrete conformations of the TFIID-DNA complex. This sort of regulated chromatin structure may in turn determine the accessibility for or recruitment rate of latter bound general transcription factors and/or RNA (holo)-polymerase. More detailed studies on TFIID-DNA contacts and activator-effected conformational changes of the initiation complex (see above) may be used to test these hypotheses.

(ii) Perspectives on transcription initiation

Class II initiation pathways

Biochemical analysis of the components required to observe accurate initiation of transcription in a cell-free *in vitro* system has resulted in the identification and purification of seven chromatographically separable general transcription factors. Earliest attempts to characterize the function of partially purified factors were most successful for TFIID, which was shown to be important in promoter recognition and template commitment (Fire et al., 1984; Hawley and Roeder, 1985). Other factors, although chromatographically separable, were shown to interact with one another in solution forming distinct complexes (namely IIB-IIE, IIE-polII, IIB-IIE-polII (Reinberg and Roeder, 1987);IIB-polII (Tschochner et al., 1992); IIF-polII (Sopta et al., 1985) in glycerol gradient centrifugation or RNApolII affinity chromatography. Physical analyses of the pathway leading to transcription initiation thus found three distinguishable steps (van Dyke et al., 1988): first template commitment by TFIID (aided by TFIIA), then preinitiation complex formation by the assembly of TFIIB, TFIIE/F and polII/F, followed by an energy dependent step leading to initiation.

Further purification of these factors and the advent of the electrophoretic gel mobility shift assay, however, allowed a more detailed analaysis of the assembly steps (Buratowski et al., 1989), thereby shifting the emphasis towards conceptualizing each general transcription factor as a physiologically soluble and distinct entity that must participate in a highly ordered multi-step assembly pathway to form a productive pre-initiation complex on the promoter.

Recently published immunoprecipitation experiments from yeast extract at physiological buffer conditions provide evidence for the existence of a large RNA polymerase II holoenzyme complex containing stoichiometric amounts of all general transcription factors except TFIID and yeast factor a

(TFIIE), as well as certain putative activation co-factors (SRB proteins), that is capable of initiating transcription accurately in an activator responsive fashion (Koleske and Young, 1994). Such a finding, if corroborated in *Drosophila* and human systems, should lead to an important revision of current pre-initiation complex assembly models in support of the much older biochemical observations with partially purified RNA polymerase and chromatin templates (Parker and Roeder, 1977) or glycerol gradient (as summarized above).

Such a model predicts a two step assembly of the initiation complex, consisting of promoter recognition by TFIID and template commitment, followed by recruitment of a holoenzyme complex (whose energy dependent initiation step may be regulated by TFIIE and TFIIH (Goodrich and Tjian, 1994; Maxam and Tjian, 1994; Ohkuma and Roeder, 1994)). A two step pathway is also favoured when considering transcriptional activation: in a multi-step pathway only one step may be limiting that can be effectively targetted by an activator; in a model which predicts the recruitment of a holoenzyme complex activators can act on multiple molecular targets to affect the rate of transcription initiation. For example, *in vitro* (Roberts et al., 1993) and *in vivo* (Colgan et al., 1993) observed activator-effected recruitment of TFIIB, which has no functional effect *in vitro* when separated general transcription factors are used (Choy and Green, 1993), may prove its functional relevance when considering this general transcription factor part of an RNA polymerase II holoenzyme complex.

Universal characteristics of initiation pathways

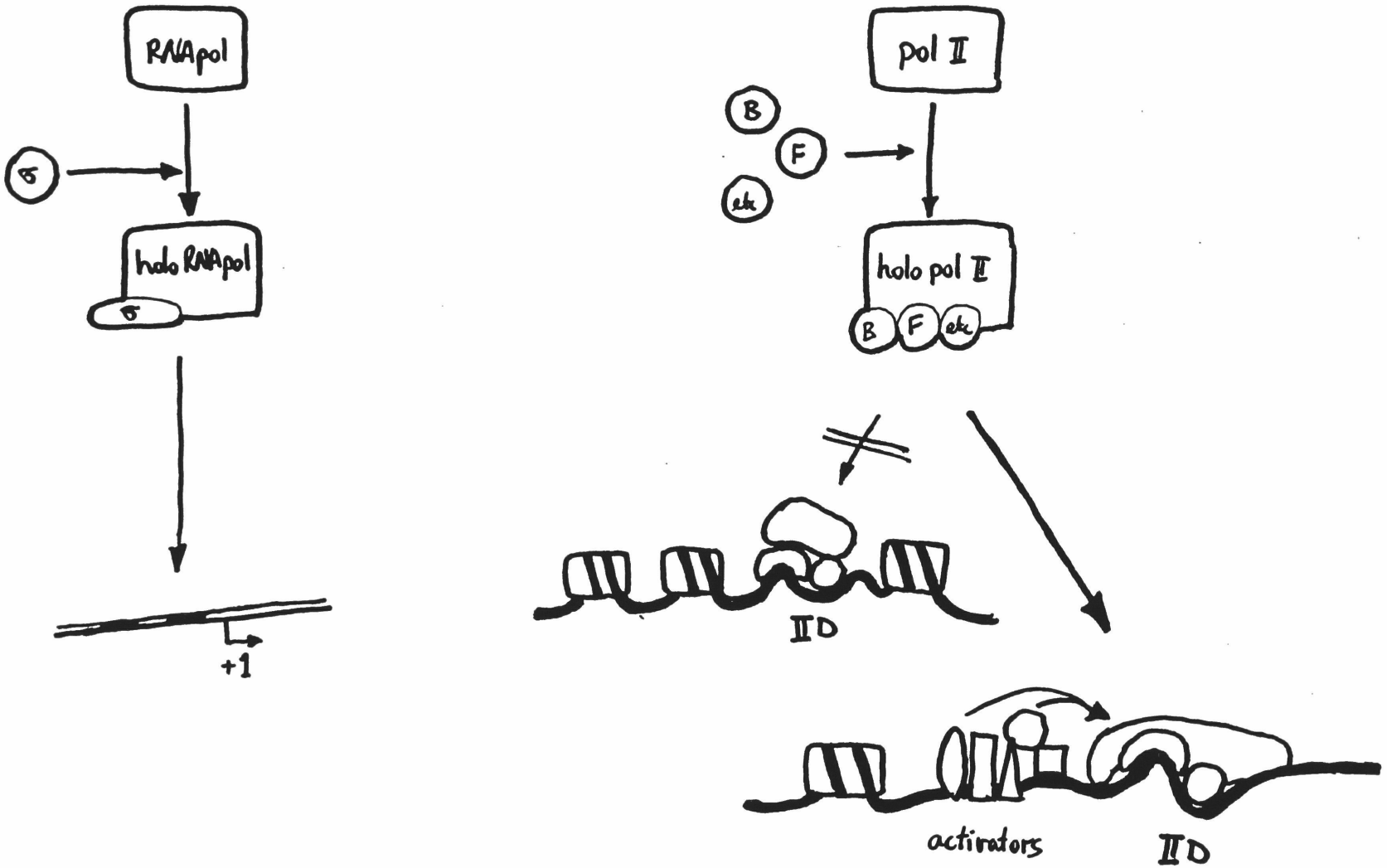
The biochemical and genetic analysis of transcription initiation and regulation in prokaryotes, primarily *Escherichia coli*, is often evoked as the ultimate paradigm for transcriptional studies in eukaryotes (Ptashne, 1986; Ptashne, 1988). Given TFIID's pre-eminent role as a promoter

recognition factor via a T/A-rich sequence element, it has understandably been viewed functionally analogous to prokaryotic sigma factors and TBP was proposed to contain a weak sequence homology to the -10 element-binding 2.4 region of sigma factors (Horikoshi et al., 1989a). However, neither functional analogy nor sequence homology between this eukaryotic and prokaryotic initiation factor has been confirmed by more detailed molecular analyses of TFIID subunits such as X-ray crystallography of TBP (chapter III(ii)).

The fundamental difference between these two proteins has turned out to be that TFIID is chromatin-associated and can thus act as a template commitment factor, while sigma associates with the RNA polymerase to form the holoenzyme that is then capable of promoter recognition. In considering these functional characteristics, it seems more appropriate to view those general transcription factors that are bound to RNA polymerase as functional equivalents to sigma. Related to this, sequence homologies between TFIIB, RAP30 (TFIIF-b) and the two subunits of TFIIE to sigma regions involved in RNA polymerase binding (2.2) and promoter melting (2.1 and 2.3) have been proposed (Malik et al., 1991). Employing recombinant proteins, it was indeed shown that TBP, TFIIB and RAP30 are sufficient for promoter recognition by highly purified RNA polymerase II (Killeen et al., 1992), and that RAP30's sigma homology domain is capable of interacting with eukaryotic as well as prokaryotic RNA polymerase (McCracken and Greenblatt, 1991).

Thus the model proposed above for a physiological pathway of initiation complex formation allows for a surprisingly high degree of analogy between eukaryotic class II and prokaryotic initiation mechanisms. In both cases RNA polymerase requires the transient association of (an) initiation factor(s) for promoter recognition, but which are dissociated from the elongating polymerase and free to complex with a newly initiating enzyme to form an initiation

Figure 70. Comparison of initiation pathways
A. Initiation in prokaryotes requires the binding of a sigma factor to the RNA polymerase to form the holoenzyme. This event induces a conformational change in sigma factor, enabling it to recognize specific sequences of proximal promoter elements, leading to transcription initiation.
B. Initiation on class II promoters is proposed to involve a similar step of holoenzyme assembly with general transcription factors binding to RNA polymerase II. This holoenzyme is able to recognize the promoter via TFIID, only when this general transcription factor is bound on the DNA in (a) particular conformational state(s), but not in (an)other(s), that functionally resemble(s) DNA packaging by nucleosomes. The transcriptionally active conformation of TFIID can be induced by a large variety of upstream-bound activators via the function of certain TAFs and soluble co-factors.


RNApol
σ
holo RNApol
σ
+1
pol II
B
F
etc
holo pol II
B
F
etc
IID
activators
IID

competent holoenzyme. TFIID has evolved in eukaryotes, as I have speculated in this chapter, as a specialized chromatin component that is required for the initiation of transcription from otherwise densely packaged chromatin templates typical of eukaryotic genomic organization. In the absence of chromatin, promoter recognition by the holoenzyme might be mediated by other DNA binding factors that interact with one or multiple components of the holoenzyme. Recent observations that YY1 can catalyze initiation by RNA polymerase II in the absence of TBP (Usheva and Shenk, 1994b), may be viewed as such an example. The proposed model for promoter recognition and for the functional roles of TFIID in initiation, furthermore, satisfies the requirements for additional levels of initiation control by evolving activator-effected mechanisms that regulate the accessibility/affinity/recruitement of the holoenzyme to the promoter region.

Whether these common features between prokaryotic and eukaryotic class II initiation are also shared by class I and class III initiation pathways remains to be seen. All eukaryotic transcription initiation requires a multi-protein complex containing TBP (SL1/TIF-1B; TFIID; $TFIIIB_2$) that is involved in template commitment. Its lack of sequence specificity in DNA interactions is compensated for by an additional factor (e.g. UBF; activators/TFIIA etc.; $TFIIIC_2$) which acts co-operatively in promoter recognition. It is currently unclear, however, whether the class I and class III complexes function in the recruitment of a (holo)enzyme complex or just RNA polymerase. The surprizingly low level of sequence identity between homologous general class I and class III factors from yeast and human suggest, though, that these may be relatively recently evolved specialized cases of transcription mechanisms that meet the particular requirements of ribosomal and transfer RNA production.

References

Abmayr, S.M., Feldman, L.D., and Roeder, R.G. (1985). *In vitro* stimulation of specific RNA polymerase II-mediated transcription by the pseudorabies virus immediate early protein. *Cell* **43:** 821-829.

Abmayr, S.M., Workman, J.L., and Roeder, R.G. (1988). The pseudorabies immediate early protein stimulates *in vitro* transcription by facilitating TFIID : promoter interactions. *Genes and Development* **2:** 542-553.

Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryptic and prokaryotic RNA polymerases. *Cell* **42:** 599-610.

Apsit, V., Freeberg, J.A., Chase, M.R., Davis, E.A., and Ackerman, S. (1993). Wheat TFIID TATA binding protein. *Nucl. Acids Res.* **21:** 1494.

Arents, G., Burlingame, R.W., Wang, B.-C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. USA* **88:** 10148-10152.

Arnosti, D.N. and Chamberlin, M.J. (1989). Secondary sigma factor controls transcription of flagellar and chemotaxis genes in *E. coli*. *Proc. Natl. Acad. Sci. USA* **86:** 830-834.

Auble, D.T. and Hahn, S. (1993). An ATP-dependent inhibitior of TBP binding to DNA. *Genes & Development* **7:** 844-856.

Auble, D.T., Hansen, K.E., Mueller, C.G.F., Lane, W.S., Thrner, J., and Hahn, S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. **:** submitted.

Axel, R., Feigelson, P., and Schütz, G. (1976). Analysis of the complexity and diversity of mRNA from chicken liver and oviduct. *Cell* **7:** 247-254.

Benoist, C. and Chambon, P. (1981). *In vivo* sequence requirement of the SV40 early promoter region. *Nature* **290:** 304-310.

Bensimhon, M., Gabarro-Arpa, J., Ehrlich, R., and Reiss, C. (1983). Physical characteristics in eukaryotic promoters. *Nucleic Acids Research* **11:** 4521-4540.

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J., and Guarente, L. (1990). Selected inhibition of activated but not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adaptors. *Cell* **61:** 1199-1208.

Berger, S.L., Piña, B., Silverman, N., Marcus, G.A., Agapite, J., Regier, J.L., Triezenberg, S.J., and Guarente, L. (1992). Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. *Cell* **70:** 251-265.

Bergman, Y., Rice, D., Grosschedl, R., and Baltimore, D. (1984). Two regulatory elements for immunoglobulin kappa light chain gene expression. *Proc. Natl. Acad. Sci. USA* **81:** 7041-7045.

Bohmann, D. and al., e. (1987). Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional

properties of transcription factor AP-1. *Science* **238:** 1386-1392.

Bohr, V.A., Smith, C.A., Okumoto, D.S., and Hanawalt, P.C. (1985). DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene in CHO cells is much more efficient than in the genome overall. *Cell* **40:** 350-369.

Brandt, W.F., de Andrade Rodrigues, J., and von Holt, C. (1988). The amino acid sequence of wheat histone H2B(2). A core histone with a novel reprtitive N-terminal extension. *Eur. J. Biochem.* **173:** 547-554.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50:** 349-383.

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190:** 576-581.

Brock, M.L. and Shapiro, D.J. (1983a). Estrogen regulates the absolute rate of transcription of the *Xenopus laevis* vitellogenin genes. *J. Biol. Chem.* **258:** 5449-5455.

Brock, M.L. and Shapiro, D.J. (1983b). Estrogen stabilizes vitellogenin mRNA against cytoplasmic degradation. *Cell* **34:** 207-214.

Bunick, D., Zandomeni, R., Ackerman, S., and Weinmann, R. (1982a). Mechanism of RNA polymerase II-specific initiation of transcription *in vitro*: ATP requirement and uncapped runoff transcripts. *Cell* **29:** 877-886.

Bunick, D., Zandomeni, R., Ackerman, S., and Weinmann, R. (1982b). Mechanism of RNA polymerase-specific transcription

in vitro: ATP requirement and uncapped runoff transcripts. *Cell* **29:** 877-886.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56:** 549-561.

Buratowski, S., Hahn, S., Sharp, P.A., and Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334:** 37-42.

Buratowski, S. and Zhou, H. (1992). A supressor of TBP mutations encodes an RNA polymerase II transcription factor with homology to TFIIB. *Cell* **71:** 221-230.

Buratowski, S. and Zhow, H. (1992). Transcription factor IID mutants defective for interaction with transcription factor IIA. *Science* **255:** 1030-1032.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K.F. (1969). Factor stimulating transcription by RNA polymerase. *Nature* **221:** 43-46.

Cairns, B.R., Kim, Y.-J., Sayre, M.H., Laurent, B.C., and Kornberg, R.D. (1994). A multisubunit complex containing the *SWI1/ADR6*, *SWI2/SNF2*, *SWI3*, *SNF5*, and *SNF6* gene products isolated from yeast. *Proc. Natl. Acad. Sci. USA* **91:** 1950-1954.

Carcamo, J., Buckbinder, L., and Reinberg, D. (1991). The initiator directs the assembly of a IID-dependent transcription complex. *Proc. Natl. Acad. Sci. USA* **88:** 8052-8056.

Caron, C., Rousset, R., Beraud, C., Monocollin, V., Eglu, J.-M., and Jalinot, P. (1993). Functional and biochemical

interaction of the THLV-1 TAX1 transcactivator with TBP. *EMBO J.* **12:** 4269-4278.

Cavallini, B., Faus, I., Matthes, H., Chipoulet, J.M., Winsor, B., Egly, J.M., and Chambon, P. (1989). Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. *Proc. Natl. Acad. Sci. USA* **86:** 9803-9807.

Cavallini, B., Huet, J., Plassat, J.-L., Sentenac, A., Egly, J.-M., and Chambon, P. (1988). A yeast activity can substitute for the HeLa cell TATA box factor. *Nature* **334:** 77-80.

Chambon, P., Gissinger, F., Mandel, J.L., Kedinger, C., Gniazdowski, M., and Meilhac, M. (1970). Purification and properties of calf thymus DNA-dependent RNA polymerase A and B. *Cold Spring Harbor Symp. Quant. Biol.* **35:** 693-707.

Charnay, P., Mellon, P., and Maniatis, T. (1985). Linker scanning mutagenesis of the 5'-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and nonerytroid cells. *Mol. Cell. Biol.* **5:** 1498-1511.

Chasman, D.I., Flaherty, K.M., Sharp, P.A., and Kornberg, R.D. (1993). Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl. Acad. Sci. USA* **90:** 8174-8178.

Chen, X., Farmer, G., Zhu, H., Praywes, R., and Prives, C. (1993). Cooperative DNA binding of p53 with TFIID (TBP): a possible mechanism for transcriptional activation. *Genes Dev.* **7:** 1837-1849.

Chiang, C.-M., Ge, H., Wang, Z., Hoffmann, A., and Roeder, R.G. (1993). Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerase II and III. *EMBO J.* **12:** 2749-2762.

Chiang, C.-M. and Roeder, R.G. (1993). Expression and purification of general transcription factors by FLAG epitope-tagging and peptide elution. *Peptide Research* **6:** 62-64.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W., and Sharp, P.A. (1988a). Human CCAAT-binding proteins have heterologous sununits. *Cell* **53:** 11-24.

Chodosh, L.A., Oleson, J., Hahn, S., Baldwin, A.S., Guarente, L., and Sharp, P.A. (1988b). A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. *Cell* **53:** 25-35.

Choy, B. and Green, M.R. (1993). Eukaryotic activators function during multiple steps of preinitiation complex assembly. **366:** 531-536.

Chuang, R.Y. and Chuang, L.F. (1975). Increased frequency of initiation of RNA synthesis due to a protein factor from chicken myeloblastosis nuclei. *Proc. Natl. Acad. Sci. USA* **72:** 2935-2939.

Coen, D.M., Weinheimer, S.P., and McKnight, S.L. (1986). A genetic approach to promoter recognition during transinduction of viral gene axpression. *Science* **234:** 53-59.

Colbert, T. and Hahn, S. (1992). A yeast TFIIB-related factor involved in RNA polymerase III transcription. *Genes & Development* **6:** 1940-1949.

Colgan, J. and Manley, J.L. (1992). TFIID can be the rate limiting in vivo for TATA-containing, but not TATA-lacking, RNA polymerase II promoters. *Genes & Development* **6:** 304-315.

Colgan, J., Wampler, S., and Manley, J.L. (1993). Interaction between a transcriptional activator and transcription factor IIB *in vivo*. *Nature* **362:** 549-552.

Comai, L., Tanese, N., and Tjian, R. (1992). The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell* **68:** 965-976.

Conaway, J.W., Bond, M.W., and Conaway, R.C. (1987). An RNA polymerase II transcription system from rat liver. *J. Biol. Chem.* **262:** 8293-8297.

Conaway, J.W. and Conaway, R.C. (1990). An RNA polymerase II transcription factor shares functional properties with Escherichia coli σ^{70}. *Science* **248:** 1550-1553.

Conaway, J.W., Hanley, J.P., Garret, K.P., and Conaway, R.C. (1991). Transcription initiated by RNA polymerase II and transcription factors in liver: Structure and function of ε and τ in initial complex formation. *J. Biol. Chem.* **266:** 7804-7811.

Conaway, R.C. and Conaway, J.W. (1988). ATP activates transcription initiation from promoters by RNA polymerase II in a reversible step prior to RNA synthesis. *J. Biol. Chem.* **263:** 2962-2968.

Concino, M.F., Lee, R.F., Merryweather, J.P., and Weinmann, R. (1984). The adenovirus major late promoter TATA box and

initiation site are both necessary for transcription *in vitro*. *Nucleic Acids Res.* **12:** 7423-7433.

Corden, J., Wasylyk, b., Buchwalder, A., Sassoni-Corsi, P., Kedinger, C., and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science* **209:** 1406-1414.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **82:** 7934-7938.

Cormack, B.P., Strubin, M., Ponticelli, A.S., and Struhl, K. (1991). Functional differences between yeast and Human TFIID Are localized to the highly conserved region. *Cell* **65:** 341-348.

Cormack, B.P. and Struhl, K. (1992). The TATA-binding protein Is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* **69:** 685-696.

Cormack, B.P. and Struhl, K. (1993). Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III transcription. *Science* **262:** 244-248.

Cornelisson, A.W.C.A., Evers, R., and Kock, J. (1988). Structure and sequence of genes encoding subunits of eukaryotic RNA polymerases. *Oxford Surv. Eukaryotic Genes* **5:** 92-131.

Coulombe, B., Killeen, M., Liljelund, P., Honda, B., Xiao, H., Ingles, C.J.and Greenblatt, J. (1992). Identification of three mammalian proteins that bind to the yeast TATA box protein TFIID. *Gene Expr.* **2:** 99-110.

Courey, A.J., Holtzman, D.A., Jackson, S.P., and Tjian, R. (1989). Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59:** 827-836.

Courey, A.J. and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55:** 887-898.

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

Crowley, T.E., Hoey, T., Liu, J.-K., Jan, Y.N., Jan, L.Y., and Tjian, R. (1993). A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature* **361:** 557-561.

Dahmus, M.E. and Kedinger, C. (1983). Transcription of adenovirus-2 major late promoter inhibited by monoclonal antibody directed against RNA polymerase IIA and IIO. *J. Biol. Chem.* **258:** 2303-2307.

Dalrymple, M.A., Peterson-Bjorn, S., Friesen, J.D., and Beggs, J.D. (1989). The product of the *PRP4* gene of *S. cerevisiae* shows homology to β subunits of G proteins. *Cell* **58:** 811-812.

Danos, O. and Mulligan, R.C. (1988). Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. *Proc. Natl. Acad. Sci. USA* **85:** 6460-6464.

Davison, B.L., Egly, J.-M., Mulvhill, E.R., and Chambon, P. (1983). Formation of stable preinitiation complex between eukaryotic class B transcription factors and promoter sequences. *Nature* **301:** 680-686.

de Andrade Rodrigues, J., Brandt, W.F., and Von Holt, C. (1988). The primary structure of the histone H2A(2) type from wheat germ. A core histone type with both, N-terminal and C-terminal extensions. *Eur. J. Biochem.* **173:** 555-560.

DeJong, J. and Roeder, R.G. (1993). A single cDNA, hTFIIA/α, encodes both the p35 and p19 subunits of human TFIIA. *Genes & Development* **7:** 2220-2234.

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.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. *Cell* **32:** 695-706.

Dombroski, A.J., Walter, W.A., and Gross, C.A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes & Development* **7:** 2446-2455.

Dombroski, A.J., Walter, W.A., Record, M.T., Siegele, D.A., and Gross, C.A. (1992). Polypeptides Containing Highly Conserved Regions of Transcription Initiaton Factor σ^{70} Exhibit Specificity of Binding to Promoter DNA. *Cell* **70:** 501-512.

Dostatni, N., Lambert, P.F., Sousa, R., Ham, J., Howley, P.M., and Yaniv, M. (1991). The functional BPV-1 E2 trans-acting protein can act as a repressor by prevemting formation of the initiation complex. *Genes & Development* **5:** 1657-1671.

Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.-C., Zawel, L., Ahn, K., Sancar, A., and Reinberg, D. (1994). Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* **368:** 769-772.

Du, H., Roy, A.L., and Roeder, R.G. (1993). Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. *EMBO J.* **12:** 501-511.

Durrin, L.K., Mann, R.K., Kayne, P.S., and Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation *in vivo*. *Cell* **65:** 1023-1031.

Dynlacht, B.D., Hoey, T., and Tjian, R. (1991). Isolation of coactivators associated with the TATA-box binding protein that mediate transcriptional activation. *Cell* **66:** 563-576.

Dynlacht, B.D., Weinzierl, R.O.J., Admon, A., and Tjian, R. (1993). The $dTAF_{II}80$ subunit of Drosophila TFIID contains b-transducin repeats. *Nature* **363:** 176-178.

Eaton, D., Rodriguez, H., and Vehar, G.A. (1985). Proteolytic processing of human factor VIII. Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. *Biochemistry* **25:** 505-512.

Egly, J.M., Miyamoto, N.G., Moncollin, V., and Chambon, P. (1984). Is actin a transcription initiation factor for RNA polymerase B ? *EMBO J.* **3:** 2363-2371.

Eisenmann, D.M., Arndt, K.M., Ricupero, S.L., Rooney, J.W., and Winston, F. (1992). SPT3 interacts with TFIID to allow normal transcription in Saccharomyces cerevisiae. *Genes Dev.* **6:** 1319-1331.

Eisenmann, D.M., Dollard, C., and Winston, F. (1989). SPT15, the gene encoding the yeast TATA binding factor TFIID Is required for normal transcription initiation In vivo. *Cell* **58:** 1183-1191.

Elliot, T. and Geiduschek, E.P. (1984). Defining a bacteriophage T4 late promoter: absence of a "-35" region. *Cell* **36:** 211-219.

Emanuel, P.A. and Glimour, D.S. (1993). Transcription factor TFIID recognizes DNA sequences downstreem of the TATA element in the HSP70 heat shock gene. *Proc. Natl. Acad. Sci. USA* **90:** 8449-8453.

Emili, A., Greenblatt, J., and Ingles, C.J. (1994). Species-specific interactions of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. *Molec. Cell. Biol.* **14:** 1582-1593.

Feaver, W.J., Gileadi, O., Li, Y., and Kornberg, R.D. (1991). CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67:** 1223-1230.

Feaver, W.J., Svejstrup, J.Q., Bardwell, L., Buratowski, S., Gulyas, K.D., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1993). Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair. *Cell* **75:** 1379-1387.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. *Nature* **355:** 219-224.

Ferreri, K., Gill, G., and Montminy, M. (1994). The cAMP-regulated transcription factor CREB interacts with a

component of the TFIID complex. *Proc. Natl. Acad. Sci.* **91:** 1210-1213.

Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R.A., and Wigler, M. (1988). Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* **8:** 2159-2165.

Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340:** 245-246.

Fikes, J.D., Becker, D.M., Winston, F., and Guarente, L. (1990). Striking conservation of TFIID in Schizosaccharomyces pombe and saccharomyces cerevisiae. *Nature* **346:** 291-294.

Fire, A., Samuels, M., and Sharp, P.A. (1984). Interactions between RNA Polymerase II, Factors, and Template Leading to Accurate Transcription. *J. Biol. Chem.* **259:** 2509-2516.

Flores, O., Ha, I., and Reinberg, D. (1990). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and subunit composition of transcription factor IIF. *J. Biol. Chem.* **260:** 5629-5634.

Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.* **267:** 2786-2793.

Flores, O., Maldonado, E., Burton, Z., Greenblatt, J., and Reinberg, D. (1988). Factors Involved in Specific Transcription by Mammalian RNA Polymerase II. *J. Biol. Chem.* **263:** 10812-10816.

Flores, O., Maldonaldo, E., and Reinberg, D. (1989). Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.* **264:** 8913-8921.

Galau, G.A., Klein, W.H., Davis, M.M., Wold, B.J., Britten, R.J., and Davidson, E.H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. *Cell* **7:** 487-505.

Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., and Chambon, P. (1979). Organisation and sequences at the 5' end of a cloned complete ovalbumin gene. *Nature* **278:** 428-434.

Gans, M., Audit, C., and Masson, M. (1975). *Gen.* **81:** 683-704.

Gardella, T., Moyle, H., and Susskind, M. (1989). A mutant *E. coli* σ70 subunit of RNA polymerase with altered promoter specificity. *J. Mol. Biol.* **206:** 579-590.

Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R.G., and Chua, N.-H. (1990). Arabidopsis thaliana contains two genes for TFIID. *Nature* **346:** 390-394.

Ge, H. and Roeder, R.G. (1994). The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA binding protein. *J. Biol. Chem.* **:** in press.

Geiger, J.H., Kim, Y., Hahn, S., and Sigler, P.B. (1994). Crystal structure of yeast TBP at 2.1Å resolution. *Biochemistry* **33:** in the press.

Gentz, R., Chen, C.-H., and Rosen, C.A. (1989). Bioassay for trans-activation using purified human immunodeficiency virus tat-encoded protein: Trans-activation requires mRNA synthesis. *Proc. Natl. Acad. Sci. USA* **86:** 821-824.

Georgakopoulos, T. and Thireos, G. (1992). Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *EMBO J.* **11:** 4145-4152.

Gerard, M., Fischer, L., Moncollin, V., Chipoulet, J.-M., Chambon, P., and Egly, J.-M. (1991). Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. biol. Chem.* **266:** 20940-20945.

Gerster, T., Balmaceda, C.G., and Roeder, R.G. (1990). The cell type-specific octamer transcription factor OTF-2 has two domains required for the activation of transcription. *EMBO J.* **6:** 1323-1330.

Gill, G., Pascal, E., Tseng, Z.H., and Tjian, R. (1994). A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. *Proc. Natl. Acad. Sci.* **91:** 192-196.

Gill, G. and Tjian, R. (1991). A highly conserved domain of TFIID displays species specificity *in vivo*. *Cell* **65:** 333-340.

Gilmour, D.S., Dietz, T.J., and Elgin, S.C.R. (1990). UV cross-linking identifies four polypeptides that require the TATA box to bind to the *Drosophila hsp70* promoter. *Molec. Cell. Biol.* **10:** 4233-4238.

Goodrich, J.A., Hoey, T., Thut, C.J., Admon, A., and Tjian, R. (1993). Drosophilla TAF40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* **75:** 519-530.

Goodrich, J.A. and Tjian, R. (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77:** 145-156.

Green, M.R., Treismann, R., and Maniatis, T. (1983). Transcriptional activation of cloned human βglobin genes by viral immediate-early gene products. *Cell* **35:** 137-148.

Greenblatt, J. and Li, J. (1981). Interaction of the sigma factor and the *nusA* gene protein of *E. coli* with RNA polymerase in the initiation-termination cycle of transcription. *Cell* **24:** 421-428.

Gribskov, M. and Burgess, R.R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SPO1 and Phage T4 are homologous proteins. *Nucleic Acids Res.* **14:** 6745-6763.

Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W., and Watson, J.D. (1961). Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*. *Nature* **190:** 581-585.

Grosschedl, R. and Birnstiel, M.L. (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. USA* **77:** 1432-1436.

Grosschedl, R. and Birnstiel, M.L. (1982). Delimitation of far upstream sequences required for maximal *in vitro* transcription of an H2A histone gene. *Proc. Natl. Acad. Sci. USA* **79:** 297-301.

Grosveld, G.C., De Boer, E., Shewmaker, C.K., and Flavell, R.A. (1982). DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. *Nature* **295:** 120-126.

Grosveld, G.C., Shewmaker, C.K., Jat, P., and Flavell, R.A. (1981). Localization of DNA sequences necessary for transcription of the rabbit βglobin gene *in vitro*. *Cell* **25:** 215-226.

Guan, C., U., P., Riggs, P.D., and Inouye, H. (1987). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67:** 21-30.

Guarente, L. (1988). UASs and enhancers: common mechansims of transcriptional activation in yeast and mammals. *Cell* **52:** 303-305.

Guzder, S.N., Qiu, H., Sommers, C.H., Sung, P., Prakash, L., and Prakash, S. (1994). DNA repair gene RAD3 of S. cerevisiae is essential for transcription by RNA polymerase II. *Nature* **367:** 91-94.

Haass, M.M. and Feix, G. (1992). Two different cDNA encoding TFIID proteins of maize. *FEBS Lett.* **301:** 294-298.

Hagemeier, C., Bannister, A.J., Cook, A., and Kouzarides, T. (1993a). The activation domain of transcription factor PU.1 binds the retinoblastoma (RB) protein and the transcription factor TFIID *in vitro*: RB shows sequence similarity to TFID and TFIIB. *Proc. Natl. Acad. Sci. USA* **90:** 1580-1584.

Hagemeier, C., Cook, A., and Kouzarides, T. (1993b). The retinoblastoma protein binds E2F residues requred for activation *in vivo* and TBP binding in vitro. **21:** 4998-5004.

Hagemeier, C., Walker, S., Caswell, R., Kouzarides, T., and Sinclair, J. (1992). The human cytomegalovirus 80-kilodalton but not the 72-kilodalton immediate-early protein transcactivates heterologous promoters in a TATA box-dependent mechanism and interacts directly with TFIID. *J. Virol.* **66:** 4452-4456.

Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L. (1989a). Isolation of the gene encoding the yeast TATA binding protein TFIID: A Gene identical to the SPT15 suppressor of ty element insertions. *Cell* **58:** 1173-1181.

Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L. (1989b). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl. Acad. Sci. USA* **86:** 5718-5722.

Hahn, S. and Guarente, L. (1988). Yeast HAP2 and HAP3: Transcriptional activators in a heteromeric complex. *Science* **240:** 317-321.

Hai, T., Horikoshi, M., Roeder, R., and Green, M. (1988). Analysis of the role of the transcription factor ATF in the assembly of a functional preinitiation complex. *Cell* **54:** 1043-1051.

Ham, J., Steger, G., and Yaniv, M. (1994). Cooperativity in vivo between the E2 transactivator and the TATA box binding protein depends on core promoter structure. *EMBO J.* **13:** 147-157.

Handa, H., Kingston, R.E., 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.

Hansen, U., Tenen, D.G., Livingdton, D.M., and Sharp, P.A. (1981). T Antigen repression of SV40 early transcription from two promoters. *Cell* **27:** 603-612.

Hashimoto, S., Fujita, H., Hasegawa, S., Roeder, R.G., and Horikoshi, M. (1992). Conserved structural motifs within the N-terminal domain of TFIIDτ from *Xenopus*, mouse and human. *Nucleic Acids Res.* **20:** 3788.

Hastie, N.D. and Bishop, J.O. (1976). The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* **9:** 761-774.

Hateboer, G., Timmers, H.T.M., Rustgi, A.K., Billaud, M., van't Veer, L.J., and Bernards, R. (1993). TATA-binding protein and the retinoblastoma gene product bind to overlapping epitopes on c-Myc and adenovirus E1A protein. *Proc. Natl. Acad. Sci.* **90:** 8489-8493.

Hawley, D.K. and McClure, R.W. (1980). Compilation and analysis of *E. coli* promoter elements. *Nucleic Acids Res.* **11:** 2237-2255.

Hawley, D.K. and Roeder, R.G. (1985). Separation and Partial Characterization of Three Functional Steps in Transcription Initiation by Human RNA Polymerase II. *J. Biol. Chem.* **260:** 8163-8172.

Hawley, D.K. and Roeder, R.G. (1987). Functional Steps in Transcription Initiation and Reinitiation from the Major Late Promoter in a HeLa Nuclear Extract. *J. Biol. Chem.* **262:** 3452-3461.

Heintz, N. and Roeder, R.G. (1984). Transcription of human histone genes in extracts from synchronized HeLa cells. *Proc. Natl. Acad. Sci. USA* **81:** 2713-2717.

Helmann, J.D. and Chamberlin, M.J.A. (1988). Structure and function of bacterial sigma factors. *Ann. Rev. Biochem.* **57:** 839-872.

Henry, N.L., Sayre, M.H., and Kornberg, R.D. (1992). Purification and characterization of yeast RNA polymerase II general initiation factor g. *J. biol. Chem.* **267:** 23388-23392.

Himmelfarb, H.J., Pearlberg, J., Last, D.H., and Ptashne, M. (1990). GAL11P: a yeast mutation that potentiates the effect of weak GAL4-derived activators. *Cell* **63:** 1299-1309.

Hiraoka, Y., Toda, T., and Yanagida, M. (1984). The NDA3 gene of fission yeast encodes beta-tubulin: A cold-sensitive nda3 mutation reversibly blocks spindle formation and chromosome movement in mitosis. *Cell* **39:** 349-358.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes & Development* **6:** 1707-1715.

Hirschhorn, R.R., Aller, P., Yuan, Z.A., Gibson, C.W., and Baserga, R. (1984). Cell-cycle-specific cDNAs from mammalian cells temperature sensitive for growth. *Proc. Natl. Acad. Sci.* **81:** 6004-6008.

Hisatake, K., Hasegawa, S., Takada, R., Nakatani, Y., Horikoshi, M., and Roeder, R.G. (1993). The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. *Nature* **362:** 179- 181.

Hochuli, E., Döbeli, H., and Schacher, A. (1987). New metal chelate adsorbant selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* **411:** 177-184.

Hoeijmakers, J.H.J. (1990). Cryptic initiation sequence revealed. *Nature* **343:** 417-418.

Hoey, T., Dynlacht, B.D., Peterson, M.G., Pugh, B.F., and Tjian, R. (1990). Isolation and characterization of the drosophila gene encoding the TATA Box binding brotein, TFIID. *Cell* **61:** 1179-1186.

Hoey, T., Weinzierl, R.O.J., Gill, G., Chen, J.-L., Dynlacht, B.D., and Tjian, R. (1993). Molecular cloning and functional analysis of drosophila TAF110 reveal properties expected of coactivators. *Cell* **72:** 247-260.

Hoffmann, A., Horikoshi, M., Wang, C.K., Schroeder, S., Weil, P.A., and Roeder, R.G. (1990). Cloning of the Schizosaccharomyces pombe TFIID gene reveals a strong conservation of functional domains present in Saccharomyces cerevisiae TFIID. *Genes & Development* **4:** 1141-1148.

Hoffmann, A. and Roeder, R.G. (1991). Purification of his-tagged proteins in non-denaturing conditions suggests a convenient method for protein interaction studies. *Nucleic Acids Res.* **19:** 6337.

Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M., and Roeder, R.G. (1990). Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature* **346:** 387-390.

Holdsworth, M.J., Grierson, C., Schuch, W., and Bevan, M. (1992). DNA-binding properties of cloned TATA-binding protein from potato tubers. *Plant Molec. Biol.* **19:** 455-464.

Hopp, T.P., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Cerretti, D.P., Urdal, D.L., and Conlon, P.J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6:** 1204-1210.

Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T., and Roeder, R. (1992). Transcription factor TFIID induces DNA bending upon binding to the TATA element. *Proc. Natl. Acad. Sci. USA* **89:** 1060-1064.

Horikoshi, M., Carey, M.F., Kakidani, H., and Roeder, R.G. (1988a). Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell* **54:** 665-669.

Horikoshi, M., Hai, T., Lin, Y.-S., Green, M.R., and Roeder, R.G. (1988b). Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* **54:** 1033-1042.

Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A., and Roeder, R.G. (1989a). Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. *Nature* **341:** 299-303.

Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A., and Roeder, R.G. (1989b). Purification of a yeast TATA box-binding protein that exhibits human transcription factor IID activity. *Proc. Natl. Acad. Sci. USA* **86:** 4843-4847.

Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P.A., and Roeder, R.G. (1990). Analysis of structure-function relationships of Yeast TATA box binding factor TFIID. *Cell* **61:** 1171-1178.

Horikoshi, N., Maguire, K., Kralli, A., Maldonado, E., Reinberg, D., and Weinmann, R. (1991). Direct interactions between adenovirus E1A protein and the TATA box binding transcription factor IID. *Proc. Natl. Acad. Sci. USA* **88:** 5124-5128.

Hu, J.C. and Gross, C.A. (1988). Mutations in rpoD that increase expression of genes in the mal regulon of E. coli K-12. *J. Mol. Biol.* **203:** 15-27.

Hu, S.-L. and Manley, J.L. (1981). DNA sequence required for initiation of transcription *in vitro* from the major late promoter of adenovirus 2. *Proc. Natl. Acad. Sci. USA* **78:** 820-824.

Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B., and Bax, A. (1992). Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256:** 632-638.

Ingles, C.J., Shales, M., Cress, W.D., Triezenberg, S.J., and Greenblatt, J. (1991). Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature* **351:** 588-590.

Inostroza, J.A., Mermelstein, F.H., Ha, I., Lane, W.S., and Reinberg, D. (1992). Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. *Cell* **70:** 477-489.

Jacob, F. and Monod, J. (1962). On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* **26:** 193-211.

Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R.A., Nordheim, A., and Stunnenberg, H.G. (1991). Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci. USA* **88:** 8972-8976.

Kakidani, H. and Ptashne, M. (1988). GAL4 activates gene expression in mammalian cells. *Cell* **52:** 161-167.

Kao, C.C., Leiberman, P.M., Schmidt, M.C., Zhou, Q., Pei, R., and Berk, A.J. (1990). Cloning of a transcriptionally active human TATA binding factor. *Science* **248:** 1646-1650.

Kashanichi, F., Piras, G., Radonovich, M.F., Duvall, J.F., Fattaey, A., Chiang, C.-M., Roeder, R.G., and Brady, J.N. (1994). Direct interaction of human TFIID with the HIV-1 transactivator Tat. *Nature* **367:** 295-299.

Käufer, N.F., Simanis, V., and Nurse, P. (1985). Fission yeast Schizosaccaromces pombe correctly excises a mammalian RNA transcript intervening sequence. *Nature* **318:** 78-80.

Kawata, T., Minami, M., Tamura, T.-a., Sumita, K., and Iwabuchi, M. (1992). Isolation and characterization of a cDNA clone encoding the TATA box-binding protein (TFIID) from wheat. *Plant Molec. Biol.* **19:** 867-872.

Keleher, C.A., Redd, M.J., Schultz, J., Carlson, M., and Johnson, A.D. (1992). Ssn6-Tup1 is a general repressor of transcription in yeast. *Cell* **68:** 709-719.

Kelleher III, R.J., Flanagan, P.M., Chasman, D.I., Ponticelli, A.S., Struhl, K., and Kornberg, R. (1992). Yeast

and human TFIIDs are interchangeable for the response to acidic transcriptional activators in vitro. *Genes Dev.* **6:** 296-303.

Kennison, J.A. and Tamkun, J.W. (1988). Dosage-dependent modifiers of *Polycomb* and *Antennapedia* mutations in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **85:** 8136-8140.

Kerr, L.D., Ransone, L.J., Wamsley, P., Scmitt, M.J., Boyer, T.G., and Zhow, Q. (1993). Association between proto-oncoprotein Rel and TATA-binding protein mediates transcriptional activation by NF-kB. *Nature* **356:** 412-419.

Kerrigan, L.A. and Kadonaga, J.T. (1992). Periodic binding of individual core histones to DNA: inadvertant purification of the core histone H2B as a putative enhancer-binding factor. *Nucleic Acids Res.* **20:** 6673-6680.

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., and Crabtree, G.R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366:** 170-174.

Killeen, M., Coulombe, B., and Greenblatt, J. (1992). Recombinant TBP, Transcription Factor IIB, and RAP30 are sufficient for promoter recognition by mammalian RNA polymerase II. *J. Biol. Chem.* **267:** 9463-9466.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365:** 520-527.

Kim, T.K., Hashimoto, S., III, R.J.K., Flanagan, P.M., Kornberg, R.D., Horikoshi, M., and Roeder, R.G. (1994). Effects of activation-defective TBP mutations on specific

TFIIB-TBP promoter interactions and TFIIB recruitment in yeast. ? **in the press:**

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365:** 512-520.

Kitajima, S., Tanaka, Y., Kawaguchi, T., Nagaoka, T., Weissman, S.M., and Yasukochi, Y. (1990). A heteromeric transcription factor required for mammalian RNA polymerase II. *Nucleic Acids Res.* **18:** 4843-4849.

Kokubo, T., Gong, D.-W., Wootton, J.C., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1994). Molecular cloning of Drosophila TFIID subunits. *Nature* **367:** 484-487.

Kokubo, T., Gong, D.-W., Yamashita, S., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1993a). Drosophila 230-kD TFIID subunit,
a functional homolog of the human cell cycle gene product, negatively regulates DNA binding of the TATA box-binding subunit of TFIID. *Genes Dev.* **7:** 1033-1046.

Kokubo, T., Takada, R., Yamashita, S., Gong, D.-W., Roeder, R.G., Horikoshi, M., and Nakatani, Y. (1993b). Identification of TFIID components required for transcriptiuonal activation by upstream stimulatory factor. *J. Biol. Chem.* **268:** 17554-17558.

Kokubo, T., Yamashita, S., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1994). Interaction between the N-terminal domain of the 230 kDa TFIID subunit and TFIIDt regulates TATA box binding. *Proc. Natl. Acad. Sci. USA* **:** in the press.

Koleske, A.J., Buratowski, S., Nonet, M., and Young, R.A. (1992). A Novel transcription factor reveals a functional

link between the RNA polymerase II CTD and TFIID. *Cell* **69:** 883-894.

Koleske, A.J. and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* **368:** 466-469.

Kolodziej, P.A. and Young, R.A. (1989). RNA polymerase II subunit RPB3 is an essential component of the mRNA transcription apparatus. *Mol. Cell. Biol.* **9:** 5387-5394.

Kretzschmar, M., Meisterernst, M., and Roeder, R.G. (1993). Identification of human DNA topoisomerase I as a cofactor for activator-dependent transcription by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **90:** 11508-11512.

Kretzschmar, M., Stelzer, G., Roeder, R.G., and Meisterernst, M. (1994). RNA polymerase II co-factor PC2 facilitates activation of transcription by GAL-AH *in vitro*. *Molec. Cell Biol.* **:** in press.

Kustu, S., Santero, E., Keener, J., Popham, D.L., and Weiss, D. (1989). Expression of σ54 (ntr)-dependent genes os probably united by a common mechanism. *Microb. Rev.* **53:** 367-376.

Langford, C.J., Klinz, F.-J., Donath, C., and Gallwitz, D. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal in yeast. *Cell* **36:** 645-653.

Laurent, B.C. and Carlson, M. (1992). Yeast, SNF2/SWI2, SNF5, and SNF6 proteins function coordinately with the gene-specific transcriptional activators GAL4 and Bicoid. *Genes & Development* **6:** 1707-1715.

Laurent, B.C., Treitel, M.A., and Carlson, M. (1991). Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proc Natl Acad Sci USA* **88:** 2687-2691.

Lech, K., Anderson, K., and Brent, R. (1988). DNA-bound fos proteins activate transcription in yeast. *Cell* **52:** 179-184.

Lee, D.K., DeJong, J., Hashimoto, S., Horikoshi, M., and Roeder, R.G. (1992). TFIIA induces conformational changes in TFIID via interactions with the basic repeat. *Molec. Cell. Biol.* **12:** 5189-5196.

Lee, D.K., Horikoshi, M., and Roeder, R.G. (1991). Interaction of TFIID in the minor groove of the TATA element. *Cell* **67:** 1241-1250.

Lee, M.G. and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature* **327:** 31-35.

Leive, L. and Kollin, V. (1967). Synthesis, utilization and degradation of lactose operon mRNA in *Escherichia coli*. *J. Mol. Biol.* **24:** 247-259.

Lescure, B. and Arcangioli, B. (1984). Yeast RNA polymerase II initiates transcription *in vitro* at TATA sequences proximal to potential nob-B forms of the DNA template. *EMBO J.* **3:** 1067-1073.

Lescure, B., Bennetzen, J., and Sentenac, A. (1981a). *In vitro* transcription of the yeast alcohol dehydrogenase I gene by homologous RNA polymerase B (II). *J. Biol. Chem.* **256:** 11018-11024.

Lescure, B., Williamson, V., and Sentenac, A. (1981b). Efficient and selective initiation by yeast RNA polymerase B in a dinucleotide-primed reaction. *Nucleic Acids Res.* **9:** 31-45.

Lesley, S.A. and Burgess, R.R. (1989). Characterization of the E. coli transcription factor σ70: Localization of a region involved in the interaction with core RNA polymerase. *Biochemistry* **28:** 7728-7734.

Li, S. and Donelson, J.E. (1993). The gene for the TATA box-binding protein of *Onchocerca volvulus*. *Molec. Biochem. Parasitol.* **61:** 321-324.

Li, Y., Flanagan, P.M., Tschochner, H., and Kornberg, R.D. (1994). RNA polymerase II initiation factor interactions and transcription start site selection. *Science* **263:** 805-807.

Lichtsteiner, S. and Tjian, R. (1993). Cloning and properties of the Caenorhabditis elegans TATA-box binding protein. *Proc. Natl. Acad. Sci. USA* **90:** 9673-9677.

Lieberman, P.M. and Berk, A.J. (1991). The Zta trans-activator protein stabilizes TFIID association with promoter DNA by direct protein-protein interaction. *Genes Dev.* **5:** 2441-2454.

Lieberman, P.M., Schmidt, M.C., Kao, C.C., and Berk, A.J. (1991). Two Distinct Domains in the yeast Transcription Factor IID and evidence for a TATA box-induced conformational change. *Molec. Cell. Biol.* **11:** 63-74.

Liljelund, P., Ingles, C.J., and Greenblatt, J. (1993). Altered promoter binding of the TATA box-binding factor induced by the transcriptional activation domain of VP16 and suppressed by TFIIA. *Mol. Gen. Genet.* **241:** 694-699.

Lin, Y.-S., HA, I., Maldonado, E., Reinberg, D., and Green, M.R. (1991). Binding of general transcription factor TFIIB to an acidic activating region. **353:** 569-571.

Lin, Y.S. and Green, M.R. (1991). Mechanism of action of an acidic transcriptional activator *in vitro*. *Cell* **64:** 971-981.

Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G., and Rutter, W.J. (1970). Specific inhibition of nuclear RNA polymerase II by α-amanitin. *Science* **170:** 447.

Liu, H.T., Gibsen, R.R., Hirschhorn, R.R., Rittling, S., Baserga, R., and Mercer, W.E. (1985). Expression of thymidine kinase and dihydrofolate reductase genes in mammalian ts mutants of the cell cycle. *J. Biol. Chem.* **260:** 3269-3274.

Liu, X., Miller, C.W., Koeffler, P.H., and Berk, A.J. (1993). p53 activation domain binds TATA-box-binding polypeptide and a neighboring p53 domain inhibits transcription. *Mol. Cell. Biol.* **13:** 3291-3300.

Lobo, S.M., Tanaka, M., Sullivan, M.L., and Hernandez, N. (1992). A TBP Complex Essential for Transcription from TATA-Less but not TATA-Containing RNA Polymerase III Promoters Is Part of the TFIIIB Fraction. **71:** 1029-1040.

Lonetto, M., Gribskov, M., and Gross, C.A. (1992). The sigma 70 family: sequence conservation and evolutionar relationships. *J. Bacteriol.* **174:** 3843-3849.

Lopez-de-Leon, A., Librizzi, M., Tuglia, K., and Willis, I.M. (1992). *PCNF4* Encodes an RNA Polymerase III Transcription Factor with Homology to TFIIB. **71:** 211-220.

Losick, R. and Pero, J. (1981). Cascades of sigma factors. *Cell* **25:** 582-584.

Losson, R., Fuchs, R.P.R., and Lacroute, F. (1985). Yeast promoters URA1 and URA3: Examples of positive control. *J. Mol. Biol.* **185:** 65-81.

Lu, H., Zawel, L., Fisher, L., Egly, J.-M., and Reinberg, D. (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358:** 641-645.

Ma, D., Watanabe, H., Mermelstein, F., Admon, A., Ogurie, K., Sun, X., Wada, T., Imai, T., Shiroya, T., Reinberg, D., and Handa, H. (1993). Isolation of a cDNA encoding the largest subunit of TFIIA reveals functions important for activated transcription. *Genes & Development* **7:** 2246-2257.

Mack, D.H., Vartikar, J., Pipas, J.M., and Laimins, L.A. (1993). Specific repression of TATA-mediated but not initiator-mediated transcription by wild type p53. **363:** 281-383.

Magasanik, B. and Neidhart, F.C. (1987). Regulation of carbon and nitrogen utilization, in *Escherichia coli* and *Salmonella typhirium*. Vol. Cellular and Molecular Biology Washington, D.C., American Society for Microbiology. pp. 1318-1325.

Maheswaran, S., Lee, H., and Sonenshein, G.E. (1994). Intracellular association of the protein product of the c-*myc* oncogene with the TATA-binding protein. *Molec. Cell. Biol.* **14:** 1147-1152.

Malik, S., Hisatake, K., Sumimoto, H., Horikoshi, M., and Roeder, R.G. (1991). Sequence of general transcription

factor TFIIB and relationships to other initiation factors. *Proc. Natl. Acad. Sci. USA* **88:** 9553-9557.

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.

Manley, J.L., Sharp, P.A., and Gefter, M.L. (1979). RNA synthesis in isolated nuclei: *in vitro* initiation of adenovirus 2 major late mRNA precursor. *Proc. Natl. Acad. Sci. USA* **76:** 160-164.

Mann, R.K. and Grunstein, M. (1992). Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene *in vivo*. *EMBO J.* **11:** 3297-3306.

Martinez, E., Chiang, C.-M., Ge, H., and Roeder, R.G. (1994). TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.* **:** in press.

Mathis, D.J. and Chambon, P. (1981). The SV40 early region TATA box is required for accurate in vitro initiation of transcription. *Nature* **290:** 310-315.

Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem* **255:** 11992-11996.

Matsumoto, S. and Yanagida, M. (1985). Histone gene organization of fission yeast: a common upstream sequence. *EMBO J.* **4:** 3531-3538.

Maxam, M.E. and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes & Development* **8:** 515-524.

McAndrew, M.B., Read, M., Sims, P.F.G., and Hyde, J.E. (1993). Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T rich human malaria parasite *Plasmodium falciparum*. *Gene* **124:** 165-171.

McCracken, S. and Greenblatt, J. (1991). Related RNA polymerase-binding regions in human RAP30/74 and *Escherichia coli* σ70. *Science* **253:** 900-902.

McGowan, J.J., Emerson, S.U., and Wagner, R.R. (1982). The plus-strand leader RNA of VSV inhibits DNA-dependent transcription of adenovirus and SV40 genes in a soluble whole-cell extract. *Cell* **28:** 325-333.

McKnight, G.S. and Palmiter, R.D. (1979). Transcriptional regulation of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. *J. Biol. Chem.* **254:** 9050-9058.

McKnight, S.L. and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. *Science* **217:** 316-324.

Meisterernst, M., Horikoshi, M., and Roeder, R.G. (1990). Recombinant yeast TFIID, a general transcription factor, mediates activation by the gene-specific factor USF in a chromatin assembly assay. *Proc. Natl. Acad. Sci. USA* **87:** 9153-9157.

Meisterernst, M. and Roeder, R.G. (1991). Family of proteins that interact with TFIID and regulate promoter activity. *Cell* **67:** 557-567.

Meisterernst, M., Roy, A.L., Lieu, H.M., and Roeder, R.G. (1991). Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* **66:** 981-993.

Mellon, I. and Hanawalt, P.C. (1989). Induction of the Escherichia coli lactose operon selectivity increases repair of its transcribed DNA strand. *Nature* **342:** 95-98.

Merino, A., Madden, K.R., Lane, W.S., Champoux, J.J., and Reinberg, D. (1993). DNA topoisomerase I is involved in boyh repression and activation of transcription. *Nature* **365:** 227-232.

Miller, J. and Reznikoff, W.S., Ed. (1978). The Operon. New York, Cold Spring Harbor Press.

Mishoe, H., Brady, J.N., Lancz, G., and Salzman, N.P. (1984). *In vitro* transcription initiation by purified RNA polymerase II within the adenovirus major late promoter region. *J. Biol. Chem.* **258:** 2236-2242.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional Regulation in Mammalian Cells by sequence-specific DNA binding proteins. *Science* **245:** 371-378.

Mizushima-Sugano, J. and Roeder, R.G. (1986). Cell-type-specific transcription of an immunoglobulin k light chain gene *in vitro*. *Proc. Natl. Acad. Sci. USA* **83:** 8511-8515.

Moncollin, V., Fischer, L., Cavallini, B., Egly, J.-M., and Chambon, P. (1992). Class II (B) general transcription

factor (TFIIB) that binds to the template-committed preinitiation complex is different from general transcription factor BTF3. *Proc. Natl. Acad. Sci. USA* **89:** 397-401.

Monod, J. and Jacob, F. (1962). General conclusions: Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harbor Symp. Quant. Biol.* **26:** 389-401.

Morgenstern, J.P. and Land, H. (1990). Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucl. Acids Res.* **18:** 3587-3596.

Muchardt, C. and Yaniv, M. (1993). A human homologue of Saccharomyces cerevisisae SNF2/SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J.* **12:** 4279-4290.

Muhich, M.L., Iida, C.T., Horikoshi, M., Roeder, R.G., and Parker, C.S. (1990). cDNA clone encoding Drosophila transcription factor TFIID. *Proc. Natl Acad. Sci. USA* **87:** 9148-9152.

Nagai, K. (1990). Cryptic initiation sequence revealed. *Nature* **343:** 418.

Nagai, K. and Thøgersen, H.C. (1984). Generation of β-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*. *Nature* **309:** 810-812.

Nakajima, N., Horikoshi, M., and Roeder, R.G. (1988). Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol. Cell. Biol.* **8:** 4028-4040.

Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Besnard, F., Roeder, R.G., and Freese, E. (1990). A downstreem initiation element required for efficient TATA box binding and in vitro function of TFIID. *Nature* **348:** 86-88.

Nikolov, D.B. and Burley, S.K. (1994). 2.1Å resolution refined structure of a TATA box-binding protein (TBP, TFIIDτ). *Nature Struct. Biol.* **:** in press.

Nikolov, D.B., Hu, S.-H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.-H., Roeder, R.G., and Burley, S.K. (1992). Crystal structure of TFIID TATA-box binding protein. *Nature* **360:** 40-46.

Nishizawa, M., Suzuki, Y., Nogi, Y., Matsumoto, K., and Fukusawa, T. (1990). Yeast Gal11 protein mediates the transcriptional activation signal of two different transacting factors, Gal4 and general regulatory factor I/repressor/activator site binding protein 1/translation upstream factor. *Proc. Natl. Acad. Sci. USA* **87:** 5373-5377.

Ohana, B., Moore, P.A., Ruben, S.M., Southgate, C.D., Green, M.R., and Rosen, M.R. (1993). The type 1 human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionarily conserved genes. *Proc. Natl. Acad. Sci. USA* **90:** 138-142.

Ohkuma, Y. and Roeder, R.G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* **368:** 160-163.

Olesen, J., Hahn, S., and Guarente, L. (1987). Yeast HAP2 and HAP3 activators both bind to the CYC1 upstream activation site, UAS2, in an interdependent manner. *Cell* **51:** 953-961.

Overduin, M., Rios, C.B., Mayer, B.J., Baltimore, D., and Cowburn, D. (1992). Three-dimensional solution structure of the src homology 2 domain of c-abl. *Cell* **70:** 697-704.

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55:** 1119-1150.

Parker, C.S., Ng, S., and Roeder, R.G., Ed. (1976). Selective transcription of the 5s RNA genes in isolated chromatin by purified RNA polymerase III. Molecular Mechanisms of gene expression. New York: Academic Press.

Parker, C.S. and Roeder, R.G. (1977). Selective and accurate transcription of yhe Xenopus laevis 5S RNA genes in isolated chromatin by purified RNA polymerase III. *Proc. Natl. Acad. Sci. USA* **74:** 44-48.

Parker, C.S. and Topol, J. (1984a). A *Drosophila* RNA polymerase II transcription factor bind to the regulatory site of an hsp70 gene. *Cell* **37:** 273-283.

Parker, C.S. and Topol, J. (1984b). A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. *Cell* **36:** 357-369.

Pascal, E. and Tjian, R. (1991). Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes & Development* **5:** 1646-1656.

Patel, D.J., Ikuta, S., Kozlowski, S., and Itakura, K. (1983). Sequence dependence of hydrogen exchange kinetics in DNA duplexes at the individual base pair level in solution. *Proc. Natl. Acad. Sci. USA* **80:** 2184-2188.

Payne, J.M., Laybourn, P.J., and Dahmus, M.E. (1989). The transition of RNA polymerase II from initiation to elongation is associated with the phosphorylation of the C-terminal domain of subunit IIa. *J. Biol. Chem.* **264:** 3310-3321.

Peterson, C.L. and Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell* **68:** 573-583.

Peterson, M.G., Tanese, N., Pugh, B.F., and Tjian, R. (1990). Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* **248:** 1625-1630.

Pinto, I., Ware, D.E., and Hampsey, M. (1992). The Yeast SUA7 Gene Encodes A Homolog of Human Transcription Factor TFIIB and is Required for Normal Start Site Selection in Vivo. **68:** 977-988.

Poon, D., Schroeder, S., Wang, C.K., Yamamoto, T., Horikoshi, M., Roeder, R.G., and Weil, P.A. (1991). The conserved carboxy-terminal domain of Saccharomyces cerevisiae TFIID Is sufficient to support normal cell groth. *Molec. Cell. Biol.* **11:** 4809-4821.

Poon, D. and Weil, P.A. (1993). Immunopurification of yeast TATA-binding protein and associated factors. *J. Biol. Chem.* **268:** 15325-15328.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989). Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243:** 629-635.

Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1989). Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. *Molec. Cell. Biol.* **9:** 1465-1475.

Prickett, K.S., Amberg, D.C., and Hopp, T.P. (1989). A calcium dependent antibody for identification and purification of recombinant proteins. *BioTechniques* **7:** 580-589.

Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequence in eukaryotic messenger RNA. *Nature* **263:** 211-214.

Ptashne, M. (1986). A genetic switch. Cambridge, Massachusetts, Cell Press/Blackwell Scientific.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature* **335:** 683-689.

Pugh, B.F. and Tjian, R. (1990). Mechanism of transcriptional activation by Sp1: evidence of coactivation. *Cell* **61:** 1187-1197.

Pugh, B.F. and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes & Development* **5:** 1935-1945.

Purnell, B.A., Emanuel, P.A., and Gilmour, D.S. (1994). TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes & Development* **8:** 830-842.

Purnell, B.A. and Gilmour, D.S. (1993). Contribution of sequences downstream of the TATA element to a protein-DNA complex containing the TATA-binding protein. *Molec. Cell. Biol.* **13:** 2593-2603.

Ragimov, N., Krauskoft, A., Navot, N., Rotter, V., Oren, M., and Aloni, Y. (1993). Wild-type but not mutant p53 can repress transcription initiation in vitro by interfering with

the binding of basal transcription factors to the TATA motif. *Oncogene* **8:** 1183-1193.

Ray, B.L., White, C.I., and Haber, J.E. (1991). The TSM1 gene of Saccharomyces cerevisiae overlaps the MAT locus. *Current Genetics* **20:** 25-31.

Reddy, P. and Hahn, S. (1991). Dominant negative mutation in yeast TFIID define a bipartite DNA-binding region. *Cell* **65:** 349-357.

Reinberg, D., Horikoshi, M., and Roeder, R.G. (1987). Factors involved in specific transcription by mammalian RNA polymerase II: functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *J. Biol. Chem.* **262:** 3322-3330.

Reinberg, D. and Roeder, R.G. (1987). Factors involved in specific transcription by RNA polymerase II: purification and functional analysis of initiation factors IIB and IIE. *J. Biol. Chem.* **262:** 3310-3321.

Reznikoff, W.S., Siegele, D.A., Cowing, D.W., and Gross, C.A. (1985). The regulation of transcription in bacteria. *Annual Review of Genetics* **19:** 355-387.

Rio, D., Robbins, A., Myers, R., and Tjian, R. (1980). Regulation of simian virus 40 early transcription *in vitro* by a purified tumor antigen. *Proc. Natl. Acad. Sci. USA* **77:** 5706-5710.

Roberts, S.G.E., Ha, I., Maldonado, E., Reinberg, D., and Green, M.R. (1993). Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. *Nature* **363:** 741-744.

Roeder, R.G. (1976). Eukaryotic nuclear RNA polymerases. New York, Cold Spring Harbor Press.

Roeder, R.G. (1991). The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem. Sci.* **16:** 402-408.

Roeder, R.G. and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224:** 234-237.

Roy, A.L., Carruthers, C., Gutjahr, T., and Roeder, R.G. (1993a). Direct role for Myc in transcription initiation mediated by interactions with TFII-I. *Nature* **365:** 359-361.

Roy, A.L., Malik, S., Meisterernst, M., and Roeder, R.G. (1993b). An alternative pathway for transcription initiation involving TFII-I. *Nature* **365:** 355-359.

Roy, A.L., Meisterernst, M., Pognonec, P., and Roeder, R.G. (1991). Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354:** 245-248.

Ruppert, S., Wang, E.H., and Tjian, R. (1993). Cloning and expression of human TAF250: a TBP-associated factor implicated in cell-cycle regulation. *Nature* **362:** 175-179.

Russell, P. (1985). Transcription of the triose-phosphate-isomerase gene of Schizosaccaromyces pombe initiates from a start point different from that in Saccharomyces cerevisiae. *Gene* **40:** 125-130.

Russell, P. and Nurse, P. (1986). *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* : A look at yeasts divided. *Cell* **45:** 7810782.

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

Samuels, M., Fire, A., and Sharp, P.A. (1984). Dinucleotide Priming of Transcription Mediated by RNA Polymerase II. *J. Biol. Chem.* **259:** 2517-.

Samuels, M. and Sharp, P.A. (1986). Purification and characterization of a specific RNA polymerase II transcription factor. *J. Biol. Chem.* **261:** 2003-2013.

Sasse-Dwight, S. and Gralla, J.D. (1988). Probing the *E. coli* glnALG upstream activation mechanism *in vivo*. *Proc. Natl. Acad. Sci. USA* **85:** 8934-8938.

Sassoni-Corsi, P., Corden, J., Kedinger, C., and Chambon, P. (1981). Promotion of specific *in vitro* transcription by excised "TATA" box sequences inserted in a foreign nucleotide environment. *Nucleic Acids Res.* **9:** 3941-3957.

Sawadogo, M. and Roeder, R.G. (1984). Energy requirement for specific transcription initiation by the human RNA polymerase II system. *J. Biol. Chem.* **259:** 5321-5326.

Sawadogo, M. and Roeder, R.G. (1985a). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quatitative *in vitro* assay. *Proc. Natl. Acad. Sci. USA* **82:**

Sawadogo, M. and Roeder, R.G. (1985b). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43:** 165-175.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H.J., Chambon, P., and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260:** 58-63.

Schekman, R., Weiner, A., and Kornberg, A. (1974). Multienzyme systems of DNA replication. *Science* **186:** 987-993.

Schmidt, M.C., Kao, C.C., Pei, R., and Berk, A.J. (1989). Yeast TATA-box transcription factor gene. *Proc. Natl. Acad. Sci. USA* **86:** 7785-7789.

Schultz, M.C., Reeder, R.H., and Hahn, S. (1992). Variants of the TATA-binding protein can distinguish subsets of RNA polymerase I, II, and III promoters. *Cell* **69:** 697-702.

Sekiguchi, T., Nohiro, Y., Nakamura, Y., Hisamoto, N., and Nishimoyo, T. (1991). The human *CCG1* gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein. *Molec. Cell. Biol.* **11:** 3317-3325.

Sekimizu, K., Kobayashi, N., Mizuno, D., and Natori, S. (1976). Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. *Biochemistry* **15:** 5064-5070.

Sekimizu, K., Yokoi, H., and Natori, S. (1982). Evidence that stimulatory factor(s) of RNA polymerase II participates in accurate transcription in a HeLa Cell lysate. *J. Biol. Chem.* **257:** 2719-2721.

Sentenac, A. (1985). Eukaryotic RNA polymerases. *CRC Crit. Rev. Biochem.* **18:** 31-90.

Serizawa, H., Conaway, R.C., and Conaway, J.W. (1993). Multifunctional RNA polymerase II initiation factor ∂ from rat liver. *J. Biol. Chem.* **268:** 17300-17308.

Seto, E., Lewis, B., and Shenk, T. (1993). Interaction between transcription factors Sp1 and YY1. *Nature* **365:** 462-464.

Seto, E., Shi, Y., and Shenk, T. (1991). YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. *Nature* **354:** 241-245.

Seto, E., Usheva, A., Zambetti, G.P., Momand, J., Horikoshi, N., Weinmann, R., Levine, A.J., and Shenk, T. (1992). Wild-type p53 binds to thr TATA-binding protein and represses transcription. *Proc. Natl. Acad. Sci. USA* **89:** 12028-12032.

Siebenlist, U., Simpson, R.B., and Gilbert, W. (1980). *E. coli* RNA polymerase interacts homologously with different promoters. *Cell* **20:** 269-281.

Siegele, D.A., Hu, J.A., Walter, W.A., and Gross, C.A. (1989). Altered promoter recognition by mutant forms of σ70 subunit of *E. coli* RNA polymerase. *J. Mol. Biol.* **206:** 591-603.

Simon, M.C., Fisch, T.M., Benecke, B.J., Nevins, J.R., and Heintz, N. (1988). Definition of multiple, functionally distinct TATA elements, one of which is the target in the hsp70 promoter for E1A regulation. *Cell* **52:** 723-729.

Simpson, R.B. (1979). The molecular topography of RNA polymerase-promoter interaction. *Cell* **18:** 277-285.

Sklar, V.E.F. and Roeder, R.G. (1977). Transcription of specific genes in isolated nuclei by exogenous RNA polymerases. *Cell* **10:** 405-414.

Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription element. *Cell* **57:** 103-113.

Smith, D.B. and Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67:** 31-40.

Sopta, M., Carthew, R.W., and Greenblatt, J. (1985). Isolation of three proteins that bind to mammalian RNA polymerase II. *J. Biol. Chem.* **260:** 10353-10360.

Spindler, S.R. (1979). Deoxyribonucleic acid dependent ribonucleic acid polymerase II specific initiation and elongation factors from calf thymus. *Biochem.* **18:** 4042-4048.

Stargell, L.A. and Gorovsky, M.A. (1994). TATA-binding protein and nuclear differentiation in Tetrahymena thermophila. *Molec. Cell. Biol.* **14:** 723-734.

Starr, D.B. and Hawley, D.K. (1991). TFIID binds in the minor groove of the TATA box. *Cell* **67:** 1231-1240.

Stein, H. and Hausen, P. (1970). A factor from calf thymus stimulating DNA-dependent RNA polymerase isolated from this tissue. *Eur. J. Biochem.* **14:** 270-277.

Stragier, P., Parsot, C., and Bouvier, J. (1985). Two functional domains conserved in major and alternate bacterial sigma factors. *Febs Lett.* **187:** 11-15.

Stringer, K.F., Ingles, C.J., and Greenblatt, J. (1990). Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345:** 783-786.

Strubin, M. and Struhl, K. (1992). Yeast and human TFIID with altered DNA-binding specificity for TATA elements. *Cell* **68:** 721-730.

Struhl, K. (1987). The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. *Cell* **50:** 841-846.

Stucka, R. and Feldman, H. (1990). An element of symmetry in yeast TATA box-binding protein transcription factor IID. *FEBS Lett.* **261:** 223-225.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzym.* **185:**

Sugden, B. and Keller, W. (1973). Mammalian deoxyribonucleic acid-dependent ribonucleic acid polymerases. Purification and properties of an α-aminitin-sensitive ribonucleic acid polymerase and stimulatory factors from HeLa and KB cells. *J. Biol. Chem.* **248:** 3777-3788.

Suzuki, Y., Nogi, Y., Abe, A., and Fukusawa, T. (1988). GAL11 protein, an auxiliary transcription activator for genes encoding galactose-metabolizing enzymes in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **8:** 4991-4999.

Swaffield, J.C., Bromberg, J.F., and Johnston, S.A. (1992). Alterations in a yeast protein resembling HIV-tat Binding

Protein relieve the requirement for an acidic activation domain in GAL4. *Nature* **357:** 698-699.

Sypes, M.A. and Gilmour, D.S. (1994). Protein/DNA crosslinking of a TFIID complex reveals novel interactions downstream of the transcription start. *Nucleic Acids Res.* **22:** 807-814.

Szoke, P.A., Allen, T.L., and deHaseth, P.L. (1987). Promoter recognition by E. coli RNA polymerase: effects of base substitutions in the -10 and -35 regions. *Biochemistry* **26:** 6188-6194.

Taggart, A.K.P., Fisher, T.S., and Pugh, B.F. (1992). The TATA-binding protein and associated factors are components of Pol III Transcription Factor TFIIIB. *Cell* **71:** 1015-1028.

Takada, R., Nakatani, Y., Hoffmann, A., Kokubo, T., Hasegawa, S., Roeder, R.G., and Horikoshi, M. (1992). Identification of human TFIID components and direct interaction between a 250 kDa polypeptide and the TATA box-binding protein (TFIIDτ). *Proc. Natl. Acad. Sci. USA* **89:** 11809-11813.

Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). *brahma*: a regulator of *Drodophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68:** 561-572.

Tamura, T., Sumita, K., Fujino, I., Aoyama, A., Horikoshi, M., Hoffmann, A., Roeder, R.G., Muramatsu, M., and Mikoshiba, K. (1991). Striking homology of the "variable" N-terminus as well as the "conserved core" domains of the mouse and human TATA-factors (TFIID). *Nucleic Asids Res.* **19:** 3861-3865.

Tanese, N., Pugh, B.F., and Tjian, R. (1991). Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes & Development* **5:** 2212-2224.

Thompson, C.M., Koleske, A.J., Chao, D.M., and Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73:** 1367-1375.

Timmers, M. and Sharp, P.A. (1991). The mammalian TFIID protein is present in two functionally distinct complexes. *Genes & Development* **5:** 1946-1956.

Travers, A.A. and Burgess, R.R. (1969). Cyclic re-use of the RNA polymerase sigma factor. *Nature* **222:** 537-540.

Travers, A.A., Ner, S.S., and Churchill, M.E.A. (1994). DNA chaperones: a solution to a persistence problem? *Cell* **77:** 167-169.

Truant, R., Xiao, H., Ingles, C.J., and Greenblatt, J. (1993). Direct interaction between the trancriptional activation domain of human p53 and the TATA box-binding protein. *J. Biol. Chem.* **268:** 2284-2287.

Tschochner, H., Sayre, M.H., Flanagan, P.M., and Feaver, W.J. (1992). Yeast RNA polymerase II initiation Factor e: Isolation and identification as the functional counterpart of transcription factor IIB. *Proc. Natl. Acad. Sci. USA* **89:** 11292-11296.

Tsuda, M. and Suzuki, Y. (1982). Efficient and strand-selective in vitro transcription initiation by purified RNA polymerase II from a unique site of the fibroin gene. *J. Biol. Chem.* **257:** 12367-12372.

Tsujimoto, Y., Hirose, S., Tsuda, M., and Suzuki, Y. (1981). Promoter sequence of fibroin gene assigned by *in vitro* transcription system. *Proc. Natl. Acad. Sci. USA* **78:** 4838-4842.

Tsukiyama, T., Becker, P.B., and Wu, C. (1994). ATP-dependent nucleosome disruption ata heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367:** 525-532.

Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., and Aloni, Y. (1992). Spesific interactions between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell* **69:** 871-881.

Usheva, A. and Shenk, T. (1994a). *Nature* **:** in press.

Usheva, A. and Shenk, T. (1994b). TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* **76:** 1115-1121.

van Dyke, M.W., Roeder, R.G., and Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* **241:** 1335-1338.

Vanfleteren, J.R., Van Bun, S.M., Delcambe, L.L., and Van Beeumen, J.J. (1986). Multiple forms of histone H2B from the nematode Caenorhabditis elegans. *Biochem. J.* **235:** 769-773.

Vogt, P.K., Bos, T.J., and Doolittle, R.F. (1987). Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein coded for by the oncogene *jun*. *Proc. Natl. Acad. Sci. U.S.A.* **84:** 3316-3319.

Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltmore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., and Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358:** 646-653.

Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D., and Kuriyan, J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexes and peptide-free forms. *Cell* **72:** 779-790.

Wang, E.H. and Tjian, R. (1994). Promoter-selective transcriptional defect in cell cycle mutant ts13 rescued by hTAFII250. *Science* **263:** 811-814.

Wasylyk, B. and Chambon, P. (1981). A T to A base substitution and small deletions in the conalbumin TATA box drastically decrease specific *in vitro* transcription. *Nucleic Acids Res.* **9:** 1813-1823.

Webster, N., Jin, J.R., Green, S., Hollis, M., and Chambon, P. (1988). The yeast UASG is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. *Cell* **52:** 169-178.

Weil, P.A., Luse, D.S., Segall, J., and Roeder, R.G. (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.

Weinzierl, R., Ruppert, S., Dynlacht, B., Tanese, N., and Tjian, R. (1993a). Cloning and expression of Drosophial

TAFII60 and human TAFII70 reveal conserved interactions with other subunits of TFIID. *EMBO J.* **12:** 5303-5309.

Weinzierl, R.O.J., Dynlacht, B.D., and Tjian, R. (1993b). Largest subunit of the Drosophila transcription factor IID directs assembly of a complex containing TBP and a coactivator. *Nature* **362:** 511-517.

White, R.J., Jackson, S.P., and Rigby, P.W.J. (1992). A role for the TATA-box-binding protein component of the transcription factor IID complex as a general RNA polymerase III transcription factor. *Proc. Natl. Acad. Sci. USA* **89:** 1949-1953.

Wilson, I.A., Niman, H.L., Houghton, R.A., Cherenson, A.R., Connolly, M.L., and Lerner, R.A. (1984). The structure of an antigenic determinant in a protein. *Cell* **37:** 767-778.

Winston, F. and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends in Genetics* **8:** 387-391.

Winston, F. and Minehart, P.L. (1986). Analysis of the yeast SPT3 gene and identification of its product, a positive regulator of Ty transcription. *Nucleic Acids Res.* **14:** 6885-6900.

Wong, J., Liu, F., and Bateman, E. (1992). Cloning and expression of the Acanthamoeba castellanii gene encoding transcription factor TFIID. *Gene* **117:** 91-97.

Workman, J.L., Abmayr, S.M., Cromlish, W.A., and Roeder, R.G. (1988). Transcriptional regulation by the immediate early protein of pseudorabies virus during *in vitro* nucleosome assembly. *Cell* **55:** 211-219.

Workman, J.L. and Roeder, R.G. (1987). Binding of transcription factor TFIID to the major late promoter during In vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* **51:** 613-622.

Workman, J.L., Roeder, R.G., and Kingston, R.E. (1990). An upstream transcription factor, USF (MLTF), facilitates the formation of preinitiation complexes during *in vitro* chromatin assembly. *EMBO J.* **9:** 1299-1308.

Wu, L., Rosser, D.S.E., Schmidt, M.C., and Berk, A. (1987). A TATA box implicated in E1A transcriptional activation of a simple adenovirus 2 promoter. *Nature* **326:** 512-515.

Xu, X., Prorock, C., Ishikawa, H., Maldonado, E., Ito, Y., and Gelinas, C. (1993). Functional interaction of the v-Rel and c-Rel oncoproteins with the TATA-binding protein and association with TFIIB. *Molec. Cell. Biol.* **13:** 6733-6741.

Yamamoto, T., Horikoshi, M., Wang, J., Hasegawa, S., Weil, P.A., and Roeder, R.G. (1992). A bipartite DNA bindining domain composed of direct repeats in the TATA box binding factor TFIID. *Proc. Natl. Acad. Sci. USA* **89:** 2844-2848.

Yokomori, K., Admon, A., Goodrich, J.A., Chen, J.-L., and Tjian, R. (1993a). *Drosophila* TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. *Genes & Development* **7:** 2235-2245.

Yokomori, K., Chen, J.-L., Admon, A., Zhou, S., and Tjian, R. (1993b). Molecular cloning and characterization of dTAFII30 α and dTAFII30 β: two small subunits of *Drosophia* TFIID. *Genes & Development* **7:** 2587-2597.

Yoshinaga, S.K., Peterson, C.L., Herskowitz, I., and Yamamoto, K.R. (1992). Roles of SWI1, SWI2 and SWI3 proteins

for transcriptional enhancement by steroid receptors. *Science* **258:** 1598-1604.

Youderian, P., Bouvier, S., and Susskind, M.M. (1982). Sequence determinants of promoter activity. *Cell* **30:** 843-853.

Zarucki-Sculz, T., Tsai, S.Y., Itakura, K., Soberon, X., Wallace, R.B., Tsai, M.-J., Woo, S.L.C., and O'Malley, B.W. (1982). Point mutagenesis of the ovalbumin gene promoter sequence and its effect on *in vitro* transcription. *J. Biol. Chem.* **257:** 11070.

Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, A.I., and Green, M.R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl. Acad. Sci. USA* **89:** 8769-8773.

Zheng, X.-M., Moncollin, V., Egly, J.-M., and Chambon, P. (1987). A general transcription factor forms a stable complex with RNA polymerase B (II). *Cell* **50:** 361-368.

Zhou, Q., Lieberman, P.M., Boyer, T.G., and Berk, A.J. (1992). Holo TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. *Genes & Development* **6:** 1964-1974.

Zhou, Q., Schmidt, M.C., and Berk, A.J. (1991). Requirements for acidic amino acid residues immediately N-terminal to the conserved domain of *Saccharomyces cerevisiae* TFIID. *EMBO J.* **10:** 1843-1852.

Ziff, E.B. and Evans, R.M. (1978). Coincidence of the Promoter and Capped 5' Terminus of RNA from the Adenovirus 2 Major Late Transcription Unit. *Cell* **15:** 1463-1475.

PRO·BONO·HUMANI·GENERIS
The Rockefeller University
1901

THE LIBRARY

19010000065515

End