Transcription regulation of human genes: novel aspects and mechanisms

Een wetenschappelijke proeve op het gebied van de Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen op gezag van de rector magnificus prof. dr. S.C.J.J. Kortmann, volgens besluit van het College van Decanen in het openbaar te verdedigen op woensdag 5 maart 2008 om 10.30 uur precies door

Sergey Denissov

geboren op 11 januari 1975 te Korsakov, Rusland
The research described at this thesis was carried out at the Department of Molecular Biology, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands. Financial support was obtained from the National Scientific Organization, National Cancer Foundation and HEROIC, an Integrated Project funded by the European Union.


Copyright © 2008 S. Denissov
All rights reserved.

Printed by PrintPartners Ipskamp
Transcription regulation of human genes: novel aspects and mechanisms

An academic essay in
Science, Mathematics and Informatics

Doctoral thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus, prof. dr. S.C.J.J. Kortmann
according to the decision of the Council of Deans
to be defended in public on Wednesday 5 March 2008
at 10.30 hours
by

Sergey Denissov

born in Korsakov, Russia
on 11 January 1975
Supervisor:  
Prof. dr. H.G. Stunnenberg

Manuscript committee:  
Prof. dr. A. Geurts van Kessel  
Prof. dr. H.T. Timmers (Utrecht University)  
Prof. dr. F.C. Holstege (Utrecht University)

The research described at this thesis was carried out at the Department of Molecular Biology, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands. Financial support was obtained from the National Scientific Organization, National Cancer Foundation and HEROIC, an Integrated Project funded by the European Union.


Copyright © 2008 S. Denissov  
All rights reserved.

Printed by PrintPartners Ipskamp
## Table of contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General introduction</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Identification of novel functional TBP-binding sites and general factor repertoires</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>ChIP-profiling implies differential TFIIA requirement for transcription of human genes</td>
<td>121</td>
</tr>
<tr>
<td>4</td>
<td>The binding profiles of RNAP I transcription factors suggest a unique three-dimensional structure of the active rDNA gene</td>
<td>137</td>
</tr>
<tr>
<td>5</td>
<td>E2F transcriptional repressor complexes are critical downstream targets of p19ARF/p53-induced proliferative arrest</td>
<td>157</td>
</tr>
<tr>
<td>6</td>
<td>General discussion</td>
<td>177</td>
</tr>
<tr>
<td>Addendum</td>
<td>Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>207</td>
</tr>
<tr>
<td></td>
<td>Samenvatting</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>Acknowledgements</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>Curriculum Vitae and publications</td>
<td>216</td>
</tr>
</tbody>
</table>
Chapter 1

General introduction
Table of contents

Prologue
1. Transcription regulation of RNA polymerase II genes 12
   1.1 General transcription factors and complexes 12
      1.1.1 TFIID 12
         TBP 12
         Structure of TFIID 13
         TAFs 13
         Variants of TAFs and TFIID 14
         Table 1. Subunits of TFIID and their properties 15
      1.1.2 TFIIA 17
      1.1.3 TFIIB 17
      1.1.4 TFIIF 18
      1.1.5 TFIIE 19
      1.1.6 TFIIH 20
      1.1.7 RNA polymerase II 21
         Structure of RNAP II 21
         CTD structure, modifications and functions 22
      1.1.8 Mediator 24
      1.1.9 TBP-containing complexes implicated in RNAP II transcription 25
      1.1.10 TBP-related factors 25
      1.1.11 TBP-free TAF-containing complexes 26
         Table 2. Characteristics of the TBP-free TAF containing complexes 27
   1.2 Core promoter of the RNAP II genes 30
      1.2.1 Structure of the core promoter 30
      1.2.2 Recognition of the core promoter by general transcription factors 30
      1.2.3 Other core-promoter elements 31
      1.2.4 Mechanisms of gene regulation via core promoter 32
   1.3 Mechanisms of the RNAP II transcription 33
      1.3.1 Initiation 33
         Preinitiation complex 33
         Open complex formation 34
      1.3.2 Elongation 35
         Promoter escape 35
         Promoter-proximal pausing 36
         Productive elongation 37
      1.3.3 Termination 37
   1.4 Transcription regulation by activators 39
      1.4.1 Gene-specific activators 39
      1.4.2 Regulation of activator function 39
      1.4.3 Transcription activation by enhancers 41
   1.5 Transcription coactivators function 41
      1.5.1 General transcription factors 41
      1.5.2 Chromatin remodeling 41
      1.5.3 Covalent histone modifications 42
## Chapter 1

1.6 Co-transcriptional regulation of different cellular processes

<table>
<thead>
<tr>
<th>Sub-section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6.1 Capping of the 5’ end of mRNA</td>
<td>45</td>
</tr>
<tr>
<td>1.6.2 Splicing</td>
<td>45</td>
</tr>
<tr>
<td>1.6.3 Processing of the 3’ end of mRNA</td>
<td>47</td>
</tr>
<tr>
<td>1.6.4 mRNA nuclear transport</td>
<td>48</td>
</tr>
<tr>
<td>1.6.5 DNA repair</td>
<td>49</td>
</tr>
<tr>
<td>1.6.6 Transcription factories</td>
<td>50</td>
</tr>
</tbody>
</table>

2. Transcription of a ribosomal RNA gene by RNA polymerase I

<table>
<thead>
<tr>
<th>Sub-section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Nucleolar sub-compartments and their functions</td>
<td>50</td>
</tr>
<tr>
<td>2.2 Transcription factors of RNAP I</td>
<td>52</td>
</tr>
<tr>
<td>2.3 Dynamics of rRNA transcription</td>
<td>54</td>
</tr>
<tr>
<td>2.4 Regulatory mechanisms and pathways of RNAP I transcription</td>
<td>55</td>
</tr>
<tr>
<td>2.5 Spatial conformation of the transcribed rDNA gene</td>
<td>56</td>
</tr>
</tbody>
</table>

3. The RNA polymerase III transcription

<table>
<thead>
<tr>
<th>Sub-section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Promoter types of the RNAP III transcribed genes</td>
<td>57</td>
</tr>
<tr>
<td>3.2 The RNAP III transcription factors</td>
<td>58</td>
</tr>
</tbody>
</table>

4. Methodology for genome research

<table>
<thead>
<tr>
<th>Sub-section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Chromatin immunoprecipitation</td>
<td>59</td>
</tr>
<tr>
<td>4.2 ChIP-cloning</td>
<td>60</td>
</tr>
<tr>
<td>4.3 ChIP-on-chip</td>
<td>60</td>
</tr>
<tr>
<td>4.4 ChIP-PET</td>
<td>61</td>
</tr>
<tr>
<td>4.5 ChIP-sequencing</td>
<td>61</td>
</tr>
<tr>
<td>4.5 CAGE</td>
<td>62</td>
</tr>
<tr>
<td>4.6 Transcription regulatory networks</td>
<td>62</td>
</tr>
</tbody>
</table>

References

62
General introduction

Prologue
The miracle of developing of an entire organism from a single cell attracts the minds from the earliest times. Although it remains to be largely unexplained, significant insight into molecular processes of life has been achieved in the last few decades. Critical cellular processes such as growth, differentiation and even death are regulated on the level of gene expression by realization of the information encoded in DNA of the genome. Gene expression is tightly regulated and controlled at every step for every gene providing an organized RNA and protein output in each cell or tissue; it is obtained by an orchestration between numerous processes such as synthesis of RNA and protein (transcription and translation), subsequent processing and chemical modifications (post-transcriptional and post-translational events), degradation and intra- and extra-cellular transport. Gene expression can be regulated at any particular moment depending on environmental signals or developmental stage.

The synthesis of RNA from the DNA template, named transcription, is the primary process in gene expression and can be split to several steps: initiation, elongation and termination. The control of the initiation event as a very first step in transcription is a very important and complex process that determines gene expression rate. It is crucial in responding to environmental changes and to execute a precise gene expression program during growth, differentiation and development. Transcription initiation and its regulation involves hundreds of proteins, many enzymatic activities and a multitude of dynamic protein/DNA/RNA interactions. Several levels of regulation can be determined, such as a coordinated assembly of a preinitiation complex at a promoter DNA region. It includes many multi-subunit protein complexes such as RNA polymerase, general transcription factors and regulators and their interplay with various regulatory DNA sites. Another regulatory level is provided from the preparation of a suitable chromatin environment that includes ATP-dependent nucleosome remodeling to change compaction and prepare promoter DNA for assembling of transcriptional machinery, and covalent histone modifications which facilitate recruitment of specific proteins to activate or repress transcription.

There are three distinct classes of genes different in their general properties and transcribed by different RNA polymerases (RNAP). The RNAP I transcribes a precursor of ribosomal RNA which is then processed to smaller species. The RNAP II transcribes genes that encode proteins or are processed to small RNA species with abundant regulatory functions. The RNAP III transcribes small structural RNA genes which are involved in many crucial processes such as splicing of RNAP II transcripts, protein transport and transcription regulation.

The sequencing of entire genome of many different organisms has opened up the way for the new genome research topics in molecular biology addressing key issues such as comprehensive identification and characterization of genes and their regulatory elements. Biochemical techniques in combination with bioinformatical tools uncover novel molecular mechanisms and underline the very limiting current knowledge in understanding of cellular processes and extreme complexity of living organisms.
1. Transcription regulation of RNAP II transcribed genes.

1.1 General transcription factors and complexes

RNA polymerase II transcribes a large number of genes. Among them are all the genes encoding proteins and a number of non-protein coding RNA genes involved in regulation of various cellular processes. RNAP II-transcribed genes undergo tight regulation which occurs largely at the step of initiation and involves coordinated interplay of many transcription factors, various DNA elements and modification of chromatin structure.

Recognition of the promoter region is largely determined by the specific affinity of general transcription factor TFIID to the certain core-promoter elements (chapter 1.2). A number of other general transcription factors (GTFs) together comprising the preinitiation complex (PIC) are found interacting with each other and RNA polymerase and were named TFIIA, TFIIB, TFIIE, TFIIF and TFIIH [1 - 3]. Most of them are stable multiprotein complexes which are highly conserved between species and were found to be essential for transcription initiation by RNAP II in reconstituted systems in vitro [1]. They are also essential for transcription of the most of the cellular genes in vivo.

1.1.1 TFIID

A key step in transcription initiation is the formation of preinitiation complex on promoters and proper positioning of RNA polymerase at the correct site. Eukaryotic RNA polymerase II itself is unable to recognize promoters and accurately position at the transcription start or to respond to a large variety of regulatory signals. These functions are performed by a large number of interacting proteins, called general transcription factors, as well as the mediator complex that together comprise PIC and have a crucial role in regulation of RNAP II functioning and transcription initiation/activation [1].

TFIID is a central complex within PIC consisting of TBP (TATA-binding protein) and 14 different TBP-associated factors (TAFs) [4]. TFIID possesses many different properties and functions which are important for transcription initiation (Table 1): enzymatic activities such as histone acetyltransferase, sequence-specific DNA-binding (binds to core promoter elements, chapter 1.2), interaction with acetylated histones via bromodomain of TAF1, interactions with a large number of transcription factors and RNA polymerase that is important for PIC assembly and for conducting upstream regulatory signals from activator proteins [7]. In addition, these interactions are required for the recruitment of TFIID to promoters via interaction with activators (coactivator function) [3, 6] that facilitates PIC assembly [5]. Some TAFs play a crucial role in transcriptional activation via protein-protein interactions, whereas other TAFs direct promoter selectivity via protein-DNA recognition. The individual TFIID subunits, their properties and functions are summarized in Table 1.

TBP

TBP is considered as the core of TFIID with a specific affinity to the TATA-like DNA sequences [114]. The highly conserved 180 amino acids C-terminal region of TBP forms a unique saddle-like fold around DNA and directly interacts via minor groove contacts [9, 10, 560]. The non-conserved N-terminal region of human TBP is dispensable for TFIID complex assembly and is not seem to be directly involved in an activator-dependent transcription by RNAP II [8] but is necessary for the recruitment of small nuclear RNA-activating protein complex (SNAPc) to the promoter of RNAP III transcribed U6 snRNA gene (chapter 3) [559]. TBP interacts with a variety of factors that are involved in both general and specific regulation of gene transcription. TBP is also a part of the SL1 and TFIIIB complexes (chapters 2.2 and 3.2) and therefore is regarded as a universal factor essential
for transcription initiation by RNAP I, -II and -III [2, 3] and for all the cellular genes with may be a few exceptions (chapter 1.1.10).

Structure of TFIID

The TFIID complex forms a molecular clamp that is composed of three to four domains joined by narrow bridges and organized into a horseshoe-like structure around a central cavity [83, 84, 114, 115]. TBP was identified in the center of TFIID facing the central cavity suggesting that this is a major site for DNA binding [83, 84, 114]. The internal interactions within the TFIID complex were extensively studied. Structural analyses have revealed that more than a half of the TAFs contain a histone fold domain (HFD). These TAFs specifically assemble into five histone-like pairs [84, 85] determining the specific TFIID structure and facilitating DNA binding [106]. A nucleosome-like octamer could be formed by yeast TAF6–TAF9–TAF4–TAF12 in vitro [136]. TAF5 is considered as a major scaffold protein that connects different structural domains of TFIID [84].

The histone fold is a fundamental interaction motif involved in heterodimerization of the core histones, H4 - H3, and H2A - H2B, and their assembly into a nucleosome octamer [77, 78]. The HFD is found not only in core histones and TAFs, but also in many other transcription factors [79]. In the histone fold, two proteins are arranged head-to-tail to allow extensive interactions between these regions as well as between the α2 helices of the two proteins leading to the formation of a tight and compact structure. This can function in determining a strict partner-interaction specificity [80, 81].

TAFs

TAFs and TAF-paralogues have specialized roles within TFIID such as maintaining complex stability, specific interactions with DNA or proteins, co-function with various transcription factors and activators. TAFs reveal significant structural and functional conservation between yeast and mammals [117]. In addition to the role in transcription initiation, much data is accumulating about specialized TAF functions in a large number of distinct cellular processes (Table 1). TAF1 is the largest subunit of TFIID (250 kDa) that possesses a multitude of functions. It has several domains interacting with other proteins (Table 2). It has two bromodomains comprised of left-handed four-helix bundles [70]. These structural modules are often present in coactivators and chromatin related factors, histone acetyltransferases (HATs) and their associated proteins. These modules bind to acetylated histone H4 that is implicated in promoter recognition. TAF1 also possesses a number of enzymatic activities: it is a bipartite protein serine kinase (consisting of N- and C-terminal kinase domains) that is capable of autophosphorylation and selective phosphorylation of histone H2B at serine 33 [74], RAP74 subunit of TFIIF [73], TFIIA [103] and the tumor suppressor protein p53 [107]. The HAT domain of TAF1 acetylates histones H3 and H4 [64] that facilitates transcription (chapter 1.5.3). It has also an important role in the gene-specific regulation: conditional inactivation of the HAT domain in yeast results in decrease of transcription of several cell cycle-related genes [72], a cell-cycle block at the G1/S transition and apoptosis [76] although it has no global effect on transcription [105]. Collectively, this suggests a specific role for the HAT activity in transcription of a subset of genes. Recently ubiquitin-activating and -conjugating activity of the TAF1 was found to be directed towards histone H1 and required for transcriptional activation in Drosophila embryos [71].

The role of various TAFs in transcription was investigated in living cells via inactivation. In yeast, individual TAFs and their combinations were inactivated using temperature-sensitive mutations, conditional depletion and targeted protein degradation [86 - 95]. The effect of inactivation was followed by measuring cellular RNA content using microarrays. The interesting result of these studies
was that TAFs were found not to be universally required for transcription. Each TAF is required for the expression of a characteristic subset of genes, ranging from 3% to 67% of the genomic genes. The broadest requirement was found for the histone-like TAF9 in yeast. The entire picture revealed that each of the TAFs and specific TAF combinations appear to be required for transcription of a particular subsets of genes.

Similar studies in mammalian cells are limited and suggest that TAFs have functions analogous to those found in yeast. Cell lines harboring a temperature-sensitive TAF1 allele do not have a global defect in RNA polymerase II transcription under inactivating conditions [96, 97]. Mouse embryonic stem cells (ES cells) deficient in TAF10 reveal changes in expression of only a subset of genes [98]. The distinct functions and selective transcriptional roles of TAFs were explained by a requirement for TBP recruitment and assembly of the preinitiation complex [101].

Gene-specific functions are supported by findings that individual TAFs mediate cell-cycle progression and inactivation of some TAFs has negative effect on cell-cycle. In yeast, TAF1 is required for transcription of genes involved in cell-cycle progression [89, 93] and inactivation of TAF1 results in G1/S arrest. Similar results were obtained observed in mammalian cells [97, 102]. Mouse cells lacking TAF10 display impaired expression of cyclin E [98] and G1/S arrest. Inactivation of yeast TAF2 and -5 [87, 88] and human TAF5 [99] results in G2/M arrest. Remarkably, TAFs themselves are differentially expressed during cell-cycle [89] and specific phosphorylation of TAFs and TBP prior mitosis leads to inactivation of RNAP II transcription [100]. A broad spectrum of diverse functions of TAFs becomes apparent and is yet incomplete (Table 1).

**Variants of TAFs and TFIID**

Previously, it was considered that the preinitiation complex is invariant at each and every promoter and composed of the same set of GTFs while TFIID contains the same set of TAFs. Many findings of proteins homologous to TBP and TAF can compose alternative TFIID forms and suggest that there are many different preinitiation complexes which can contribute to gene-specific transcription regulatory mechanisms.

The TAF variants were found in different organisms (mostly studied in yeast, Drosophila, mouse and human) and are mostly expressed from separate genes but can also result from alternative mRNA processing. They reveal a cell- or tissue-specific expression patterns and have specialized functions in regulation of different cellular processes. They can be part of TFIID and change the properties of the complex, and, therefore, direct its activity/affinity towards promoters of a specific subset of genes. In addition, TAF variants are involved in TBP-free complexes such as TFTC, STAGA, or PCAF/GCN5 (chapter 1.1.11).

A number of TAF variants and their properties are described in Table 1. Amongst the most studied is TAF4b [501 - 505]. It is expressed at high level specifically in ovary, testis and lymphoid B cells and is part of a TAF4b-containing TFIID that activates a subset of genes. Similarly, TAF9b and TAF7L also can replace their conventional counterparts within TFIID and in this way regulate transcription of specific genes [482, 520]. In other case, external apoptotic stimuli can lead to alternative splicing of TAF6 resulting in a 10 amino acids truncated form named TAF6 delta. It was found as a part of a TFIID-like complex that lacks TAF9 and selectively alters gene expression and promotes apoptosis [506]. Additionally, the variants of TAF1, TAF5 and TAF10 were found in different organisms [536 - 539]. Indeed, different TFIID complexes with or without TAF10 were found present simultaneously in cells and were shown to possess distinct functional properties [534]. In addition, the core TAFs were found expressed at very different levels in distinct cell types or tissues [529, 535]. Collectively,
these data suggest that the composition of TFIID in different cells can vary and that the different TFIIDs can modulate transcription initiation in response to specific activators resulting in expression of specific genes [533].

Table 1. Subunits of TFIID and their properties.

<table>
<thead>
<tr>
<th>Subunits(^1)</th>
<th>Properties and functions of the subunits</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBP (38 kDa*)</td>
<td>Binds to TATA-box DNA sequence. Interacts with TAFs, TFIIA, TFIIIB and other transcription factors. Part of distinct complexes: TFIIIB [52], SL1 [53], TAC [51], B-TFIID [50].</td>
</tr>
<tr>
<td>TAF1 (250 kDa)</td>
<td>Interacts with many transcription factors and with chromatin. Required for cell cycle progression [89, 93, 97, 102]. Functions as a scaffold for TFIID [67 - 69]. Contains a number of functional domains: TBP-interacting domain [63], TAF-TAF interaction domain [62], histone acetyltransferase (HAT) domain [64], promoter recognition domain [66], two bromodomains that can bind to acetylated histones H3 and H4 [70, 557], ubiquitin-activating and -conjugating activity domains [71]. The kinase domain can phosphorylate histone H2B at ser33 [74], RAP74 subunit of TFIIF [73], TFIIA [103] and p53 [107].</td>
</tr>
<tr>
<td>TAF2 (150 kDa)</td>
<td>Binds to core-promoter (Inr element) [22, 23]. Interacts with TAF1 and TBP [75]. Required for cell cycle progression [87].</td>
</tr>
<tr>
<td>TAF3 (140 kDa)</td>
<td>Contains histone fold domain [108]. Highly expressed in testis [482].</td>
</tr>
<tr>
<td>TAF4 (130 kDa)</td>
<td>Contains histone fold domain [84, 85]. Critical role in stability of TFIID complex [494 - 496]. Coactivator for the transcription factors Sp1, CREB (cAMP signalling pathway), receptors for retinoic acid (RAR), vitamin D3 (VDR) and thyroid hormone (TR) [80, 484 - 487, 489]. Negative regulator of transcription factor ATF7 (stress induced JNK2 kinase pathway and cAMP signalling pathway) [490, 491]. Negatively regulates the TGF-beta pathway by competition with TAF4b [483]. Loss of TAF4 leads to deregulation of many genes, high TGF-beta expression and induces proliferation [492, 493].</td>
</tr>
<tr>
<td>TAF4b (105 kDa)</td>
<td>Can replace TAF4 within TFIID providing different functions [483, 500]. Strongly expressed and has specific functions in ovary, testis and lymphoid B cells, whereas weakly expressed in many other tissues [501, 502, 505]. Coactivator for the components of the TGF-beta pathway, antagonist of TAF4 [483, 492, 504].</td>
</tr>
<tr>
<td>TAF5 (100 kDa)</td>
<td>Critical role in stability of TFIID complex via stable interaction with TAF6 and TAF9 [494 - 496]. SUMO-modification at Lys-14 inhibits TFIID promoter binding [104]. Important for transcription of a large number of genes [497]. Interacts with TFIIF [498]. Required for cell cycle progression [88, 99].</td>
</tr>
<tr>
<td>TAF6 (80 kDa)</td>
<td>Interacts with DPE promoter element [22, 23]. Contains histone fold domain [84, 85]. Essential for transcription of many genes in different tissues and during development in</td>
</tr>
</tbody>
</table>
TAF6 delta (79 kDa) | Expressed in response to apoptotic stimuli; part of TFIID-like complex lacking TAF9; induces expression of specific genes and apoptosis [506].

TAF7 (55 kDa) | Interacts with several TAFs [507], nuclear receptors for vitamin D(3) and thyroid hormone [508]. Inhibits HAT activity of TAF1 [509]. Binds/modulates transcription activator c-Jun [510].

TAF7L (52 kDa) | Can replace TAF7 within TFIID. Expressed in testis and essential for differentiation of male germ-cell [482].

TAF8 (34 kDa) | Contains histone fold domain [84, 85]. Positive regulator of adipogenic differentiation [556].

TAF9 (31 kDa) | Contains histone fold domain [84, 85]. Essential for structural integrity if TFIID [515]. Interacts with DPE [22, 23]. Essential for transcription of most genes [101, 514, 516]. Required during embryogenesis [519]. Interacts with TFIIB and many acidic activators (co-activator function) [512, 513, 517], interacts and stabilizes tumor suppressor protein p53 [518].

TAF9b (28 kDa) | Can replace TAF9 within TFIID and is involved in regulation of a subset of genes [520]. Functions in transcriptional repression and gene silencing [515].

TAF10 (30 kDa) | Contains histone fold domain [84, 85]. Required for cell cycle progression [98]. Interacts with TAF3 and TAF8 via HFD [531]. Highly expressed in testis [482]. Required for TFIID stability [529]. Essential during early embryogenesis [529]. Has a specific function in keratinocytes [532]. Regulated via methylation by SET9 [530].

TAF11 (28 kDa) | Contains histone fold domain [109]. Interacts with TFIIA [56, 61]. Provides critical functional contacts with TBP and TAF13 in the preinitiation complex [101, 516]. Essential for transcription of the majority of genes. Coactivator for nuclear receptors for vitamin D3 and thyroid hormone [57].

TAF12 (20 kDa) | Contains histone fold domain [84, 85]. Required also for transcription of rRNA genes by RNAP I [561]. Involved in regulation of chromatin remodeling [527]. Coactivator for ATF7 in cAMP signalling pathway [491].

TAF13 (18 kDa) | Contains histone fold domain [109]. Provides critical functional contacts with TBP and TAF11 in the preinitiation complex [101, 109].

TAF15 (68 kDa) | Contains distinct domains that bind to RNA and DNA, and also RNAP II [523]. Contains potent transcription activator domain [524, 526]. Has impact on oncogenic transformation [524, 525].

---

1 TAFs names correspond to the unified nomenclature [54].

*The molecular weight is indicated for human proteins.
1.1.2 TFIIA

Transcription factor TFIIA in mammalian cells consists of three subunits: TFIIA\(\alpha\), TFIIA\(\beta\) and TFIIA\(\gamma\) whereas TFIIA in yeast has only 2 subunits TOA1 and TOA2 [250]. The \(\alpha\) and \(\beta\) subunits are post-translationally processed from a single polypeptide via proteolytic cleavage [116] which has an effect on stability of the complex [257].

TFIIA has been initially characterized as a general factor required for basal transcription in vitro [118 - 120]. Additional studies revealed that TFIIA is not essential for accurate initiation of transcription when highly purified fractions are used in reconstituted experiments [460, 461]. Nevertheless, TFIIA is required for basal transcription when partially purified cell extract fractions are added [3, 118 - 120, 242]. Therefore, it was suggested that TFIIA can function as antirepressor via antagonizing transcriptional repressors present in the cell extracts, possibly via displacing them from promoter and blocking their function.

Detailed investigation of TFIIA function in PIC assembly on promoters has been addressed by approaching a crystal structure of the TFIIA/TBP/DNA complex [121, 122]. It was found that TFIIA contacts DNA region immediately upstream of the TATA-box; such a structure is appeared to be remarkably conserved between yeast and human. These structural data suggest that TFIIA can contribute to stability of TFIID at promoter and, therefore, positively affect formation of PIC [118 - 122].

TFIIA likely plays a general role in transcription initiation and PIC assembly via direct interactions with different general factors such as TFIID (subunits TBP, TAF1 and TAF11), TFIIE and TFIIF [56, 61, 103, 166, 250 - 252]. These interactions may stabilize TFIID-DNA complex and, therefore, facilitate PIC formation and stimulate both TFIIE and TFIIF [166]. TFIIA-TBP interactions are also important for transcription in vivo [119, 123, 124] because TBP mutants which do not interact with TFIIA are unable to activate transcription [135]. A specific complex containing TBP and uncleaved form of TFIIA\(\alpha\beta\) is also capable to facilitate transcription [51]. In addition, interactions with TBP-related factors TRF1, TRF2 and TRF4 in vivo underline a general role of TFIIA in transcription regulation of broad diversity of genes [253 - 256, 462].

TFIIA also can function as a coactivator via direct interaction with several activators such as AP-1 and SP-1, and also a number of cofactors and other regulatory proteins [63, 113, 125, 127 - 134, 463 - 466]. The coactivator function is supported by studies in yeast where disruption of interactions between TBP with TFIIA results in impairment of transcription in response to activators only for a subset of genes [135]. Depletion of TFIIA in vivo also did not affected significantly an overall level of cellular transcription [119, 129] while specific genes might be deregulated because the depleted cells are specifically blocked in G2/M cell cycle phase [129].

In addition to TFIIA, human cells contain TFIIA\(\alpha\beta\)-like factor (ALF) that has been found specifically in testis [457, 459]. ALF can stabilize TBP binding to the promoter in complex with TFIIA\(\gamma\) subunit and possibly can regulate some specific genes [458].

1.1.3 TFIIB

TFIIB is a monomeric protein of 38 kDa. It has significant sequence homology among mammals, Drosophila, yeast and Archea [1, 3, 475 - 477]. Numerous structural and mutational studies revealed that TFIIB contains several domains which conduct distinct functions with important implications for the regulation of PIC assembly [3, 145, 146, 163, 164]. The C-terminal core domain, comprising about 2/3 of the protein sequence, interacts with TBP and specific DNA elements (BRE, chapter 1.2.1) upstream and also downstream of the TATA box [10 - 12, 137, 140, 141]. These interactions
Chapter 1

stabilize the binding of the TFIID complex at promoter region [20, 478] and contribute to transcription start site selection [1, 3, 145].

The N-terminal domain contains a zinc binding motif and is required for intermolecular TFIIB structure [147 - 149, 279]. This domain can form extensive and precise interactions with subunits Rpb1, -2, and -9 of RNA polymerase II [143 - 145]. In addition, this domain interacts with the RAP30 subunit of TFIIF [480]. These interactions facilitate TFIID stability at promoter [155] and are important for proper setting of the DNA direction in the preinitiation complex and also for positioning of RNAP II. This also suggests that TFIIB acts as a physical bridge between TFIID and RNAP II/TFIIF within PIC [152]. Similar interactions may also take place between conserved N-terminal domains of TFIIB-related factors and RNAP III or archaeal RNA polymerase [146].

The two other domains of TFIIB are the B-finger and linker [147]. The B-finger domain contains a highly conserved sequence block and in the RNAP II-TFIIB complex is located within an active center of elongating RNA polymerase II that coincides with the position of the 8-basepair DNA-RNA hybrid [147, 150, 151]. This might be important for start site selection, promoter escape and proper conformation of the active site of RNAP II during transcription. A post-initiation inhibitory role of this domain was suggested based on competition between TFIIB and RNA in the active center of RNAP II that can cause premature failure of transcription elongation [147, 258]. The B-finger domain is also essential for the modulation of conformational changes of TFIIB upon interaction with DNA or activators. Mutations in the B-finger, however, do not affect PIC assembly and interactions with other general factors such as TBP, TFIIF, RNAP II, or DNA [18, 155, 157, 161]. The linker domain was suggested to have a connective role between TFIIB domains that facilitates the proper conformation of TFIIB in a complex with RNAP II [147].

The dynamic conformation of TFIIB and intramolecular interactions between N- and C-terminal domains appear to be important regulatory factors in transcription initiation [153, 155 - 157]. TFIIB undergoes conformational changes upon integration into a complex with TBP and promoter DNA [160, 161] as well as upon binding to promoters with different architecture that modulates promoter-specific PIC response to activator proteins [18, 161].

TFIIB has been shown to interact with various transcriptional activators. These interactions can facilitate recruitment of TFIIB to the promoter [138] and also affect the intramolecular structure of TFIIB. Consequently, this activates PIC formation via enhancing interactions with RNA polymerase and other GTFs [154, 156, 157, 161]. The viral transcription activator VP16 from herpes simplex virus 1 (HSV1), for example, has been shown to change the intramolecular conformation of TFIIB and facilitate its DNA recognition and recruitment at certain types of promoters [20, 154, 162].

An enzymatic activity of TFIIB has recently been discovered underlining the general function of different covalent modifications in transcription. TFIIB possesses autoacetylation activity that has implications for transcription activation [165].

1.1.4 TFIIF

TFIIF is a heterotetrameric complex containing two large (RAP74, 74 kDa) and two small (RAP30, 30 kDa) subunits initially identified in mammalian cells as RNA polymerase II associated proteins essential for transcription [170 - 172]. TFIIF was found to be a three-subunit complex in yeast [468, 469]. TFIIF subunits contain many structural domains [115, 179 - 182, 208] that are responsible for the integrity of the complex and extensive interactions with the general factors TFIID, TFIIA and TFIIB, RNA polymerase II, different transcription activators as well as with DNA [146, 183 - 194, 203].
TFIIF tightly binds to the extended surface of RNAP II and specific interactions were detected with RPB4/RPB7 dimer and with the RPB5 and RPB9 subunits [259, 363, 470, 471]. These interactions have an effect on suppression of nonspecific binding of RNA polymerase to DNA and strong stabilization of the polymerase at the promoter in a complex with other GTFs [138, 167, 173, 174, 195, 196, 259].

The structure of TFIIF within the TBP-TFIIB-TFIIF-RNAP II-DNA complex has been resolved by crosslinking studies [177, 178]. TFIIF appears to be essential for inducing a specific DNA topology with tight wrapping of promoter DNA around RNAP II within the PIC [202]. Both subunits of TFIIF were found to contact DNA near the TATA-box that results in specific conformational changes of the complex and proper positioning of RNAP II on DNA [175, 176]. In this way TFIIF plays an integral role in PIC formation and stability and is required for accurate initiation and start site selection [1, 3, 18, 167 - 168]. The binding of TFIIF to promoter is a critical prerequisite for the entry of TFIIE and TFIIH into the preinitiation complex [167, 196 - 198] and for open complex formation prior to the synthesis of the first phosphodiester bond of nascent transcripts [199 - 201].

Following transcription initiation, TFIIF is required for efficient escape of RNA polymerase II from the promoter and for efficient transcription elongation [205, 206, 209, 240]. At early elongation phases (immediately after transcription initiation) TFIIF in cooperation with TFIIH suppresses the premature stalling of RNAP II [206]. Furthermore, TFIIF is directly involved in the ternary elongation complex via interaction with RNAP II and different elongation factors [211, 217 - 220]. It potently activates the rate of elongation because it is required for efficient NTP-driven forward translocation of RNAP II [474]. In addition, TFIIF functions to suppress a transient pausing of RNAP II at DNA [3, 205, 209, 210, 217, 473] similarly to other general elongation factors [205, 206].

Several studies revealed that TFIIF seems predominantly associate with certain regions within the long transcribed genes [212, 213] and transiently associates with paused polymerase to induce optimal elongation rate [211]. The molecular mechanism underlying this activation is unclear but appears to occur in part via interaction with FCP1 protein (chapter 1.1.7) and regulation of FCP1 phosphatase activity towards the RNAP II C-terminal domain (CTD), or independently on enzymatic activity of the FCP1 [163, 184, 207, 215].

The activity of TFIIF itself is modulated during initiation and elongation processes by selective covalent modifications. The subunit RAP74 can be phosphorylated by TAF1 and TFIIH [221 - 224] and also auto-phosphorylated [225]. In addition, the subunit RAP30 is auto-acetylated [226].

### 1.1.5 TFIIE

TFIIE, like TFIIF, also consists of two large (TFIIEα, 56 kDa) and two small (TFIIEβ, 34 kDa) subunits [115]. Both subunits contain different structural domains [5] that interact with general transcription factors TFIIB, TFIIF and TFIIH [3, 166, 233, 454, 455], RNA polymerase II [152, 171, 228] and transcription activators [234, 235]. TFIIE interacts with double-stranded and single-stranded DNA of the promoter possibly via its central core domain [152, 229, 244, 145]. Protein-DNA crosslinking studies revealed that TFIIE binds upstream of the transcription start site approximately between −10 and +1 and that these interactions may facilitate and stabilize the separation of the DNA strands during open complex formation [146, 202, 229, 244, 245].

TFIIE is involved in late events of transcription initiation and formation of the preinitiation complex. In stepwise assembly models, which are based on in vitro complementation assays, TFIIE enters PIC after RNAP II and can be recruited by transcription activators. TFIIE interacts directly with TFIIF, TFIIB, RNAP II and promoter DNA [171, 233, 248, 245, 454, 456]. TFIIE functions in close
connection with TFIIH: it regulates recruitment of TFIIH to the promoter and also stimulates its ATPase, kinase and helicase activities [145, 196, 222, 227, 237 - 240, 242]. TFIIIE cooperates with TFIIH in ATP-dependent melting of promoter DNA and formation/maintenance of the open promoter structure that is prerequisite to the formation of the first phosphodiester bond and facilitates the transition of RNAP II to the elongation phase [115, 196, 200, 240, 246, 248]. Additionally, TFIIIE is also involved in conformational changes at the active center of RNAP II upon its interaction with DNA [152].

1.1.6 TFIIH

TFIIH is a multiprotein complex consisting of 10 subunits: p89/XPB, p80/XPD, p62, p52, p44, p40/CDK7, p38/cyclin H, p34, p32/MAT1 and p8/TFB5 [290 - 296]. Two subcomplexes within TFIIH have been identified: a tight core complex consisting of five subunits (XPB, p62, p52, p44, p34), and a second complex of cyclin-activating kinase (CAK) consisting of three subunits (cyclin H, CDK7, and MAT1). The XPD subunit mediates the binding of the two subcomplexes into one entity [297, 298]. The p8/TFB5 subunit contributes to the stability of TFIIH because mutations in the gene coding for TFB5 increased degradation of TFIIH. In addition, this subunit regulates ATPase activity of XPB and is essential in DNA damage response [295, 296, 299]. TFIIH possesses three enzymatic activities that are essential for transcription: DNA-dependent ATPase [290, 300], ATP-dependent helicase [301 - 303] and CTD kinase [237, 304].

The recruitment of TFIIH to promoters occurs via tight interaction with TFIIIE during late stages of transcription initiation. These two factors cooperate in formation of an open promoter structure before elongation can be started [196, 200, 240, 246, 248, 115]. The DNA-dependent ATPase and helicase (XPB subunit) activities of TFIIH are directly required for promoter opening before the formation of the first phosphodiester bond. TFIIH makes contacts with promoter DNA upstream and downstream of the transcription start sites and catalyzes ATP-dependent unwinding of DNA in the 3'-5' direction and, as a consequence, separation of DNA strands [199, 244, 263, 264, 267, 271, 273, 275, 276, 307]. This is supported by the finding that transcription can be initiated without TFIIH from negatively supercoiled or premelted DNA templates [200, 311].

The catalytic activities of TFIIH are also involved in the promoter escape – the process of the transition of RNA polymerase from preinitiation to the productive elongation stage [268]. It takes place after synthesis of the first phosphodiester bond of RNA and may be a rate-limiting step in transcription because the short transcripts (<10 nt in length) are unstable and susceptible to premature transcriptional arrest [268 - 274]. TFIIIF, TFIIIE and TFIIH cooperate to suppress this early elongation arrest via mechanism that requires helicase activity of XBP subunit [269 - 273]. Indeed, without TFIIH the fraction of stalled RNA polymerase close to promoters was significantly higher than in the presence of TFIIH and ATP [268, 272, 273]. In addition, the kinase activity of TFIIH plays a critical role in promoter escape: the CDK7 subunit phosphorylates serine 5 residues on the heptarepeat of the CTD that leads to the release of RNAP II from the PIC, exchange of its interacting factors and transition of RNAP II to the elongation stage [278, 313 - 316].

TFIIH has also been shown to interact with diverse transcription activators including hormone receptors [283]. This suggests a gene-specific coactivator function because activators can enhance the recruitment of TFIIH to promoters and stimulate the enzymatic activities of TFIIH [283, 481]. On the other hand, TFIIH itself can regulate the function of activators [283 - 285]. For example, kinase activity of CDK7 subunit can phosphorylate RARa and, consequently, regulate the activity of this hormone receptor [318]. The kinase activity, in turn, can be regulated by other subunits of TFIIH.
In addition to its regulatory role in RNAP II transcription, TFIIH is required for ribosomal RNA synthesis by RNAP I [288, 289]. Furthermore, apart from transcription TFIIH complex is involved in cell cycle control during the transition from G2 to M phase [138, 281, 282] and nucleotide excision repair (NER) in which damaged DNA is replaced by newly synthesized based on complementary strand [303, 320 - 322]. Defects in repair function can cause several diseases and premature aging [286, 287]. Thus, TFIIH is a multifunctional factor that is involved in three essential functions of the cell: transcription, DNA repair and cell-cycle control.

1.1.7 RNA polymerase II
RNA polymerase II is a central part of the transcription machinery that carries out the fundamental reaction of synthesis of an RNA chain based on the DNA template. It transcribes all protein coding genes and most of the non-coding RNA genes. RNAP II is a target for extensive regulation during every step of the transcription process (initiation, promoter escape and elongation, termination) via interaction with various transcription factors, and also by itself it has many regulatory functions that are important for the different processes.

In the eukaryotes the core RNAP II is a complex of 12 subunits that possess different properties. The subunits are designated as RPB1 to RPB12 by decreasing order of their molecular weight [323]. The eukaryotic subunits exhibit high structural and functional conservation and most human subunits can be substituted by the corresponding counterparts in yeast [348, 349]. Moreover, RPB1, -2, -3 and -11 share similarity with bacterial RNA polymerase core enzyme [3, 350 - 353]. Five subunits (RPB5, -6, -8, -10 and -12) are shared with RNA polymerase I, -II and -III and four other subunits (RPB1, -2, -3 and -11) have homologous counterparts in RNAP I and RNAP III [3, 5, 323, 354, 355]. Only three subunits (RPB4, -7 and -9) are unique to RNAP II. The Rpb4 and Rpb7 were found to be unstably associated with the main 10-subunit entity of RNA polymerase, they are required for initiation and PIC formation but are dispensable for elongation [78, 79, 115, 138, 143, 145, 276, 280, 314, 375, 377]. Another unique structure on the large subunit RPB1 is the carboxy-terminal domain (CTD) that consists of highly conserved tandem repeats with a consensus heptapeptide sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser [3, 138, 324]. The number of heptapeptide repeats varies from 26 in yeast to 52 in humans and is critical for the functioning of the RNAP II.

Structure of RNAP II
The extensive investigations of RNA polymerase structure by various techniques [5] visualized a highly complex network of molecular interactions and provided insight into the organization of the transcription machinery and mechanisms of transcription regulation [115, 145, 357, 361]. The structure of RNA polymerase appears to be highly conserved among different organisms implying high functional conservation [70 - 72, 361, 362]. Generally, several modules can be distinguished that form a main channel which comprises a catalytic domain and accommodates single stranded DNA and 8 base pairs of the DNA-RNA hybrid [80, 81]. In addition, there is a second channel which serves for the entry of NTP substrates selectively depending on complementary DNA nucleotide [145, 325]. The detailed molecular structure and interactions in the active center have been resolved at high resolution in a three-dimensional view [362, 150]. The active center has the binding sites for NTP and 3’ RNA terminus and accommodates two Mg$^{2+}$ ions that are essential for its catalytic function - the transfer of a nucleotidyl moiety from the NTP substrate to the 3’-hydroxyl group on the RNA
terminus [151, 327]. This transfer induces a translocation of the DNA/RNA hybrid to the +1 position relative to the RNA polymerase in the elongation complex. The essential feature of the polymerase is that this translocation can be reversed when the active center performs hydrolysis of the previously synthesized phosphodiester bond via intrinsic endonuclease activity [328 - 331]. The specific molecular mechanisms and regulatory determinants of this backtracking have been investigated and revealed that this proofreading activity of RNAP II is significantly higher in case when non-complementary NTP is misincorporated [332 - 334].

A structural view on RNAP II interactions with other general factors has been achieved. The contacts with TFIIB have been resolved in isolated system [147] and in the context of fully assembled PIC [144, 146]. It has been identified that the finger domain of TFIIB coincides with the position of the 8-basepair DNA-RNA hybrid within the active center of RNA polymerase, and the linker and core domains of TFIIB also interact with RNAP II and contribute to the regulation of PIC [147, 150, 151, 258]. The structures of RNA polymerase complexes with other initiation factors such as TFIIF and mediator have also been resolved [363, 364] as well as the architecture of elongating RNAP II in complex with elongation factor TFIIIS [327, 365].

Although RNAP II is mainly referred to the catalytic entity, different subunits have been shown to possess other specialized functions such as transcription start site selection, stabilization of the transcription complex, regulation of elongation and interactions with various transcription factors [3, 138, 115].

CTD structure, modifications and functions

Critical regulation at all transcription stages occurs via the RNAP II carboxy-terminal domain (CTD). The ability of the CTD to interact with various regulatory proteins is dependent on the dynamic exchange of specific covalent modifications [335 - 337]. The CTD only consists of hexamer repeats and does not fold into a specific, stable structure. Although the extended CTD is very long (up to 120 nm) its actual shape may be much more compact possibly due to a tendency to form beta-turns [150, 151, 336, 338 - 340]. Two forms of RNAP II can be distinguished based on the phosphorylation state of the CTD [5]. One form (named IIA) contains a hypo- or unphosphorylated CTD and is involved in PIC assembly and transcription initiation [302, 367]. The other form (named IIO) is present during elongation and termination phases and has CTD phosphorylated mainly at serine residues 2 and 5 [371 - 374] with, on average, one phosphate per repeat [370].

Several mammalian protein kinases possess an activity towards the CTD: the CDK7, CDK8 and CDK9, that have also yeast homologs. They belong to the cyclin-dependent kinase family (CDK) and their regulation occurs via association with cyclins via formation of CDK7-cyclin H, CDK8-cyclin C and CDK9-cyclin T pairs. The structural features of these complexes were addresses in recent studies but the determinants of their specificity towards CTD are still poorly understood [336]. The CDK7/cyclin H pair associates with the MAT1 protein to form the CDK-activating kinase (CAK). The CAK forms a complex with the general transcription factor TFIIH and phosphorylates the CTD at serine 5 residues during transcription initiation [93, 276, 375, 377 - 380]. Upon interaction with Mediator complex during initiation, the kinase activity of TFIIH is enhanced and subsequently facilitates the transition to elongation and recruitment of RNA processing factors [315, 381]. Another complex that targets serine 5 phosphorylation is the pair of CDK8 and cyclin C proteins. It forms a distinct module within the Mediator complex via association with the subunits MED12 and MED13 [383 - 386] and can influence transcription both positively and negatively. The positive function is linked to ATP-dependent dissociation of preinitiation complexes triggered by CDK8 [387] or
phosphorylation of the MED2 subunit in the Mediator complex [388]. Repression of transcription by CDK8 can occur via premature phosphorylation of the CTD that prevents PIC formation [388]; it can also deregulate CDK7 via phosphorylation of its associated protein cyclin H [277]. Phosphorylation of the CTD at the serine 2 residues is mediated by the CDK9/cyclin T complex. It is a core component of the positive transcription elongation factor P-TEFb [390 - 395]. P-TEFb facilitates RNP II function [396, 397] and counteracts the negative effect of DSIF and its cofactor NELF during elongation [398, 399]. The function of CDK9/cyclin T is subject to regulation by 7SK snRNA [392 - 395].

The dynamic state of CTD phosphorylation is mediated by the action of specific phosphatases that is involved in transcriptional process. The human small CTD phosphatase 1 protein (SCP1) [400], the plant proteins AtCPL1 and AtCPL2 [401] and yeast Ssu72 phosphatase [402] are able to dephosphorylate CTD at serine 5. The Ssu72 is highly conserved and essential throughout transcription. It binds to TFIIIB and can enhance its in initiation defects in mutational studies [156, 403, 405, 406]. Ssu72 is also involved in mRNA elongation, as its mutation increases RNP II pausing [406] and termination of snRNA and mRNA genes partly via its involvement in cleavage and polyadenylation of 3'-ends of mRNA. [407, 408].

Dephosphorylation of serine 2 is mainly carried out by TFIIF-associated CTD phosphatase 1 (FCP1) [409, 411] that has been isolated from yeast [411, 412] and human cells [163, 413]. The RNAP II-associated factor TFIIF can stimulate FCP1 phosphatase activity and, therefore, may accelerate reinitiation of transcription by enhancing the conversion of RNP II from the elongating IIO form back to the initiating IIA form [184, 185, 207, 413, 418, and chapter 1.1.4). FCP1 also interacts with TFIIIB [184, 185], the RPB4/7 subunits of RNAP II [184, 412, 419] and also with the phosphorylated CTD [423]. Via these interactions, FCP1 may stay involved in elongation process and is able to dephosphorylate transcription elongation complexes [184, 420 - 422]. Highly purified Fcp1 is also able to dephosphorylate serine 5 [422].

In addition to phosphorylation, the CTD can be modified by glycosylation via covalent linkage of a monosaccharide, N-acetylglucosamine (GlcNAc), onto the side chain hydroxyl group of serine or threonine residues in the heptapeptide repeat throughout the entire CTD domain [433]. This process is carried out by an O-GlcNAc transferase (OGT), and the reverse deglycosylation reaction - by an O-GlcNAc-specific b-N-acetylglucosaminidase (O-GlcNAcase) [424]. Glycosylation is only found in the IIA form of RNP II (with unphosphorylated CTD) and appears to be mutually exclusive with phosphorylation on the CTD [424, 425]; it is possible that glycosylation regulates phosphorylation during transcription initiation via blocking kinase accessibility to the CTD. Additionally, glycosylation may alter CTD conformation and consequently may regulate dynamic interactions with different proteins.

RNAP II, and in particular the CTD, plays a central role in regulation of transcription and transcription-linked processes such as RNA processing and DNA repair (chapter 1.6). Many studies addressed the questions how the simple heptapeptide sequence can be specifically recognized by a large variety of proteins. The known covalent modifications of the CTD were found to be crucial for establishing specific structural configurations. The dynamic changing of these modifications during transcription allows efficient regulation and exchange of interacting proteins at different stages of transcription process. Thus, phosphorylation of serine 5 occurs in promoter-proximal regions and leads to initiation of transcription and recruitment of mRNA-capping enzyme guanylyltransferase [280, 313 - 316, 429, 430]. Phosphorylation of serine 2 occurs mainly in regions that are more distal from the promoter and is required for the recruitment of 3'-end polyadenylation factors [314, 341 -
345, 409]. Although no simple structural basis for these interactions was found, extended beta-strand-like conformations may be recognized in the CTD structure; hydrophobic interactions together with direct or indirect recognition of phosphorylated serines may be the principal interaction forces [336]. In addition to phosphorylation, CTD undergoes cis-trans isomerization at the two proline residues (P3 and P6) adjacent to both serines on the C-terminal side. The proline isomerase Pin1 acting at those prolines in connection with phosphoserines has been implicated in regulation of phosphatase Fcp1 [431] and formation of 3’ end of mRNA [346, 347]. In all described structures of interacting CTD both prolines displayed trans conformation and their isomerization to cis would likely impair the binding [338]. The architecture of CTD-binding domains and structural features of their interactions were described in details [336].

1.1.8 Mediator
Mediator is a large multiprotein complex consisting of about 30 subunits most of which share significant structural and functional conservation between yeast and mammalian homologs [565]. Mediator is found essential for RNAP II transcription and the main function currently considered is the major coactivator that transmits diverse regulatory signals (e.g. from gene-specific DNA-binding transcription factors) to the general transcription factors at the promoter, i.e. serves to enable activators to regulate transcription initiation [566, 567]. The mediating strategy, possibly, provides possibility for fine tuning of transcription instead of direct interaction of activators with general factors and RNA polymerase.

Mediator is organized into several sub-complexes that form different modules. The central module consists of about 20 tightly associating proteins and is characterized originally as an independent mediator-like complex [569, 581 - 583]. Additional subunits were also found to be required for mediator function [567] although the precise stoichiometric composition of Mediator is not defined precisely.

The modular structure of mediator is established based on structural microscopic studies [575, 584]. The head module is able to directly interact with subunits of RNAP II and some activators [568 - 570]. The middle module also interacts with several subunits of RNAP II [385, 364, 572] and both middle and tail modules interact with activators and various nuclear receptors [570,574]. An additional module consisting of 4 subunits (MED12, MED13, CDK8, CycC) can enter mediator via interaction with head and middle modules but also exists as an independent complex [384, 572]. The CDK8 subunit is able to phosphorylate serines 2 and 5 of the RNAP II CTD which has inhibitory effect on transcription if occurs prior to PIC formation [384, 388, 575]. The repression role of the CDK8 module was also deduced from its ability to phosphorylate TFIID [277] and further supported by observation that the inactivation of CDK8 can result in transformation of inactive Mediator to the transcriptionally active form and activation of a subset of genes [578, 93]. The modular multisubunit structure can make Mediator able to integrate diverse transcription regulatory signals.

Mediator is assumed to function as coactivator that can associate with the chromatin via interaction with activators and can further promote recruitment of RNAP II via its head module. It positively affects the assembly of a preinitiation complex at promoters resulting in transcription activation. This model is further supported by the finding that mediator interacts with the unphosphorylated form of CTD [586, 591] and these interactions are disrupted upon CTD phosphorylation when RNAP II enters into its elongation state [380]. Interactions with activators can also induce conformational changes of the Mediator that can contribute to its function [575]. It was found that different subunits of mediator are involved in activator-specific transcription regulation as inactivation of specific subunits affects
expression of certain genes and results in abolishing of selective signalling and differentiation pathways [566]. In addition, in yeast Mediator can acetylate histones H3 and H4 contributing to transcription activation [587]. Mediator is also able to enhance basal transcription independently of activators [588, 589].

1.1.9 TBP-containing complexes implicated in RNAP II transcription

In addition to TFIID, TBP was found in several distinct complexes without TAFs. The human B-TFIID complex consists of TBP and BTAF1 (yeast Mot1), a Snf2/Swi2-related ATP-ase [540, 541]. BTAF1 can regulate binding of TBP to DNA and is able to dissociate TBP from DNA using ATP hydrolysis [542 - 544]. Transcriptional profiling analysis in yeast suggests that Mot1 can positively and negatively affect expression of a subset of genes [545, 546].

Another complex, TAC, consists of TBP, the unprocessed form of TFIIA\alpha\beta and TFIIA\gamma subunits of TFIIA. It was found specifically in undifferentiated cells and implicates in active RNAP II transcription [51, 553].

Negative cofactor 2 (NC2) can form a complex with TBP and block binding of TFIIA and TFIIIB that is implicated in inhibition of basal transcription [547, 548]. In addition, NC2 can stimulate transcription from different promoters [31, 550] and in vivo is involved in both positive and negative regulation of transcription in yeast [551]. The recent finding of an interaction between NC2 and BTAF1 suggests a functional cooperation between both factors in transcription regulation [552].

1.1.10 TBP-related factors

TBP is an essential and universal factor that is required for transcription initiation by all three RNA polymerases in eukaryotes. A number of proteins structurally and functionally similar to TBP were identified in different organisms and called TBP-related factors (TRF). Recent studies revealed an important regulatory role of these factors in transcription of a subsets of genes.

The TBP-related factor 1 (TRF1) is expressed in the nervous system and germ cells of Drosophila [593]. The C-terminal core DNA-binding domain has 63% identity to the TBP in amino acid sequence. TRF1 is similar to TBP as it interacts with TFIIA and TFIIIB, can bind to TATA-box DNA and direct RNAP II transcription in vitro [253]. Cytologically, TRF1 was found associated with a limited number of sites distinct from TBP on Drosophila chromosomes suggesting that it functions in transcription of a few tissue-specific genes. It was found to drive transcription from an alternative promoter of the Tudor gene as a part of the multiprotein complex with TRF1-associated factors that are distinct from the TAFs is TFIID [595]. DNA footprinting revealed that it binds to a TC-rich sequence at −22 to −33 relative to the transcription start site. Interestingly, the Tudor gene has a second promoter that is TBP-dependent and not responsive to TRF1 suggesting a mechanism for alternative gene expression in response to different pathways and stimuli [596]. In addition to its role in regulation of RNAP II genes, TRF1 is mainly present in a stable complex with BRF, a component of the TFIIB complex [597]. The TRF1-BRF, but not a TBP-BRF complex, is essential for RNAP III transcription of the tRNA, 5S RNA and U6 RNA genes. This is different in most of the other species and cells types characterized thus far where RNAP III genes are dependent on TBP.

The TBP-related factor 2 (TRF2), in contrast to TRF1, was found in many different organisms including Xenopus, mouse and man [598]. It has about 50% similarity with the C-terminal DNA-binding domain of TBP, but the amino acids crucial for interaction of TBP with the TATA-box are substituted and, therefore, TRF2 can not bind the TATA-box but might instead recognize different sequences [254, 599]. However, similarly to TBP, TRF2 associates with TFIIA and TFIIIB [254, 256].
and can be a part of functional preinitiation complex required for transcription of specific genes [602], and may have a redundant role in transcription of the TFIID-dependent genes [606]. It can also function as a transcription repressor possibly via competition with the general factors TFIID and TFIIB [256, 602]. Depletion of TRF2 in C. elegans and Xenopus leads to defects in early development as a result of a decrease in transcription of specific genes [599, 604, 605]. Removal of TRF2 gene from mouse cells does not affect the viability, but results only in defects of expression of testis-specific genes [607, 608]. TRF2 can be present in complex with components of the NURF chromatin remodeling complex and with DREF, the DNA replication-related element (DRE)-binding factor [609]. DREF drives transcription from an alternative promoter of PCNA (proliferating cell nuclear antigen) gene possibly by functioning as a core promoter-selectivity factor.

TRF3 has 93% amino acid sequence identity to TBP at the C-terminal core region and can interact with the TATA-box as well as with TFIIA and TFIIB. TRF3 is expressed in various tissues in different vertebrate organisms and is present in a multiprotein complex, but its transcription properties are not yet characterized [610].

### 1.1.11 TBP-free TAF-containing complexes

A number of distinct multiprotein complexes containing TBP-associated factors, but lacking TBP, were found in cells from different organisms. The complexes such as TBP-free TAF-containing complex (TFTC), TFTC-related PCAF/GCN5 complexes, Spt-Ada-Gcn5 acetyltransferase (SAGA), SAGA-like complex (SLIK), Spt3-TAF9-GCN5L acetylase (STAGA), and polycomb complex 1 (PRC1) possess multiple and essential functions in general transcription regulation of RNAP II genes [5]. These complexes are implicated in gene activation mainly due to their histone acetyltransferase (HAT) activity and coactivator function of TAFs, whereas PRC1 is implicated in gene silencing. The subunits composition and the biological functions are listed in Table 2.

The TBP-free TAF-containing complex, TFTC, contains many histone fold TAFs which are required for the integrity of the complex. In addition, it contains a number of large proteins such as GCN5 and TRRAP which are implicated in transcription activation. The three-dimensional structure of TFTC revealed specific domains similar to those found in TFIID which can bind to DNA and have similar to TFIID functions [611]. Indeed, TFTC can drive basal and activator-mediated transcription in vitro from TATA-box containing and TATA-less promoters [617]. The coactivator function of TFTC is mediated by its ability to support basal transcription in response to direct interactions with activators, nuclear hormone receptors [612 - 615] and to acetylate histone H3 via the GCN5L subunit which is implicated in active transcription [616].

The yeast Spt-Ada-Gcn5 acetyltransferase (SAGA) complex contains subunits common with TFTC, such as Tra1 (the homolog of human TRRAP), Ada3, GCN5, Spt3, TAF5, TAF6, TAF9, TAF10, and TAF12, but also has specific subunits with unique functions (Table 2). SAGA has an essential regulatory role in RNAP II transcription. Among the prominent properties of SAGA are coactivator function, interaction with TBP and methylated histone H3, HAT and deubiquitinating activities [5]. The complex forms five globular domains similar to those found in TFTC [631]. The subunits Sp7, Spt20, Ada1, TAF5, TAF10, TAF12, and Sgf73 are necessary for the structural integrity of the SAGA complex and TAF5 has a central structural role [622, 628, 631, 642, 643]. The largest subunit Tra1 forms an outer domain which is able to interact with transcriptional activators [632 - 634] similarly to the functions of human TRRAP [620, 621]. A recent ChIP-on-chip study revealed broad occupancy of TRRAP at large number of human promoters suggesting that this protein has a rather general role in transcription [561].
About 10 - 20% of yeast genes are regulated by SAGA [635, 636]. Remarkably, these genes are mainly of highly regulated type (e.g. stress response pathways) that contain a TATA-box in their promoters rather then housekeeping type that mainly lack TATA sequence. These genes also require the TAF-free form of TBP that functions on target promoters with SAGA via interaction with the Spt3 and Spt8 subunits [130, 637]. The preinitiation complex and TBP can be stabilized via recruitment of SAGA by transcription activators and possibly also mediated via histone H3 acetylation by the GCN5 subunit [622, 640]. In addition, the Chd1 subunit of SAGA binds to methylated lysine 4 of histone H3 at active promoters [644, 645] and enhances GCN5-mediated acetylation of histone H3 [629]. The SAGA complex can also function in transcription silencing via binding to TBP and inhibition of its association with promoters or with TFIIA [130, 637].

The SAGA-like complex, SILK, has similar composition as SAGA except for Spt8 and with the addition of the SLIK-specific subunits Rtg2 and a C-truncated Spt7 protein [625, 628, 629, 646, 647]. SILK displays similar functions as SAGA such as interaction with activators, TBP and methylated histone H3, HAT activity, and deubiquitination of histone H2B although there are minor differences in gene-specific regulation [647, 648].

STAGA, the Spt3-TAF9-GCN5L acetylase complex is the counterpart of yeast SAGA in human cells. It contains subunits common with TFTC and homologs counterparts found in SAGA (Table 2). The structural similarity with SAGA suggests common functions. For example, structural integrity of SAGA can be maintained by ataxin-7 and histone fold proteins [614]; TRRAP, GCN5L, Ada1, Ada2, Ada3, and ataxin-7 are involved in coactivator function targeting GCN5 HAT activity to promoters [615, 650, 651].

The PCAF complex consists of the PCAF protein (p300/CBP-associated factor) and about 20 other proteins [658]. The complex reveals similarity to SAGA/STAGA in its subunit composition suggesting that it performs similar functions [661]. PCAF protein binds to the p300/CBP coactivator and can be competed out by the adenoviral E1A oncoprotein which deregulates normal cell growth and binds to pRB and p300/CBP [659, 660]. The PCAF protein is homologous to the yeast GCN5 subunit of SAGA and possesses similar functions. Its HAT activity can acetylate histones and as well as transcription factors [661]. The p400 subunit is identical to TRRAP protein [662]. p300/CBP and RNAP II are weakly associated with purified PCAF complex suggesting their functional interactions at promoters in response to gene-specific activators [656, 663].

The TAF-containing polycomb repressive complex 1, PRC1, was characterized in Drosophila embryos as containing many TAFs amongst about its 30 subunits [652 - 654]. In both flies and vertebrates this complex functions to maintain a transcriptionally silent state of the developmental genes such as the HOX-family. Repression occurs via specific activities associated with the subunits of the complex, such as a histone deacetylase (HDAC) activity, chromatin remodeling ATPase and several other transcription inhibitory activities [652, 655 - 657]. A large number of subunits with different functions including gene-specific transcription factor Zeste suggest that the RPC1 complex is involved in a variety of repression pathways.

<table>
<thead>
<tr>
<th>Name</th>
<th>Subunit composition</th>
<th>Functions of the complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFTC</td>
<td>TAF2, TAF4, TAF5, TAF5L, TAF6, TAF6L, TAF7, TAF9, TAF9b</td>
<td>Forms direct contacts with activators via multiple subunits, such as TAF5, TAF6, TAF10, SAP130, Spt3, GCN5L, TRRAP, ataxin-7 [612 - 615].</td>
</tr>
</tbody>
</table>
complex) TAF10, TAF12; TRRAP (transformation-transactivation domain-associated protein), SAP130 (spliceosome-associated protein 130), GCN5L (HAT enzyme), Ada3 adaptor protein, Spt3, and ataxin-7 [520, 617].

TRRAP interacts with nuclear hormone receptors ERa, ERb, VDR, and PPARg [612]. Is an essential cofactor for c-Myc and E2F [620] and recruits GCN5 [621]. GCN5L acetylates histone H3 [616]. Ataxin-7 is essential for the structural integrity of TFTC and STAGA [614]; linked to SCA7 neurodegenerative disorder [615].

| SAGA (Spt-Ada-Gen5 acetyltransferase) | Subunits in common with TFTC: Tra1 (homolog of human TRRAP), Ada3, GCN5, Spt3, TAF5, TAF6, TAF9, TAF10, and TAF12. Unique subunits: Ada1, Ada2, Spt7, Spt8, Spt20, Chd1 (for chromo-ATPase/helicase-DNA binding domain 1), Ubp8 (ubiquitin-specific processing protease 8), Sgf11 (SAGA-associated factor 11 kDa), Sgf29, and Sgf73 [7, 622, 624 - 630]. Required for transcription of 10 - 20% of highly regulated yeast genes that have usually TATA-less promoters [635, 636]. Functions at promoters cooperatively with TAF-free TBP [130, 632, 637]. Acetylates histone H3 [622]. Sgf73 subunit is required for regulation of GCN5 HAT activity in SAGA [628], important for structural integrity of the complex [614]. SAGA can be recruited to promoters by transcription activators [622, 640] and via interaction of Chd1 subunit with methylated lysine 4 of histone H3 [629]. Tra1 interacts with transcriptional activators such as Gal4 and GCN4 [632 - 634]. HAT activity and non-specific DNA-binding of SAGA are enhanced by proteasome regulatory 19S subcomplex [630]. Ubp8 and Sgf11 subunits function in gene activation via removal of the monoubiquitin residue from lysine 123 of histone H2B [624 - 627, 630]. Negative regulation of transcription by Spt3 and Spt8 subunits via binding to TBP and preventing TBP from binding to promoter DNA and TFIIA [130, 637]. |
| SLIK (SAGA-like complex) | Common subunits with SAGA: GCN5, Ada1, Ada2, Ada3, TAF5, TAF6, TAF9, TAF10, TAF12, Tra1, Spt3, Spt20, Chd1, Ubp8, Sgf11, Sgf29, and Sgf73 [646, 647]. Has similar functions as SAGA such as: interaction with activators, TBP and methylated histone H3, HAT activity, deubiquitination of histone H2B. Differ from SAGA in gene-specific regulation [625, 628, 629, 646 - 648]. |
### General introduction

<table>
<thead>
<tr>
<th>Complex</th>
<th>Description</th>
<th>Subunits</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCAF (p300/CBP associated factor (PCAF) acetylase containing complex)</td>
<td>More than 20 subunits, including PAF400 (identical to TRRAP), Ada2, Ada3, Spt3, TAF9, TAF10, TAF12, PCAF (Gcn5-related HAT), TAF-related factors PAF65a and PAF65b</td>
<td></td>
<td>Structurally and functionally similar to SAGA/STAGA [658 - 663]. Interacts with p300/CBP and RNAP II [659, 663]. HAT activity acetylates histones as well as transcription factors [661].</td>
</tr>
<tr>
<td>PRC1 (polycomb repressive complex 1)</td>
<td>Contains about 30 subunits, such as TAF1, TAF4, TAF5, TAF6, TAF9, TAF11, PH, PSC, PC (polycomb), RING1, Zeste, HSC4, SMRTER, Mi-2, Sin3A, Rpd3, p55, Sbf1, DRE4/Spt16, p90, HSC3, Modulo, Reptin, DNA topoisomerase II, p110, tubulin, actin, RS2, RL10</td>
<td></td>
<td>Subunits have various functions and the complex can repress transcription via different pathways [652 - 657]. Rpd3 is histone deacetylase (HDAC); Mi-2 is chromatin remodeling ATPase; SMRTER, Sin3A, p55 are general transcription repressors. PH, PSC, PC, RING1 subunits can form a repressive sub-complex.</td>
</tr>
</tbody>
</table>
Chapter 1

1.2 Core promoter of the RNAP II genes

1.2.1 Structure of the core promoter

The central place for assembly of the preinitiation complex is the core promoter region spanning about 60 bp around the transcription start site. It is responsible for a key event of proper positioning of RNA polymerase II and initiation of transcription at the correct start site. The core promoter comprises several critical DNA elements that are common in the majority of the cellular genes transcribed by RNAP II. These elements include the TATA-box (the binding site for TBP), BRE (TFIIB-recognition element), Inr (initiator element), DPE (downstream promoter element) and MTE (motif ten element) [82]. Most of the promoters contain one or more of these elements, but no one element is absolutely essential for promoter function. The sequence of these elements, though relatively conserved between the genes and species, shows a certain degree of structural diversification.

The TATA box is located near position -30 to -25 relative to the transcription start site (+1) and has a consensus sequence TATA(A/T)A(G/A) but also a wide variety of AT-rich sequences were found to function as TATA-box. This sequence is recognized by TATA-binding protein (TBP) [112, 260, 262, 555, 558]. Although the sequence of the TATA-box is not symmetrical and has the potential to dictate the direction of transcription, TBP does not bind the sequence with high orientation specificity [65] therefore it contributes little to directionality and mainly functions as a scaffold to anchor larger transcription complexes at promoters.

The BRE element is located immediately upstream of the TATA-box in about 12% of TATA-containing promoters. It has a consensus sequence (G/C)(G/C)(G/A)CGCC and is specifically recognized by the helix-turn-helix motif at the C-terminus of TFIIB [82, 261]. In addition, TFIIB can recognize another element (BRE) localized downstream of TATA-box with the consensus sequence (G/A)T(T/G/A)(T/G)(T/G)(T/G)(T/G). This interaction stabilizes a TFIIB–TBP–promoter complex and was shown to modulate promoter strength both in vitro and in vivo [137].

The Inr core element is located at transcription start site [13] and has a consensus sequence PyPyA(+1)N(T/A)PyPy with typically an adenosine at +1 position. The function of Inr does not always dedicate transcription to begin at the +1 nucleotide and RNAP II can be redirected to alternative start sites by e.g. altering the spacing between the TATA-box and Inr if the promoter contains both of these elements. In this case the Inr can increase the efficiency of transcription initiation from alternative sites [14]. The spacer of 25–30 bp separating the TATA-box and Inr is critical for these two elements to function synergistically in transcription activation [15].

The conserved DPE element was identified downstream of the transcription start site precisely at position -28 to -32 relative to the A(+1) nucleotide in the Inr motif [16]. DPE has the consensus sequence (A/G)G(A/T)(C/T)(G/A) and was typically found together with Inr in TATA-less promoters. DPE and Inr elements function cooperate together by stabilizing TFIID binding to the promoters.

The MTE is generally found in promoters as a conserved core element [11] located precisely at positions +18 to +27 relative to A(+1). This element is important for basal (activator-independent) transcription. It shows synergism with the TATA-box and DPE but also can compensate for loss of these elements and act independently in the presence of Inr element.

1.2.2 Recognition of the core promoter by general transcription factors

One of the primary steps in transcription initiation is PIC assembly and proper positioning and directionality of RNA polymerase. It is directed via interaction of specific transcription factors with
core promoter elements. A number of interaction mechanisms were identified although the complete picture of the interacting factors with core promoters still needs to be unraveled.

The TATA-box is recognized by TBP which forms a saddle-like structure on DNA [117]. This molecular saddle consists of two symmetrical domains each containing ~90 amino acids. The binding of TBP does not occur in a polar manner and is not sufficient for determine promoter directionality, thus the other core elements are required. TFIIIB can bind to BRE that enhances PIC assembly and together with RNA polymerase II dictates the distance between the TATA-box and the transcription start site [18]. The role of TFIIIB in this process was revealed by mutational analysis showing that the N-terminal charged domain can shift the location of the start sites by a few nucleotides [19]. The possible mechanism of this shift could be the result of a conformational change in the TFIIIB or RNAP II complex. Remarkably, the TFIIIB-BRE interaction does not only stimulate transcription in reconstituted assays in vitro [261] but also represses a basal transcription in vitro as well as in vivo [20].

TFIID, the TBP and TAFs containing complex (chapter 1.1.1), plays a basic role in the recognition of the two other core elements – Inr and DPE [106]. Analysis of TFIID binding to core promoters containing both these elements revealed that the Inr is essential for stable TFIID binding [21, 22] which may occur mainly via stable and preferential interaction between Inr and a sub-complex of TAF1 and TAF2 [23]. Another sub-complex of TBP-TAF1-TAF2 was found to be sufficient for Inr activity in reconstituted transcription assays in vitro [24]. In addition to TFIID, RNAP II has a weak intrinsic preference in recognition of Inr-like sequences and initiates transcription inefficiently from the Inr-containing promoters via forming a stable complex with TBP, TFIIIB, and TFIF [25]. Thus, both TATA and Inr elements can be recognized via different subunits of TFIID complex which could be sufficient for proper RNAP II positioning. This model is supported by the known functional synergy between TATA and Inr when separated by 25 - 30 bp spacer in the promoter [15] although these elements alone are not sufficient for transcription initiation [27].

The DPE does not function independently (unlike TATA and Inr) and shows a strong dependence on the presence of the Inr element in DPE-dependent promoters. TFIID efficiently binds to both DPE and Inr elements that is sufficient for active transcription. Sequence alterations of the elements or changes in the spacing between them lead to significant decrease in the TFIID binding and transcriptional activity [21, 16]. Recognition by TFIID occurs via TAF6 and TAF9 subunits [21, 28, 29]. Furthermore, TAF9 preferentially associates with the DPE-containing human IRF-1 promoter relative to the DPE-less version of the promoter [30]. In addition to TFIID, DPE-dependent transcription requires also the protein kinase CK2 and the coactivator PC4 [36]. The negative cofactor-2 (NC2) can specifically differentiate between TATA- and DPE-promoters functioning in transcriptional activation and repression, respectively [31].

The MTE element can function both in conjunction and independently of the TATA-box and DPE. It was suggested to interact with general transcription factors because it can compensate for a loss of other core elements and weak recognition by TFIID [11]. These results suggest that the MTE contributes to the binding of TFIID to promoters.

1.2.3 Other core-promoter elements

Small nuclear RNA genes (snRNA) belong to a specific group of the genes which are transcribed by RNAP II but contain a promoter structure distinct from the majority of other RNAP II transcribed genes. They contain an essential, highly conserved proximal sequence element (PSE) at the position -45 to -60 that dictates the location of the transcription start site of snRNA genes [32]. An interesting
Chapter 1

feature is also that the PSE is present in core promoters of snRNA genes transcribed by RNAP III; it seems to be important for determination of polymerase specificity. Addition of a TATA-box at the position 15–20 nucleotides downstream of the PSE switches transcription from RNAP II to RNAP III whereas the genes are transcribed by RNAP II in the absence of PSE [33]. The PSE is recognized by small nuclear RNA-activating protein complex, SNAPc (chapter 3.2). The human complex is essential for the activity of both RNAP II and III-transcribed snRNA genes [34]. The subcomplex containing SNAP190, -43 and -50 binds to the PSE in a sequence-specific manner. The other two subunits, SNAP19 and -45 are important for complex integrity and also have other functions [35]. SNAPc-directed transcription requires TBP which binds to the snRNA promoters containing a TATA box in addition to a PSE [35].

The above described elements have rather constant positions relative to the +1 start site which is consistent with their specific roles in transcription initiation complex and PIC assembly. Additionally, there is a variety of other DNA sequence elements that are located at variable positions relative to the start site and that contribute to the core promoter activity. Many of the elements are associated with specific types of genes involved in related biological pathways during cell growth and differentiation.

The downstream core element (DCE) is present in a large number of different promoters and is mutually exclusive with the DPE [17, 249]. DCE contains three sub-elements: SI, SII, and SIII, with a core sequence of CTTC, CTGT, and AGC respectively. Each of these sub-elements can interact with TAF1.

Another element is the MED-1 (multiple start site downstream element) that is identified in TATA-less promoters containing multiple transcription start sites [37].

A distinct class of promoters was found to associate with CpG islands. There are about 30,000 CpG islands in the human genome. Promoters of about 80% of protein-coding genes in mammalian genomes are located in the CpG islands [38]. Despite of this abundance, the core elements that are essential for their promoter function remain poorly defined. CpG islands usually lack consensus TATA boxes, DPE or Inr elements [39]. In addition, they are often characterized by the presence of multiple transcription start sites. A common feature of the CpG islands is the presence of multiple binding sites for transcription factor Sp1 at 40-80 bp upstream of the transcription start sites [40] suggesting that Sp1 may direct the general factors to form a preinitiation complex [41]. The TAFs subunits of TFIID that recognize other core-elements may be especially important in regulation of the CpG island-type promoters [561].

1.2.4 Mechanisms of gene regulation via core promoter

The various core-promoter elements can function as equivalent and interchangeable recognition sites for TFIID. Their combinatorial diversity in eukaryotic promoters can have a significant contribution to the gene regulatory strategies in response to upstream activators [2]. In this case a limited number of transcription factors can support a much larger repertoire of gene expression patterns via combinatorial compositions of core-elements so that a particular gene can respond to specific activators or enhancers only if their core promoters contain certain elements.

Bioinformatic analysis of eukaryotic promoters revealed that less than 22% of the human genes contain a TATA-box, amongst which 62% have an Inr, 24% - DPE, and 12% - BRE [562]. From the more than 78% TATA-less promoters, 45% possess an Inr, 25% have DPE, and 28% have BRE. The exact functional properties of these promoters classes remain to be investigated. In general, promoters of human housekeeping genes, growth factors and transcription factors often lack a TATA-box [563]. It is important to note that only known promoters (mainly from protein-coding genes) were accessible.
for the analysis, whereas a large number of promoters from novel non-coding genes identified in several recent studies [561] may lead to a revision of the current notion.

There are several examples showing selective communication between a core promoter and specific transcription factors or enhancers. A core promoter selectivity in response to activators is found in mammalian cells for hsp70 and myoglobin genes. The replacement of native TATA-boxes with a viral analogs abolished transcription activation of these genes [42, 43]. Another study shows that the AE1 enhancer in Drosophila preferentially stimulates transcription from TATA-containing promoters rather than from Inr-DPE promoters [44]. The selective response of TATA-Inr and Inr-DPE promoters to enhancers is also illustrated in the context of the same genomic locations [45]. A contribution of the core promoter in responding to DNA-specific activators was found for the mouse terminal transferase promoter containing an Inr element. Its conversion to a TATA-box promoter abolished the response to Elf-1 activator which bound 60 bp upstream of the start site [46]. In addition, members from the large family of Ets transcription factors were found to have intrinsic Inr preference [47].

The mechanisms of the preference for the core promoters are not well understood. It is possible that the affinity of TFIID binding for the core elements is an important factor for certain promoters. For example, the transcription factor VP16 efficiently activates promoters containing two core elements [48]. On the other hand, the activator SP1 can stimulate transcription with equal efficiency from promoters containing either a TATA-box or Inr. SP1 is capable of recruiting TFIID and therefore be less dependent on the affinity of TFIID to the core elements. Other mechanisms may involve conformational changes of the TFIID complex as it was described upon its binding to different core promoters [49].

1.3 Mechanisms of the RNAP II transcription

1.3.1 Initiation

Transcription initiation is a complex process consisting of many molecular events that are directed towards the assembly of a PIC at promoters followed by initiation and elongation. The preinitiation complex consists of a large number of different proteins with various functions, such as multiple enzymatic activities directed towards transcription factors and histones, structural regulation, chromatin remodeling and transduction proteins for upstream regulatory signals. With respect to such a diversity of the factors it has been suggested that transcription of each and every gene, or a group of closely related genes, is individually regulated. Nevertheless, the core part of the PIC is considered to have relatively uniform composition consisting of a number of general transcription factors (GTFs) and RNA polymerase II (chapter 1.1).

Preinitiation complex

The initial step in PIC assembly is the specific recognition of the core promoter by TFIID (chapter 1.1 and 1.2). Together with other general transcription factors and Mediator the factors establish a recognition scaffold for stabilization of the RNAP II binding and its proper positioning at core promoter and start site selection [145, 240]. The studies that investigated the architecture of PIC revealed that extensive protein-DNA interactions extend over the promoter DNA between positions -43 and +24 relatively to the transcription start site and RNAP II subunits Rpb1 and Rpb2 make extensive DNA contacts over 60 bp [145, 177, 178, 244, 245]. TFIIB and TFIIF contact DNA both upstream and downstream of the TBP-binding site (TATA-box). TFIIE and TFIIF interact with DNA upstream of the TATA-box. The XPB helicase subunit of TFIIH interacts with DNA downstream and upstream of the transcription start site. It was found that promoter DNA is tightly wrapped around the
Chapter 1

PIC and that it is extensively interacts with general factors [178, 202, 240, 666, 683]. X-ray studies revealed the structural organization and topology of the RNAP II in complex with DNA and RNA [150, 361, 362, 671, 674]. The structures of initiation-competent RNAP II in complex with TFIIB, TBP, TFIIF or Mediator also have been described [140 - 143, 358, 364, 675, 678].

Two principal pathways, though non-mutual excluding, were suggested for PIC assembly. In one model, sequential assembly of PIC is suggested in which GTFs enter the complex in a stepwise manner [1, 138]. In vitro experiments in which purified components are sequentially added to the reaction suggested that the most efficient pathway is triggered by initial binding of TFIID to promoter followed by TFIIA and TFIIB recruitment, stabilization of promoter-bound TFIID and, finally, recruitment of RNAP II/TFIIF complex. Subsequently, TFIIE and TFIIH enter the PIC. This stepwise model is supported by ChIP (chromatin immunoprecipitation) studies showing an ordered recruitment of GTFs to the promoters of the genes in vivo upon transcription activation [687]. An interesting implication of these studies is that promoters of inactive genes may be occupied by several GTFs prior activation and that this partially pre-assembled PIC can mark the genes for potential activation [694].

Another pathway of PIC formation is based on the findings of the large protein complexes (holoenzymes) consisting of RNAP II and a variable number of GTFs, Mediator and other proteins involved in chromatin remodeling, DNA repair and mRNA processing [138]. A common feature of the complexes was the absence of TFIID. These holoenzymes are suggested to be recruited as one entity to the promoter via interaction with various gene-specific activators in which TFIID can serve as a promoter targeting and recognition factor [694].

Open complex formation

Following the formation of preinitiation complex, the extensive interactions of RNAP II with GTFs need to be destabilized in order to release the polymerase for the elongation. Different mechanisms of the process of PIC disassembly involve extensive covalent modifications of the proteins and conformational changes.

A crucial event during initiation that is essential for RNAP II to start transcription is the formation of the transcription bubble, the specific molecular structure of local separation of DNA strands. When formed, one strand of the DNA associates with the active center of RNAP II. ATP-dependent helicase activity of XPB, a DNA-bound subunit of TFIIH, catalyzes the unwinding of 11 - 15 bp of promoter DNA around the transcription start site resulting in the formation of a transcription bubble which begins 20 bp downstream from the TBP-binding site (TATA-box) [244, 258, 264, 688, 693]. In experiments with pre-melted promoters, transcription initiation does not possess a critical requirement of an ATP and TFIIH requirement [691]. In addition, the ATP-dependent action of the XPB DNA-helicase may also function to release the tight binding of RNAP II/TFIIF complex to the DNA which can be an impediment to promoter escape [691]. This could explain the transcription stimulatory role of TFIIH during early elongation (chapter 1.3.2). Once the transcription bubble is formed at the promoter it is maintained during elongation along with the moving RNA polymerase. Phosphorylation of CTD of RNAP II is disrupting the interaction of CTD with Mediator [336] and, therefore, allows the RNA polymerase to leave the promoter and enter the elongation phase. Upon release of RNA polymerase, the PIC (entire of partial) can remain bound at the promoter and serve as a scaffold for the formation of the next transcription initiation complex in the next round of transcription reinitiation [695].
1.3.2 Elongation

Promoter escape

Transcription initiation integrates numerous events that are prerequisite for elongation, i.e. a process of RNA synthesis on DNA templates. Elongation by RNAP II is highly efficient and robust in eukaryotic cells with the rate of about 1500 nucleotides/min and is able to transcribe very long, or even extremely long, genes (up to 2000 kb for dystrophin pre-mRNA) without dissociating from DNA or releasing the nascent RNA transcript. Such a productive elongation requires many specialized factors and is dependent on the specific regulatory mechanisms. The initial process of early elongation differs significantly from late productive elongation. Transition from initiation to elongation, called promoter escape or promoter clearance, is a critical and tightly regulated stage of transcription regulation. During this stage the RNAP II undergoes dramatic structural and functional changes and interaction mode with promoter DNA, exchanges general transcription factors and loose contacts with preinitiation complex [269, 699, 700].

When the RNA transcripts are smaller than 8 - 9 nucleotides in length, the early RNAP II elongation complex is unstable and is frequently prevented from further transcription and releases the short transcript [270, 271, 704, 705]. The stability of the elongation complex is greatly increased when the length of RNA transcript reaches 9 nucleotides [704]. The early instability can possible be due steric collision of the growing RNA transcript with the B-finger of TFIIB located in the active center of RNAP II [147]. The active center of elongating RNAP II accommodates 8-basepair DNA-RNA hybrid (chapter 1.1.7), and the B-finger has to be displaced which may explain the slow rate of promoter escape, transcription pausing and polymerase backtracking during early elongation [704, 707, 708]. RNA of 9 - 10 nucleotides starts to exit from the main channel of RNAP II and interacts with specific proteins which stabilize the transcription complex and separate the RNA transcript from the DNA template [151]. Another event that occurs during promoter escape is a collapse of the upstream part of the transcription bubble when it becomes 18 nucleotides long after synthesis of 7 - 9 nucleotides of RNA transcript [258, 707]. This collapse suppresses polymerase pausing caused by the B-finger of TFIIB and stimulates formation of a stable elongation complex. The rate of promoter escape is also dependent on the sequence of the transcribed DNA as a weak RNA–DNA hybrid is inefficient in promoter escape [708].

The general transcription factor TFIIF cooperates with the ATP-dependent activity of the XPB DNA helicase of TFIIH so that they are able to suppress premature transcription stopping at the early elongation phase [268, 269, 272, 273, 714 - 716]. It is suggested that the ATP-dependent activity performs structural rearrangements in the preinitiation complex and transforms it into initiation- and promoter escape-competent conformation that results in open complex formation. The positive effect of TFIIF on transcription can be explained by its ability to increase the rate of RNAP II processivity. In addition, TFIIF can facilitate unwinding of promoter DNA wrapped around the polymerase and general initiation factors in the preinitiation complex. TFIIF makes tight contacts with double-stranded DNA upstream and downstream of the TATA box and apparently is required for the bending of the DNA [202, 683]. The XPB DNA helicase of TFIIH can increase the efficiency of promoter escape by disrupting these interactions via unwinding of downstream DNA.

During very early elongation, the CTD of RNA polymerase becomes extensively phosphorylated at serine 5 by the Cdk7 subunit of TFIIH or the Cdk8 subunit of the mediator complex (chapters 1.1.7 and 1.1.8). Following this modification, different proteins become involved in elongation via interaction with the CTD and replacing the initiation-specific factors.
Chapter 1

Although the promoter escape is a rate limiting step during the synthesis of the first 9 nucleotides, the premature elongation complex is frequently susceptible to transcription pausing and backtracking until the RNA transcript is about 23 nucleotides long [704]. At this stage transcription factor TFIIS is required for resuming elongation [719]. The RNA polymerase is stabilized via interaction of the Rpb7 subunit with the nascent RNA transcript until it is about 40 nucleotides in length [720]. Interaction of nascent RNA with CTD also can affect elongation [721].

Promoter-proximal pausing

After promoter escape, the elongation complex does not enter productive elongation stage but meets another rate-limiting process – promoter-proximal pausing at the 5' region of the gene. This transcription pause is found to be an important regulatory step for various genes in eukaryotes [699, 722, 723]. The genes, such as MYC, FOS and heat-shock proteins which are activated immediately in response to stress signals, are found to harbor an active but paused polymerase at a promoter proximal region around +20 to +40 region [709, 725, 732 - 734]. This pausing is regulated by transcription factor NELF (negative elongation factor) which dissociates upon gene activation [735]. It enables dynamic and rapid alterations in transcriptional output and makes the early pausing an efficient step of gene regulation [699].

Promoter-proximal pausing possibly functions as a mRNA quality control checkpoint prior to productive elongation. It may be required to allow completion of the capping reaction on the 5' end of a nascent RNA and assembly of pre-mRNA processing factors (chapter 1.6.1) occurring during this pause [725]. Capping enzymes are recruited via direct interactions with the Spt5 subunit of DSIF (DRB sensitivity-inducing factor) and with the CTD that is phosphorylated at serine 5 [280, 314, 315, 335, 337, 791, 797, 802]. Both DSIF and NELF were originally found to be required for inhibition of transcription by an elongation inhibitor DRB [390, 699]. DSIF is a heterodimeric complex consisting of the two subunits Spt4 and Spt5 that are conserved among yeast and human [727]; NELF consists of five conserved subunits [728]. DSIF stably interacts directly with RNA polymerase II and NELF binds to RNA and to the DSIF-RNAP II complex and cooperates with DSIF to repress transcription elongation [315, 316, 729, 730, 805 - 808]. Completion of the capping might be a signal for releasing from the pause and for resuming elongation as the capping complex is able to counteract NELF [724, 809].

The transcription pause and release of RNAP II to the stage of progressive elongation is also dependent on the function of several other transcription factors which counteract the negative effects of DSIF and NELF. The P-TEFb (positive transcription-elongation factor-b) complex consists of the protein kinase CDK9 and either cyclin T1, cyclin T2a, cyclin T2b, or cyclin K [391]. P-TEFb is recruited to the promoters of active genes containing a paused polymerase, and phosphorylates DSIF, NELF and the CTD of RNAP II specifically at Ser2 that facilitates release from the pause, activates transcription and coordinates pre-mRNA processing [343, 390, 699, 737 - 738]. The activity of the transcription factor TFIIS is critical for efficient resuming of RNAP II elongation from the pause [740]. The transcription pause is accompanied with the backward movement, or backtracking, of RNAP II and transcription bubble along the DNA template for several nucleotides [701, 730]. This leads to misalignment of the 3'-OH end of the RNA transcript within the RNAP II catalytic site and hence this end cannot be further extended by addition of another ribonucleotide. Direct and extensive interaction of TFIIS with RNAP II induces conformational changes and stimulates the intrinsic endonucleolytic RNA-cleavage activity of the polymerase active center that results in creation of a
new 3’-OH terminus that is correctly aligned to the DNA template and enables RNAP II to enter the productive elongation phase of transcription [719, 730, 740].

**Productive elongation**

After release from the promoter proximal pause, DSIF remains associated with RNAP II also during productive elongation where it has a positive role in preventing premature termination and pausing [701, 744, 745]. DSIF also associates with elongation factors TFIIF, TFIIS, and CSB and a number of chromatin regulation factors [701]. TFIIS and P-TEFb also associate with RNAP II during elongation and are present over the transcribed region of a gene [213]. They assist RNAP II to overcome transcription blocks and to pass through barriers such as specific DNA sequences and DNA-bound proteins [699 - 701, 719]. Although mainly involved in transcription initiation and efficient escape of RNA polymerase II from the promoter, TFIIF can stimulate efficient elongation via suppressing transient pausing of RNAP II, release from elongation pauses as well as potently activating the rate of nucleotide addition [205, 206, 209 - 211, 240, and chapter 1.1.4). TFIIF was found to associate within transcribed regions with sites that likely correspond to the places of paused polymerase [212, 213]. The elongation factors ELL and Elongin are associated with the elongation complex at the productive stage and function to increase RNAP II processivity and suppress pausing [730]. The CSB factor is a DNA-dependent ATPase that associates with the elongation complex plays a role in transcription-coupled DNA repair (TCR) that is required to resume elongation of the RNAP II complex when it is stalled at sites of DNA damage (chapter 1.6.5).

Several factors have been identified to function in maintenance of a specific chromatin structure and facilitation of elongation through the nucleosome environment of chromatin. ATP-dependent chromatin remodeling factors from ISWI, SNF2 and CHD families (chapter 1.5.2) are associated with transcribed regions of active genes and regulate nucleosome movement along DNA [747 - 751]. The factors FACT and Spt6 are found over the coding regions of genes and function in nucleosome disassembly-reassembly [752 - 754]. FACT destabilizes the nucleosomes via removal of H2A–H2B dimer [755]. In addition, it interacts with DSIF and can function together with P-TEFb to release RNAP II from the promoter proximal pausing [730]. Spt6 interacts with histones H3 and H4 and is involved in maintaining the chromatin structure during elongation and also stimulates the transcription rate of RNA polymerase [753, 756, 760]. Spt6 interacts with RNAP II and other elongation factors such as DSIF, TFIIS and FACT and may be involved in quality control of a pre-mRNA via recruitment of the exosome [701, and chapter 1.6.4].

The PAF complex is associated with RNAP II during elongation and serves to recruit factors with different enzymatic activities such as set1 and set2 proteins which methylate lysines of histone H3 at positions 4, 36 and 79 within the transcribed region. Histone methylation is also important in transcription regulation via modulation of the association of various transcription factors with chromatin. In addition, recruitment of Rad6 results in histone H2B monoubiquitination that is required for H3-K4 and H3-K79 methylation [701, 758], and for the recruitment to the chromatin of the proteasome that degrade proteins conjugated to polyubiquitin chains but is also reporter to be required for efficient transcription elongation (19S regulatory particle of the proteasome) [759, 760]. A number of the pre-mRNA processing events associated with the stage of productive elongation are described in chapter 1.6.
1.3.3 Termination

Termination of transcription by RNA polymerase II is an essential event that protects downstream sequences from being transcribed which can lead to impairing effects on the cellular processes. Termination does not occur at specific sites but rather randomly between 0.5 and 2 kb downstream of the 3’ end of the pre-mRNA, and is regulated by unique mechanisms. This continued processivity possibly reflects a highly robust transcription by RNAP II ensuring that long genes will be efficiently transcribed through the termination-like sequences without premature failure of an elongation. Most of the RNAP II transcribed mammalian genes have a conserved sequence element (AAUAAA) and G/U-rich sequence at the end of the transcribed part. These sequences are recognized by multisubunit factors that are involved in cleavage and polyadenylation of a 3’ end of pre-mRNA (chapter 1.6.3) and also are required for termination [342, 761 - 763, 769].

Two different models describe termination process. The ‘torpedo model’ suggests that after cleavage of the mRNA at the poly(A) site, the 5’→3’ exonuclease can bind the exposed 5’ end of the nascent RNA transcript. This exonuclease degrades the RNA until it catches up the elongation complex and triggers its destabilization that is followed by dissociation of the polymerase from DNA and the actual transcription termination [761, 762]. This mechanism is based on observations in yeast where several factors involved in the RNA cleavage reaction but not in polyadenylation are required for termination [763, 764]. The 5’→3’ RNA exonucleases Rat1 in yeast and Xrn2 in human are essential for termination because their inactivation together with some other exonucleases does not significantly affect cleavage or polyadenylation but results in a termination defect and accumulation of the RNA transcripts downstream of the poly(A) site [765, 766, 771, 772]. The Rat1/Xrn2 exonuclease digests the nascent downstream RNA cotranscriptionally but is not a dedicated termination factor. It is recruited to the places of 3’ pre-mRNA processing via Rat1-associated protein Rtt103, which interacts with the cleavage-polyadenylation factor CPF (chapter 1.6.3). Furthermore, Rat1/Xrn2 exonuclease itself facilitates recruitment and function of cleavage/polyadenylation factors.

The ‘allosteric model’ proposes that transcription over the RNA processing elements at the 3’ end of a gene triggers conformational changes of the elongation complex and its destabilization resulting in transcription termination. This model is supported by the finding that termination can take place without prior cleavage although it is dependent on the presence of the poly(A) signal [778, 779, 785]. The CTD of RNAP II is required for efficient RNA cleavage and termination [773]. It can function via cotranscriptional recruitment of specific factors involved in cleavage and polyadenylation [774 - 777]. These factors can also directly interact with RNA and recognize the poly(A) signal in the RNA. Therefore transcription over this signal can trigger a negative effect on RNA polymerase and the entire elongation complex resulting in transcription termination independently on cleavage/polyadenylation [771]. It is also possible that cleavage and polyadenylation factors can transmit quality control signals such as successful cleavage and processing of nascent mRNA. In addition, transcription termination can be triggered by dissociation of specific anti-termination and anti-pausing factors from the elongation complex [781, 782].

A specific chromatin structure can apparently also be an important regulatory determinant of termination. Chromatin-remodeling enzyme Chd1p that is involved in elongation [783] is required to form a specific nucleosome structure at the termination region of some yeast genes [750]. A spatial conformation of the transcribed chromatin possibly may also have an important regulatory role (chapter 1.6.6).
1.4 Transcription regulation by activators
The high complexity of the molecular mechanisms of gene-regulation is underlined by the fact that the tens of thousands of protein coding genes follow their individual regulatory programs which are coordinated with each other and responding to various internal and external signals during cell life, development and differentiation. Taking into account that the total number of genes can exceed hundred of thousands [561, 946, 947] the regulatory mechanisms may be highly complex and unattainable for understanding. Gene-specific regulation at the transcription level is considered to be regulated by specialized proteins that bind to specific DNA sequence elements either in a proximal promoter or in the distal transcriptional regulatory regions such as enhancers, locus control regions (LCRs) and silencers. These proteins function via interaction with general transcription factors, RNAP II and various co-regulators that possess different activities. Among these activities are remodeling of nucleosomes within chromatin, enzymatic activities that covalently modify histones and transcription factors, such as acetylation, phosphorylation, methylation, ubiquitination, ADP-ribosylation and their reverse activities. The diversity is also reflected on genomic level about 5 - 10% of the protein coding genes in mammalian genomes are dedicated to transcription regulation [514].

1.4.1 Gene-specific activators
Transcription activators have a modular structure and typically consist of two major domains: one domain that binds to the specific DNA sequences and another domain that binds to transcription (co)factors and/or regulates their activity [138, 949]. The DNA-binding domains are structurally conserved and can be separated to several classes based on their molecular structure, such as helix-turn-helix (e.g. homeodomain), different types of zinc finger, leucine zipper, helix-loop-helix and HMG domain [950]. The transcription activation domains have significant structural diversity and depending on their amino acid composition can be divided into acidic, glutamine-rich, proline-rich and hydrophobic β sheets. The thousands of gene-specific activators can be sorted into different families based on similarity in their DNA binding specificity but may yet have distinct activation functions, such as SP1, AP1, p53, nuclear receptors, and other families.

The binding of a particular activator to DNA is generally not highly specific both in vitro and in vivo and a degenerated sequence motif can be recognized with similar affinity [971]. A high degree of specificity essential for precise control of gene transcription is often achieved by combination of recognition sites for multiple factor in a genomic region in the promoter (proximal promoter region at about -50 to -200 bp relative to the transcription start site) or with distal location from the target gene (enhancers, LCRs). These clusters contain usually unique combination of several short (6 - 8 bp) DNA recognition sites which rarely can be found in other genomic region although a single recognition site can be found in many places. Multiple activators accommodated at these clusters can function synergistically and activate transcription much stronger than a single factor alone [972, 973]. Such an organization of multiple activators in a single complex allows low concentrations of activator proteins to regulate a large range of transcription activation and provides a capacity to integrate multiple regulatory signals into a single output.

1.4.2 Regulation of activator function
Activators are crucial regulators which maintain a tightly controlled gene expression as an essential prerequisite of normal cellular life and development. The gene-specific activators are subjected to a tight control via various mechanisms. Deregulation of activators can result in severe cellular dysfunctions such as apoptosis or, in opposite, malignant transformation.
The regulation of activator function often occurs through specialized domains. For example, nuclear receptors contain a binding domain for specific ligands, i.e. small molecule such as hormones, and are able to function as activators only upon binding of the cognate ligand. In other cases, the regulatory modules can be separate proteins [969]. For example IκB functions as a detachable regulatory subunit that modulates the activity and cellular location of the transcription factor NF-κB [970].

Posttranslational modifications in their combinatorial diversity provide another regulatory dimension. For example, phosphorylation of CREB (cyclic AMP response element binding protein) by protein kinase A turns it to an active form which is able to activate target genes [951]. Acetylation has also an important regulatory role. For example in case of p53 it increases the binding affinity to DNA [954]. A number of factors such as Sp1, AP-1, AP-2, CTF/NF-I, and Zeste carry O-linked N-acetylgalactosamine (GlcNAc) monosaccharide residues which might have regulatory implication [949, 952, 953].

Polyubiquitination (an addition of a chain of ubiquitin residues to the protein lysines) has been implicated in degradation of the proteins by the proteasome [955, 956, 963]. It was noted that the most potent activators expressed in cells are the least stable and more efficiently degraded via polyubiquitination [964]. Monoubiquitination does not lead to protein degradation but functions to enhance the activity of transcription factors [965]. Furthermore, in combination with many deubiquitinating enzymes which function to remove ubiquitin from the target proteins, the ubiquitin system can serve as an additional system in regulation of the stability and function of different transcription factors.

It is important to note that many general transcription factors or their subunits (chapter 1.1) also carry a number of enzymatic activities which can work in positive or negative feedbacks in response to activators. An example of such a mechanism investigated in details [957] shows that upon RNAP II gene initiation the Mediator-associated kinase phosphorylates the transcription activator GCN4 that marks this activator for subsequent ubiquitination and degradation. This ability of general transcription factors to mark activators for destruction provides an efficient mechanism to limit the time window in which a DNA-bound activator can activate transcription. Similar mechanisms have been suggested also for other potent activators, such as E2F-1, Myc, Jun, Fos, and p53 [958 - 962].

Covalent modification also can regulate subcellular location of transcription factors so that they are imported to the nucleus or exported to the cytoplasm in response to regulatory stimuli. This mode of regulation is documented for activator NFκB which is kept outside the nucleus by its interaction inhibitor IκB. NFκB dissociates after phosphorylation of IκB upon specific signals and become accessible for nuclear transport.

Covalent linkage of the small ubiquitin-related modifier (SUMO), a 101 amino acid polypeptide, to the proteins is also implicated in regulation of subnuclear localization as well as stabilization of transcription factors [955, 966]. Addition of SUMO leads to transport of the target proteins into the nuclear bodies where they are stored until when certain physiological stimuli induce SUMO-1 proteases and thereby nuclear redistribution of the proteins. Such a regulation was observed for transcription factors PML, LEF1, Sp3 [966 - 969].

Gene-specific negative transcription regulation can also be achieved by repressors which can bind to an activation domain and suppress its function or compete for the binding to a specific DNA site [138]. For example, the chaperone Hsp90 binds to the heat shock transcription activator Hsf1, or the yeast activator Gal4 can be inactivated by repressor Gal80 protein that binds to the activation domain of Gal4 [974]. In general, the level of transcription of a cellular gene can be determined by a balance function of positive and negative regulators.
1.4.3 Transcription activation by enhancers

Enhancers are usually composed of several specific DNA binding sites for different transcriptional activators [975]. Enhancers function to activate transcription of their target genes upon certain stimuli, e.g. at a particular stage in development or differentiation in a specific tissue. A unique characteristic of enhancers is that they can target a gene located at a long distance in the genome (up to >100 kb upstream or downstream from the promoter) and function independently of their sequence orientation. Sequence-specific activators and architectural proteins organize an enhanceosome when they are assembled at enhancer [138]. Enhanceosome can be spatially positioned at close proximity to a target promoter via DNA looping involving extensive protein-protein and protein-DNA interaction and modification of the chromatin structure [977]. Selectivity for a particular target promoter is suggested to be dependent on the specific interactions between enhanceosome and promoter-specific factors, and also on boundary elements (insulators) which can block unwanted interactions of the enhancer with unrelated promoters [976]. Being in direct contact with the target promoter, the enhanceosome can facilitate the assembly of a transcription initiation complex at the core promoter via direct recruitment of general transcription factors, RNAP II, and coactivators that can function in providing specific chromatin modifications. Another possible mechanism is that the general factors and RNAP II can be assembled at enhancers and upon specific stimuli can be relocated to promoters [978].

The Locus Control Regions (LCRs) are similar to enhancers in that they consist of multiple activator binding sites [979]. LCRs are able to activate transcription independently of the genomic context whereas the function of enhancers can be suppressed by the inactive state of the chromatin. Also the functioning of LCRs is dependent on orientation and is limited in distance from a target gene.

1.5 Transcription coactivators function

1.5.1 General transcription factors

Extraordinary diversity in regulation of the gene transcription is provided by a coordinated function of a large number of gene-specific activators and different repression mechanisms in combination with diversified general transcription factors and core-promoters (chapters 1.1.1 and 1.2.1). General transcription factors and RNAP II can function as coactivators and directly interact and receive the activation signals from activators that in this way can enhance an assembly of the preinitiation complex (chapter 1.1). In addition, activators can affect conformation of PIC. Extensive interactions with activators were characterized for TAFs subunits of the multiprotein complexes TFIID and Mediator therefore revealing their co-activator function [694, 981]. Mediator is considered as a molecular bridge for transmission of signals from the gene-specific activators to the PIC and RNA polymerase (chapter 1.1.8). Interaction with TAFs also indicates that activators can recruit TBP-free TAF-complexes at the specific genes. In addition, activators interact with a large number of various proteins which affect the PIC via different mechanisms serving in stabilization and scaffold formation, anti-repression, providing various covalent modifications of transcription factors and chromatin [633].

1.5.2 Chromatin remodeling

A large group of coactivators direct their activities towards chromatin which is an essential regulator in of transcription. These cofactors include chromatin-remodeling complexes and a variety of histone-modifying enzymes. They appear to be primarily recruited by activators or repressors bound to
specific DNA sites to modify chromatin structure and provide the appropriate environment for assembly of the preinitiation complex or change the chromatin into compact repressive state. Chromatin-remodeling factors utilize ATP hydrolysis to catalyze the mobilization and repositioning of nucleosomes via disrupting and displacement of the histone-DNA contacts [751]. This activity is necessary for altering the condensed repressive chromatin structures into an open state that is accessible for transcription factors. There are many chromatin-remodeling complexes but most of them contain a common subunit that belongs to Snf2-like family of ATPases. These complexes were sorted to 7 subfamilies: SNF2, ISWI, CHD1, INO80, CSB, RAD54, and DDM1 [751]. Most of the complexes consist of several subunits which are required for regulation of the ATPase subunit and also possess other functions.

A well studied member of SNF2 subfamily is a SWI/SNF, a large complex of about 2 MDa and 12 subunits conserved between yeast and human [984]. This complex is recruited by activators, e.g. nuclear receptors, and is required for transcription of about 5% of yeast genes. Another complex within the SNF2 subfamily is RSC which is closely related to SWI/SNF but appeared to have more general requirements in cellular transcription [985, 986]. The members of the ISWI subfamily include complexes such as NURF, RSF and NoRC [987 - 989]. They function to assemble nucleosomes or to enhance the stability of chromatin structure in contrast to SWI/SNF which disrupts nucleosome structure. RSF and NURF increase activated transcription [990] whereas NoRC has been implicated in epigenetic silencing [991]. The NuRD and Mi-2 complexes are members of CHD1 subfamily containing chromodomain and DNA-binding motifs [992 - 994]. These complexes possess a histone deacetylase (HDAC) subunit suggesting a cooperative function of chromatin remodeling and deacetylation of histones in transcription repression. Members of the other families also contain subunits with distinct functions and are required for different processes such as homologous DNA recombination, DNA repair, DNA methylation and maintenance of the genome stability.

1.5.3 Covalent histone modifications

Covalent histone modifications appear to play a crucial role in transcription by altering the chromatin environment to make it suitable either for active transcription or repression. Well studied modifications include acetylation, phosphorylation, methylation, ubiquitynation, and ADP-ribosylation. There are many proteins known to carry these enzymatic activities or that can reverse these modifications which can be recruited to the chromatin as co-activators/co-repressors via association with specific transcription activator/repressor complexes. The dynamic functioning of the chromatin modifying enzymes provides an appropriate combination of histone modifications regulating a gene transcription in response to different signals. Different modifications can function either in cooperative or antagonizing mode via affecting their recognition by the specific proteins or regulating the activities of histone modification enzymes.

**Acetylation** usually occurs at multiple lysine residues and has been correlated with actively transcribed genes. It is mainly located at promoter regions [1003] whereas transcriptionally-silent heterochromatin lacks acetylated histones [995, 1002]. Histone acetyltransferases (HATs) and deacetylases (HDACs) are found to associate with many coactivator and corepressor complexes, respectively. Subunits of the general factors TFIID and SAGA are able to acetylate histones (chapter 1.1). In addition, many other proteins have HAT activity, for example, CREB-binding protein (CBP) and the related protein p300 which appear to be generally required for transcription of RNAP II genes [561], and the p160 family of coactivators targeted by nuclear receptors [996 - 998].
In general, HAT activities have substrate selectivity such as particular histones or lysine residues, (e.g. PCAF and hGCN5 acetylate only histone H3). Acetylation might regulate higher order chromatin architecture by decreasing charge-dependent compaction of histone/DNA fiber [1005]. Alternatively, histone acetylation can regulate transcription in response to certain signal transduction pathways or increase regulatory diversity in combination with other histone modifications. For example, acetylation of lysine 16 on histone H4 (H4K16Ac) is involved in chromatin decondensation [1005] whereas its deacetylation by SirT2 may induce the condensation of the chromatin during metaphase transition [1007]. Acetylated histones can also serve as recognition sites for proteins with bromodomains and therefore can direct their recruitment; bromodomains are present in many transcription factors including the TAF1 subunit of TFIID.

Apart from transcription, histone acetylation is involved in regulation of other processes. The HBO1 acetyltransferase that acetylates histone H4 is a central factor in DNA replication and functions in organizing the pre-replicative complex at the origins of replication and is required for the initiation of S phase [1008, 1009]. Acetylation of lysine 56 at histone H3 (H3K56) in yeats by Rtt109 enzyme and lysine 12 at histone H4 (H4K12) has been implicated in DNA repair [1010-1013].

**Methylation** of the histones can occur at the arginine or lysine residues. Histone methylases (HMTs) are recruited to chromatin as co-factors and, unlike HATs, modify precisely a single residue at a specific position on a particular histone [999]. Methylation marks can be recognized by the chromo-like domains and PHD domains and therefore can serve to recruit these specific proteins. The arginine methylation is mediated by recruitment of protein arginine methyltransferases (PRMTs) which can modify several residues on histones H3 and H4 [1000, 1001]. These modifications correlate with transcription activation and can be reversed via conversion to citrullines by the PADI4 enzyme that, therefore, antagonizes the positive effects of arginine methylation [1004].

Methylation on different lysine residues is the most extensively studied so far and is found to be associated with actively transcribed genes as well as with transcriptionally silent loci [1000]. Methylation on lysines 4, 36, and 79 of the histone H3 (H3K4, H3K36, and H3K79) is implicated in transcription activation. The H3K4 trimethylation (H3K4me3) mark is localized at the promoters of the active genes [645]. H3K4me3 can be established by MLL1, MLL2 and Set1 methyltransferase complexes that are recruited to the chromatin via association of their common subunit WDR5 with mono- and di-methylated lysine 4 on histone H3 (H3K4me1 and H3K4me2) [1019]. A NURF chromatin-remodeling complex can recognize H3K4me3 via a PHD domain of BPTF subunit and implicates in transcription activation [1018]. Upon DNA-damaging stress transcription of the active genes can be suppressed by recruitment of the repressive Sin3a HDAC complex via recognition of the H3K4me3 by the PHD-finger containing subunit ING2 [1021, 1022]. In addition, the demethylation and acetylation activities can be linked to the chromatin via binding of the JMJD2A and CHD1 proteins to the H3K4me3 [629, 1035].

The H3K36me3 is accumulating at the transcribed region of active genes and functions in suppression of transcription initiation at inappropriate sites within the coding region [1015-1017]. Mechanistically, H3K36me3 is recognized by the EAF3 protein which is a part of the HDAC complex Rpd35 that deacetylates the histones and therefore prevents initiation factors to assemble on the gene-internal initiation sites.

Methylation at H3K9, H3K27, and H4K20 is correlated with transcriptional repression. Methylated H3K9 is recognized by the HP1 protein that results in formation of condensed heterochromatin and is accompanied by recruitment of the repression proteins [1024, 1025]. This modification is a characteristic of centromeric heterochromatin. In addition, H3K9 methylation and HP1 were found
within a coding region of the active genes suggesting a dual role for this histone mark [1023]. H3K27 methylation is required for programmed transcription repression of a large number of different genes including developmental HOX genes, and also for inactivation of the X chromosome during genomic imprinting. This mark is connected with polycomb group (PcG) proteins which function in chromatin condensation and also transcription repression that do not involve chromatin compaction [1026]. The active and repressive chromatin methylation marks, H3K4me and H3K27me, can be present at the same time on specific genomic regions in the embryonic cells [1039]. These bivalent marks correspond to low-level expression of the developmental transcription factors. After cell differentiation these modifications diverge and only one of them remains to either activate or to repress the corresponding genes. The repressive function of the H4K20 methylation is implicated in the formation of heterochromatin and is required for chromosome condensation [1027]. The mono- and dimethylated H4K20 are present at sites of DNA repair and are required, along with phosphorylation of H2AX, to create a recognition sites and recruit the cell-cycle checkpoint protein Crb2 to the sites of DNA damage [1040 - 1042].

The removal of methyl groups from lysines is performed by two types of histone demethylases. The LSD1-demethylase reverses the H3K4 methylation that consequently represses a transcription [1028]. In complex with androgen receptor, LSD1 demethylates H3K9 that results in transcription activation [1029]. There are several JmjC-domain demethylases which can reverse H3K9 and H3K36 methylation [1030 - 1034, 1037]. The demethylases are selective to certain sites and are important players in dynamic maintenance of the modifications that play a crucial role in transcription regulation [1038].

Phosphorylation is provided by various protein kinases and can target the serine (S) and threonine (T) residues [1050]. Phosphorylation on H3S10 is involved in transcription activation of the NFκB-regulated genes and the early response genes such as Fos and Jun [1000]. The H3S10 phosphorylation also functions in chromatin condensation during mitosis regulating the binding of HP1 to methylated H3K9 [1024]. This modification can be recognized by a specific domain in 14-3-3 signal transduction proteins [1044, 1045]. Upon DNA damage, phosphorylation of the histone variant H2AX takes place over long (several kb) genomic regions around the sites of DNA damage and is required for recruitment of the chromatin remodeling INO80 complex [1046, 1047]. The phosphorylation of histone H2A at S129 and T3 and histone H4 at S1 is involved in the double-strand break repair and non-homologous end joining [1048, 1051]. Although little functional data is available, recent finding revealed that many specific kinases can be recruited to the specific genes upon certain specific stimuli [1043]. This suggests that the signal transduction cascades directly influence transcription via chromatin regulation.

Monoubiquitination of histone H2B (H2Bub1) correlates with active transcription and is localized at promoters as well as within transcribed regions of the genes [1052 - 1054]. This modification enhances the elongation rate via cooperation with the FACT elongation factor (chapter 1.3.2) and is also required for the methylation of H3K4 [1055, 1056]. In contrast, ubiquitination of K119 at histone H2A correlates with transcription repression [1057]. Ubiquitination of the histones H3 and H4 by CUL4-DDB-Roc1 complex has been implicated in DNA repair via regulation of recruitment of the XPC repair protein to the sites of DNA damage [1058, 1059].

1.6 Co-transcriptional regulation of different cellular processes

The process of precise translation of genetic information to the proteins requires a coordinated action of different mechanisms responsible for a proper synthesis of mRNA and its delivery to the
cytoplasm. In this way transcription, mRNA processing and DNA repair appear to be closely connected and co-regulated. The mRNA processing steps comprise of an addition of a cap-structure at the 5’ end and poly(A)-tail at the 3’end of mRNA, and excision of introns from the pre-mRNA. Processed mRNA molecules can be exported from the nucleus via systems of nuclear transport. Although all these processes are biochemically isolated, they appear to be tightly linked and efficiently coordinated. Moreover, significant evidence accumulated suggesting that transcription is responsible for this coordination. The current view is that much integration and coordination occurs via the RNAP II C-terminal domain (CTD). Indeed, CTD truncation causes defects in capping, splicing and cleavage/polyadenylation [773]. The CTD carries different modifications and it can interact with proteins required at all stages of transcription, from initiation to termination [337, 787]. There are positive and negative feedbacks so that one reaction can enhance the next or, lead to repression to prevent the further usage of wrongly processed RNA molecules [909, 910]. In addition to functional interconnection between the processes a physical connection between the factors has been found to exist in the large mega-complexes named ‘transcription factories’ [335].

1.6.1 Capping of the 5’ end of mRNA
When RNAP II escapes from the promoter and transcription is being initiated (chapter 1.3.2) the first processing event is the addition of the cap-structure at the 5’ end of pre-mRNA to provide resistance to 5’-3’ RNA exonucleases [792]. Three enzymatic activities are involved [790]. First, RNA triphosphatase hydrolyzes the phosphate of the first pre-mRNA nucleotide to a diphosphate. Second, RNA guanylyltransferase catalyzes the fusion of a GMP moiety from GTP to the first nucleotide of the pre-mRNA via a 5′-5′ triphosphate linkage. Finally, RNA (guanine-7-) methyltransferase adds a methyl group to the N7 position of the transferred GMP to form the m7G(5')ppp(5')N cap. The enzymes from the first and second steps are combined in a single protein in mammalian cells. After addition, the cap structure is then recognized by the cap binding complex (CBC) and upon export from nucleus to cytoplasm this complex is exchanged by translation initiation factor eIF-4E [791, 793, 802] that facilitates efficient recognition of mRNA by ribosomal complexes. The cap-modification occurs during early elongation when RNA transcripts are about 25 - 30 nucleotides in lengths [788, 789]. At this stage the specific factors DSIF and the negative elongation factor NELF are recruited to the transcription complex to suppress transition to the productive elongation stage [729, 794 - 795]. This pausing is possibly required for completion of capping on the nascent RNA transcripts. The capping enzymes enter a transcription complex via direct interaction with CTD phosphorylated at serine 5 [280, 314, 315, 335, 337, 791, 797, 802] that is modified during early elongation by several enzymes such as CDK7 of the TFIIF complex (chapter 1.1.6). In addition, triphosphatase and guanylyltransferase interact with Spt5, a subunit of DSIF [316, 727, 745, 805]. Upon binding to transcriptional complex, the capping activity is stimulated by both phosphorylated CTD and Spt5 [316, 429, 805 - 806, 808]. In addition, the capping complex counteracts NELF that facilitates promoter escape to resume elongation [809]. In general, mRNA capping is a critical event that affects downstream steps such as RNA stability, splicing, transport, and translation [794, 810].

1.6.2 Splicing
The majority of mammalian genes have a specific structural organization where protein-coding sequences (exons) are interrupted by non-coding sequences (introns). An average gene contains 9 introns of about 3000 bp and 10 exons of about 200 bp [815]. Such an organization may provide additional possibilities to control cellular processes since genes in the simpler organisms often lack
introns. For example, in yeast only ~3% of the genes contain introns, but, remarkably, these genes are most active and comprise ~30% of the total primary RNA pool [816, 817].

Introns must be removed from the primary transcribed pre-mRNA in order to generate functional mRNA that is suitable for nuclear export and translation into proteins [818]. The process of intron removal (splicing) takes place in large macromolecular structures named spliceosomes. They consist of small nuclear ribonucleoprotein particles (snRNPs) associated with five snRNA species U1, U2, U4, U5, U6, and many other proteins [819 - 821]. The transcribed pre-mRNA contains specific elements essential for the splicing reaction: the canonical splice sites at the exon-intron junctions comprise consensus sequences AG-GURAGU at the 5′ and YAG-RNNN at the 3′ end of the introns (R, purine; Y, pyrimidine) and at ~100 nt upstream of the 3′ splice site there is the branchpoint CURAY [337, 822]. These elements are recognized by spliceosomal snRNPs. In addition, there are exonic and intronic splicing enhancer and silencer elements which regulate splicing at the adjacent 5′ and 3′ sites. They are recognized by the proteins of serine/arginine-rich protein family and also by heterogeneous proteins bound to the pre-mRNA molecules (hnRNP) [823, 824].

In a stepwise model, the initial step of spliceosome assembly starts from recognition of the 5′ and 3′ splice sites by U1 snRNP and the splicing factor U2AF, accordingly. These events lead to recruitment of a branchpoint binding protein (BBP) and U2 snRNP to the branchpoint site on the pre-mRNA. Further, U5 and U6 enter the spliceosome and displace U1 snRNP. At this step U2 and U6 interact directly with 3′ and 5′ splice sites at the pre-mRNA and also with each other via base-pair homology. The assembly of the U2, U5 and U6 snRNPs together with other protein components results in formation of a catalytic spliceosome entity and positioning of the 5′ and 3′ splice sites in close proximity to each other [825 - 828]. Alternative models propose an assembly of the spliceosome from large complexes, such as penta-snRNP which contains the U1, U2, U4, U5 and U6 snRNAs [831] or the 200S lnRNP (large nuclear ribonucleoprotein particle), which contains other proteins in addition to U snRNP [832].

Regulation of splicing is functionally linked to transcription. Initial findings revealed that transcribed pre-mRNA is spliced cotranscriptionally since the splicing factors are localized at sites of active transcription [833, 835, 836]. A number of other cytological and biochemical studies further suggested cotranscriptional pre-mRNA processing and functional interactions between RNAP II and the splicing complex [797, 837, 841]. The phosphorylated CTD can directly activate splicing and the IIO form of RNAP II is able to significantly increase formation of spliceosomal complexes [836, 844, 845, 856]. In addition, increased association of splicing factors is observed at the sites of active RNAP II transcription only with intact CTD (not truncated) [839] and over-expression of CTD-containing RPB1 subunit in mammalian cells induces nuclear reorganization of splicing factors [340]. Furthermore, the conformational changes in CTD induced by the proline isomerase Pin1 result in splicing inhibition in vitro [431].

A number of transcription factors associated with spliceosome have been found by comprehensive proteomic analysis [842] suggesting a tight links between transcription and splicing. Several multiprotein complexes were found to contain splicing proteins together with transcription factors such as general elongation factors P-TEFb, TAT-SF1 and TFIIF [843, 846 - 849]. These factors can interact with RNAP II and, therefore, can provide functional links between splicing and transcription elongation. Indeed, interaction of U-snRNPs with TAT-SF1 strongly stimulates elongation [850]. Splicing factors can affect elongation because the presence of a 5′-splice site downstream of a promoter significantly enhances the transcription rate [851]. One explanatory model suggests that splicing factors can associate with both pre-mRNA transcript and components of transcription.
General introduction

elongation and, possibly, also transcription initiation because the U1 snRNA can interact with the general transcription factor TFIH [279, 319]. Additionally, yeast U1 snRNP is recruited to intron-containing genes during transcription [870].

Connection is found between RNAP II elongation processivity and alternative splicing. The factors affecting elongation rate, such as mutations in RNAP II, specific inhibitors [853 - 855] or DNA sites that cause polymerase pausing [852], result in reduction of the skipping of exons during splicing [337, 852]. This effect is possibly due to a shift in kinetic balance between the rate of splicing reactions and the rate of appearance of additional 3’ splice sites during pre-mRNA synthesis.

Gene-specific regulation of splicing can occur in connection with promoter structure although the mechanisms of this regulation are not well understood [857]. It is possible that the strength of a promoter can affect elongation rate that can influence alternative splicing [858]. Taking into account abundance of tissue-specific alternatively spliced genes the actual mechanisms can vary depending on promoter structure and the presence of transcription regulatory elements occupied by specific transcription factors [859 - 863]. In line with this suggestion, different transcription activators are able to enhance elongation rate and also provide specific effects [862, 863]. The rate of splicing is positively affected by class IIB activators (e.g. VP16) that stimulate both transcription initiation and elongation, but little effect is found from the class I and IIA activators [865], which only stimulate transcriptional initiation or require synergistic components for their function [866]. Transcriptional coregulators and coactivators can also have implication in gene-specific splicing depending on promoter structure [867 - 869]. Specific histone modifications at promoter regions that are required for active transcription, such as acetylation, may have a positive impact on splicing [866].

Capping reaction is an additional regulator of splicing. The cap structure significantly enhances the formation of a spliced mRNA and the excision of the first cap-proximal intron. One possible mechanism suggests that the cap binding complex positively affects interactions of U6 snRNA with 5’ splice site and facilitates displacement of U1 snRNP [829]. Interaction with hnRNP F may also have an effect on splicing [830].

1.6.3 Processing of the 3’ end of mRNA

The formation of the 3’ end of transcribed RNA occurs at specific sites and in several steps. Most of the RNAP II transcribed genes contain at their 3’ ends a conserved poly(A) site (sequence AAUAAA) and a G/U-rich sequence located 10–30 nucleotides upstream and downstream around the pre-mRNA cleavage site. These sites are specifically bound by multisubunit cleavage and polyadenylation specificity factor (CPSF) and by cleavage stimulatory factor (CstF) [344, 837, 871, 872, 876]. Together with cleavage factors I and II (CFI and CFII) [874, 875, 878] they are responsible for endonucleolytic cleavage of the transcribed RNA at a site located between the poly(A) signal and G/U-rich region. CFI, in addition, is implicated as a regulator of poly(A) site selection that may selectively suppress inappropriate poly(A) site usage [879, 880]. Following the cleavage of a nascent pre-mRNA, the poly(A) polymerase (PAP), in complex with CPSF, adds a poly(A) tail (a uniform 3’ end consisting of about 200 adenosine residues) to the 3’ OH end of RNA [881]. Histone mRNAs, snRNAs and snoRNAs, although transcribed by RNAP II, have different mechanisms of the 3’ end formation that do not require poly(A) addition but dependent on U7 snRNP complex, cleavage factors and CTD [787, 881, 899, 900].

The binding of cleavage/polyadenylation factors is found not only at a 3’ end of a gene but also at a promoter and throughout the coding region [314, 341, 335, 884, 885]. This phenomenon can be explained by involvement of the Ser2-phosphorylated CTD in 3’ end maturation because inactivation
Chapter 1

of the Ser2/5 kinase CTK1 [888] inhibits the function of 3’ end processing factors [345]. Such a phosphorylated CTD can be specifically recognized by the yeast 3’-end mRNA processing factors Pcf11 and Rtt103 via their CTD-interacting domains [341, 765]. In addition, CPSF and CstF can bind CTD irrespectively of phosphorylation [773, 881]. Also the CTD itself is able to directly activate 3’ cleavage [774] and polyadenylation [341, 777]. In addition, the Rna15/CstF64 and a processing/termination factor Ssu72 (also known as CTD phosphatase, chapter 1.1.7) [893] interact with initiation factor TFIIB and the coactivator Sub1/PC4 [885]. CPSF can interact with TFIID during initiation and remains associated with RNAP II during elongation [797, 895].

A link has been found between 3’ processing and transcription termination. Initially, the presence of an intact poly(A) signal was considered to be essential for RNAP II termination [761, 769]. Later studies have shown that certain subunits of the cleavage/polyadenylation factors are required for RNAP II termination [342, 763] but the event of cleavage itself may not be required [778, 779]. Furthermore, the termination factor Rat1/Xrn2 (it is the 5’–3’ RNA exonuclease) binds to the CTD-bound 3’ end processing factor Rtt103 and it can degrade an exposed 5’ end of the nascent RNA after cleavage that eventually triggers the release of the polymerase from DNA [765, 766].

1.6.4 mRNA nuclear transport

After processing (capping, splicing, 3’-end processing and polyadenylation) mRNAs is assembled into ribonucleoprotein complexes (mRNP) and subjected to transport that is responsible for their export through the nuclear pore complexes to cytoplasm [911]. mRNA transport is a complex process involving an extensive network of interactions and it appears to be connected to the pre-mRNA processing steps and to transcription [818, 911]. This connection is possibly required to ensure that only fully processed and intact mRNAs can reach the cytoplasm [909, 910].

mRNA transport was found to have a number of connection lines to transcription. The mRNA export adaptor molecules (Aly/RE and ATPase/RNA helicase UAP56) associate with RNA during transcription [912] and their binding can be detected over a gene including promoter and at higher level at a 3’ end [887]. The Aly/RE interacts with tetrameric complex THO that is involved in transcription elongation and mRNA export [913 - 917]. Both Aly/RE and THO are components of the transcription/export complex (TREX) which is recruited to an active gene during transcription elongation and is proposed to directly connect transcription to mRNA export [916]. The TREX complex interacts also with the proteins involved in splicing suggesting an integral role in mRNA transcription, splicing and nuclear export [842, 918, 919, 921]. An additional link was found by discovery of the connecting protein Sac3p that interacts with components of THO, mRNP and nuclear pore complex [922, 923]. Another protein, Sus1p, interacts with both Sac3p/THO and TAF-containing histone acetylase complex SAGA (chapter 1.1.11). Sus1p is recruited to the transcription initiation sites of SAGA-dependent genes suggesting an additional tight link between transcription and mRNA transport [924]. This protein may function by facilitating interactions of Sac3p/THO with both nascent transcripts and nuclear pore complex and, therefore, it can organize SAGA-dependent genes to the nuclear periphery for an efficient access of the newly made transcripts for the nuclear export system [911].

TREX also regulates the nuclear exosome, a large complex of 3’-to-5’ exonucleases involved in RNA processing, mRNP assembly during elongation, degradation of mRNA with unprocessed 3’ ends and disassembly of improper mRNP particles [818, 911]. The complex of Aly/RE, UAP56 and TREX functions also during co-transcriptional mRNP assembly - the packaging of mRNA into the stable and exportable mRNP particles [915, 917]. The negative regulation reveals that functional errors during
the mRNP assembly and transport can suppress mRNA synthesis and/or stability [913, 926] underlining the interdependence of gene expression processes.

1.6.5 DNA repair

Every cell faces a constant flow of DNA damage originating either from endogenous (cellular metabolism) or exogenous (chemicals, ionizing radiation and UV light) factors. Various DNA repair mechanisms are crucial for genome stability and intact state of the gene sequences ensuring a proper gene expression and genome replication. There are nucleotide excision repair (NER) or base excision repair (BER) pathways responsible for the reparation of local DNA damages. DNA lesions block transcription elongation and hence RNA polymerase turns to a stalled state because it can not bypass a DNA lesion until it is repaired [927, 928]. Interestingly, RNAP II itself appeared to play an active role in DNA repair. Initially it was shown that transcribed genes are repaired much faster than non-transcribed loci and other genomic sites [930, 931]. This difference was found to be due to a specific mechanism of transcription-coupled repair (TCR) which is dependent on the presence of a stalled elongating RNAP II [927] and also transcription activators irrespectively of elongation [931, 932].

The mechanism of TCR may rely on the process of chromatin rearrangement via histone modifications and nucleosome remodeling that is associated with transcription and is required to facilitate assembly of the transcription pre-initiation complex (chapter 1.5). Several histone acetyltransferases (TFTC/SAGA, NuA4, p300/CBP, TIP60 complexes, chapter 1.5.3) and ATP-dependent chromatin remodeling factors (SWI/SNF and RAD families, chapter 1.5.2) can be found at the DNA damage sites and are implicated in efficient DNA repair [933, 934]. They possibly function by facilitating displacement of the stalled RNA polymerase complex and by increasing access of the repair factors to the places of DNA damage. In addition, they can recruit multiple factors required for backtracking of the RNA polymerase to make the DNA damage site accessible, for DNA repair and for restarting the stalled RNA polymerase.

Phosphorylation of the histone variant H2AX by specific kinases of the PIKK family (phosphatidylinositol-3 kinase-related kinase) is an early event that appears within several minutes in response to DNA damage. This modified histone can be recognized by certain factors from the DNA repair pathway [933].

Ubiquitination of RNAP II is another posttranslational modification that is associated with DNA repair. Significant level of ubiquitination can be detected when cells are stressed, particularly in response to DNA damage or to a block of RNAP II elongation by other factors. UV irradiation or peroxide treatment cause DNA damage resulting in polyubiquitination of stalled RNAP II at the sites of DNA lesions that is implicated in activation of DNA repair pathways. Serine 5 phosphorylation of CTD is, possibly, a prerequisite modification for ubiquitination [436, 439, 440, 453]. RNAP II ubiquitination is mediated by a number of different E3 ubiquitin ligases [263, 434 - 440]. Some of them, i.e. BRCA1 and CSA, associate with elongating RNAP II [438 - 440] and BRCA1 also interacts with the 3’ cleavage stimulatory factor CstF [797]. Although ubiquitination of RNAP II may lead to proteasome-mediated degradation, there are examples of nonproteolytic pathways that can be activated in response to distinct patterns of polyubiquitin chains: e.g. endocytosis [441], activation of the NFkB pathway [446, 447], enhancement of ribosome translational efficiency [443] and DNA repair [442, 444, 445]. Degradation-independent polyubiquitination pattern of RNAP II can be found in response to a transcriptional block induced by alpha-amanitin which blocks nucleoside triphosphates incorporation into the nascent transcript [448, 449].
1.6.6 Transcription factories

The transcribed genes and their corresponding genomic loci appeared to have unique spatial organization which may be important for a gene regulation. In yeast, the long genes form gene loops where promoter and termination regions are physically linked [935, 936]. The loop structures of the transcriptionally active regions are observed also in cells from different species [937]. The existence of the DNA loops and spatially close positioning of the distal functional elements such as promoters, enhancers and locus control regions (LCRs) is investigated by chromosome conformation capture (3C) studies [940 - 945].

Based on numerous data, the so called transcription factories are suggested to be a general structural feature of an active gene transcription [937 - 939]. These factories are the main sites of cellular transcription and are organized as major protein complexes attached to the nuclear fibrils and containing various transcription factors and a large number of the active RNA polymerases. In actively growing mammalian cells it is estimated that there are several thousands of these factories each of each organizes transcription of several RNAP II genes. The nascent transcripts can be visualized not as spread fibrils but as foci located compactly at the sites of transcription. The number of factories is variable and can correlate with the rate of cellular transcription. Upon transcription inhibition the number of factories is decreasing and the gene loops are disassembled that suggests that RNA polymerase is a major structural factor of the factories [937 - 939].

Organization of the cellular transcription into factories suggests many ways for gene regulation including functioning of enhancers, insulators, locus control regions and heterochromatin barriers [937 - 939]. Also it provides a basis for the known communication and cross-regulation between different stages of transcription and pre-mRNA processing.

2. Transcription of ribosomal RNA gene by RNA polymerase I

Cell growth, proliferation and different living processes require coordination of multiple processes including transcription, translation and replication. Protein synthesis in actively growing cells requires several thousands ribosomes per minute, or up to several millions per cell cycle [1060 - 1064]. Such an enormous quantity of ribosomes requires the coordination of a large number of genes and proportional amounts of building blocks: ribosomal particles are assembled from more than 70 ribosomal proteins (RP) transcribed by RNA polymerase II, and 4 ribosomal RNA species (rRNA) such as 28S, 18S, 5.8S and 5S that accounts for about 50% of cellular transcription. The first three rRNA species are transcribed by RNA polymerase I (RNAP I) as a single pre-rRNA, 45S pre-rRNA transcript and a small 5S rRNA – by RNA polymerase III. In addition, a complex network of different cellular mechanisms is required for processing of rRNAs and proteins, and formation and export of ribosomes. How such a perfect and extremely complex coordination takes place is very far from being understood.

2.1 Nucleolar sub-compartments and their functions

Synthesis of 28S, 18S, 5.8S ribosomal RNA by RNAP I and biogenesis of ribosomes take place in the nucleolus, a prominent non-membrane nuclear compartment in eukaryotic cells. The human diploid genome contains about 400 copies of ribosomal genes which are organized in tandem repeats and distributed among the short arms of the acrocentric chromosomes 13, 14, 15, 21 and 22 [1070]. These gene clusters are named nucleolus organizer regions (NORs) as they appear as distinct structures on metaphase chromosomes during mitosis. Three main nucleolar sub-compartments can be observed as
morphologically distinct areas in cells from different species. A central part, called fibrillar center (FC), is seen in electron microscope as a transparent circular area of variable diameter that corresponds to NOR [1071 - 1073, 1082]. The number of FCs per nucleolus can vary from one to more than ten. The FC contains little RNA, but does contain DNA corresponding to one or several rRNA genes [472, 1074]. The FC is the basis of NOR and in actively growing cells, individual NORs can give rise to many FCs within one large nucleolus [1063, 577]. RNAP I transcription factors, some other proteins such as topoisomerase I and transcription start sites were found localized mainly within the FC underpinning it as a place of active transcription [1074, 1079].

The electron dense area at the periphery of FC was named dense fibrillar components (DFC) and consists of tightly packed fibrils [1076]. DFC is composed mainly of ribonucleoproteins (RNPs) and is considered to be the site of pre-rRNA synthesis, initial processing and modification [1075, 1076, 1079, 1063]. The initial processing in DFC occurs in several steps. First, the pre-rRNA undergoes 2 types of modifications: 2-O-ribose-methylation and pseudouridylation of several hundred nucleosides at important structural regions of the 18S and 28S genes [1067, 1068]. The synthesis of these modifications is guided by a very large number of 60 to 300 nucleotide long small nucleolar RNAs (snoRNAs) which associate with different proteins, including the enzymes pseudouridine synthases and methyltransferases, to form small nucleolar RNPs (snoRNPs). In addition, the snoRNAs are also involved in modification and intron excision of tRNA, modification of splicing-associated snRNA (chapter 1.6.2) and possibly mRNA [1093]. The functional importance of these modifications remains elusive. Second, the pre-rRNA precursor is cleaved at specific sites so that 18S, 28S and 5.3S ribosomal RNAs become released. The cleavage events require endonucleases and exonucleases and also a number of snoRNAs such as U3, U8, U14, RNase P and RNase MRP transcribed by RNAP II and –III [1062, 1084 - 1089]. A large snoRNP complex has been purified and characterized to contain U3 snoRNA and about 30 proteins [1084]. This complex is required for cleavage of 18S rRNA and assembly of the small 40S ribosomal subunit. Some components of this complex are also required for RNAP I function, therefore suggesting a communication/coordination between rRNA transcription and processing [1062].

After the early processing events, the rRNA molecules in complex with different snoRNPs are transported to the third nucleolar compartment named granular components (GCs). The GC consists of granules of about 15 nm in diameter which corresponds to pre-ribosomal particles at different stages of assembly up to complete ribosomes [1076]. The small, 40S ribosomal particle is processed via the intermediate 90S pre-ribosome containing U3 snoRNP [1090, 1091]. The large, 60S ribosomal particle undergoes a distinct pathway of formation [1092]. In general, ribosome formation is a highly complex multi-step process with a large variety of proteins and RNAs that associate with and guide ribosomal particles through different steps during their traveling from early processing through nucleolar compartments, nucleus and nuclear pores to cytoplasm [1066]. In addition, certain molecules are required for transport of ribosomal components to the nucleolus and their integration with other components, for regulation of the nucleolar morphology e.g. assembly/disassembly during cell cycle and other processes [1063, 1066 - 1069]. Such a complexity is underlined by proteome analysis that identified several hundreds of proteins in nucleoli of human cells [1064, 1065, 576]. This complexity is in line with numerous findings that in addition to the synthesis of rRNA and ribosome assembly, the nucleolus is involved in regulation of various cellular processes such as cell growth and cell cycle control, aging, maturation of non-ribosomal RNAs and proteins [1063].

In addition to the FC, DFC and GC compartments, the nucleolar chromatin appears as electron dense islands either within or outside of the nucleolus (correspondingly, intranucleolar and perinucleolar
Chapter 1

This chromatin corresponds to the non-transcribed intergenic spacer and also to inactive copies of rDNA genes.

2.2 Transcription factors of RNAP I

The complete 43 kb human rRNA repeat unit consists of two major parts: a transcription unit of about 13 kb, and an intergenic non-transcribed spacer (~30 kb) [312, GeneBank U13369]. The transcription unit contains 5’ and 3’ external and two internal transcribed spacers separating the sequences of 18S, 5.8S and 28S rRNA molecules. All four spacers and three rRNAs are transcribed as a single pre-rRNA, the 45S molecule which then is further subjected to processing steps.

The promoter of the rDNA gene contains several essential functional elements: a core element from −45 to +18 relative to the start site and an upstream control element (UCE) from −156 to −107. The distance between these elements and their relative direction are crucial for initiation [1094]. In addition, there is an upstream enhancer region activating rDNA gene transcription [1094, 1095]. Transcription initiation at the promoter requires assembly of a pre-initiation complex (PIC) comprising of RNA polymerase I and a number of other transcription factors. Promoter selectivity factor 1 (SL1 in human cells or transcription initiation factor (TIF)-IB in mouse cells) is a protein complex which consists of the TATA-binding protein (TBP) and several associated factors (TAFs): TAF110, TAF63 and TAF48 [1096, 1097]. SL1 specifically recognizes the core promoter via its TAF subunits, and not via TBP [26, 410].

Upstream binding factor (UBF) binds to the UCE as a dimer [873, 894, 1069]. It binds to the minor groove of DNA via several high-mobility-group domains (HMG boxes) [672, 696] which are similar to the domains found in HMG1/2 proteins – the important structural elements of chromatin and chromosomes [801]. UBF is highly conserved between species and has a number of critical functions. It is located in fibrillar centers and is considered as an architectural protein because as a dimer it can turn every 140 bp DNA to form a loop of 360°, thereby placing the core promoter element and UCE in close proximity [920]. This structure can facilitate interactions between UBF and SL1 which are associated with these promoter elements and, consequently, also the formation of a PIC [12]. The looping may indicate the importance of precise distance/positioning between promoter elements. In addition, the architectural role of UBF and its involvement in maintaining the structural conformation of the nucleolus is suggested from its binding across the entire rDNA transcription unit [669]. It has also been suggested to be a major force for the formation and maintenance of NOR structures, because genomic integration of arrays of UBF binding sites resulted in formation of structures morphologically similar to NORs even outside of the nucleolus [55]. Moreover, RNAP I transcription factors appeared to be recruited to the non-natural NORs independently of transcription and possibly via their interaction with UBF. Indeed, UBF can activate transcription via interaction with the SL1 subunits TAF48 and TBP and stabilize its promoter binding via interaction with PAF53 subunit of RNAP I and, in addition, via competing with nonspecific DNA-binding proteins, such as histones [12, 68, 111, 126, 873, 1069].

RNA polymerase I is a multiprotein complex consisting of about 14 subunits [1098]. There are two distinct populations of RNAP I. RNAP Iα comprises the majority (>90%) of the total cellular polymerase I pool; it is unable to initiate transcription from rDNA promoter and likely correspond to the elongation form of the polymerase. RNAP Iβ comprises <10% of cellular polymerases and is able to direct transcription initiation at the rDNA promoter and is associated with regulatory factor TIF-IA (Rrn3p in yeast) via PAF67 (pol-I-associated factor of 67 kDa) and RPA43 subunits of RNAP I [139,
The RNAP Iβ can be engaged in initiation at the rDNA promoters via interaction of TIF-IA with the TAF$_63$ and TAF$_{110}$ subunits of the SL1 complex [158, 159]. A number of RNAP II factors are also found to be required for transcription of rRNA genes. PCAF, which is an interaction protein of general RNAP II coactivators CBP and p300 and a subunit of PCAF TBP-free complex (chapter 1.1.11), has been shown to activate rRNA transcription possibly via acetylation of TAF$_{68}$ subunit of the SL1 complex [169]. A general transcription factor TFIH which has many functions in RNAP II transcription (chapter 1.1.6), and TAF12, the component of TFIID complex (chapter 1.1.1), have also been uncovered to have a role in rRNA transcription [289, 561].

Recent findings revealed that nuclear actin and nuclear myosin-I (NM1) are located in the nucleolus and that they interact with different components of the RNAP I transcription and are required for transcription of rDNA genes [308 - 310]. Nuclear actin is an ATPase that cycles between monomeric (G-actin or β-actin) and polymerized (F-actin) states and is associated with myosin [310]. These proteins also appear to be broadly involved in transcription of RNAP II genes via interaction with numerous transcription factors and complexes. Their precise function in rRNA transcription is not yet clear, but both proteins seem to be associated with initiation and elongation as they are present on promoters and transcribed regions. One obvious function could be that rRNA associated actin-myosin can act as transcription-coupled molecular motors required for mobility of nucleolar structures such as processing particles, or for a chromatin architectural function via anchoring of RNAP I transcription components to rDNA gene. Actin and NM1 are required for transcription in vivo but their physical location in the nucleolus is not yet established.

Elongation of RNAP I is a highly efficient process. When RNAP I is released from promoter into the elongation phase, TIF-IA dissociates whereas UBF and SL1 remain promoter bound thereby recruiting another RNAP Iβ and facilitating reinitiation of transcription from the same promoter and supporting multiple rounds of transcription. UBF, distributed over the entire rDNA gene, might have a regulatory role in elongation. A number of elongation factors are required including those which have function in RNAP II transcription [243]. Topoisomerase I is a DNA-bound enzyme that catalyzes topological changes in the DNA. It cuts one of the DNA strands allowing rotation of the other strand resulting in the relaxation of both negative and positive supercoils [265]. This enzyme is crucial for rRNA transcription and the majority of cellular pool is found in the nucleolus with a preferential location in fibrillar centers [230 - 232, 1079]. It functions to release topological stress in the rDNA during transcription such as torsional strain with positive supercoils ahead of the polymerase and negative supercoils behind. Topoisomerase I is highly sensitive to various negative factors that result in loss of the enzyme from fibrillar centers [376].

Cytological studies in different species have revealed that a large number (50 - 100) of polymerases can simultaneously be engaged in transcription of one rDNA gene unlike to RNAP II transcribed genes [472]. High polymerase loading gives rise to fir-tree like structures in the electron microscope (EM) [241, 1063]. These trees represent fully extended rDNA genes loaded with polymerases connected to nascent pre-rRNA transcripts of increasing length radiating away. At the 5’ ends of these transcripts ‘knob’ particles are observed which correspond to U3 snoRNP showing that processing occurs co-transcriptionally [1084]. The U3 processing particles appear to be very stably associated with rRNA because these EM sample are prepared under non-physiological conditions such as low ionic strength buffers and detergents leading to dissociation of most of the proteins. Indeed, processing factors specific for 18S and 28S molecules that are located in DFC sub-compartment of nucleolus cannot be detected in these preparations. The transcribed area of active rDNA does not
contain nucleosomes (also not in metaphase) whereas the non-transcribed intergenic spacer is packed in chromatin structures (nucleolar chromatin) [577].

Another co-transcriptional process which takes place during rRNA synthesis is transcription-coupled DNA repair (chapter 1.6.5) [266]. The process is mediated by a number of proteins involved in DNA repair, namely by TFIIH, CSB and XP, which are involved in RNAP I transcription [288, 289]. These factors are found in the protein complexes with RNAP I and SL1. Remarkable, ATPase, helicase and kinase activities of TFIIH are not required in these processes.

Termination of RNAP I transcription occurs at a termination region located at the 3′ end of the transcribed region of the rRNA gene [236]. This region consists of 3 groups (R1, R2 and R3) of specific elements, called Sal boxes. The specific factor TTF-I binds to the terminator sequences and cooperates with other factors to stop and release elongating polymerase from rDNA genes [214].

### 2.3 Dynamics of rRNA transcription

Distinct pathways of transcription initiation have been considered for rRNA. Initially, a number of studies characterized large multiprotein complexes in cell extracts form different species [873]. These complexes, named holoenzymes, contained RNAP I and many other factors and have been shown to be sufficient for transcription initiation and termination, rRNA synthesis and processing in vitro. Such holoenzymes are thought to be recruited to the sites of rDNA transcription. An alternative pathway is the stepwise assembly of transcription initiation complex on promoters from separate components.

Recently a mechanism of the rRNA transcription has been investigated with much greater attention for the dynamics of the processes. The movement of several individual subunits of RNAP I and basal factors UBF, TAF\textsubscript{I}48 and TIF-IA, which were tagged with green fluorescent protein (GFP), was monitored in living cells by FRAP technique (fluorescent recovery after photobleaching) [305]. This method allows analysis of the kinetics of transcription initiation and elongation. The RNA polymerase I transcription factors appear to form a highly dynamic protein complex that stochastically assembles from freely diffusible subunits. The majority of the proteins, including the subunits of RNAP I, enter into the nucleolus independently as distinct proteins rather than as a preassembled holoenzyme. All the components are steadily and rapidly exchanged between the nucleoplasm and FC in the nucleolus.

The data suggest that each subunit passes through the nucleolus with an average rate of several thousand molecules of per second with only a few second residence times within the nucleolus. The major fraction of each polymerase subunit (>90%) is not engaged in transcription, and less then 10% is incorporated into an elongating complex suggesting that the efficiency of incorporation is low. The residence times of RPA subunits at promoter is transient and observed between 0.2 and 1.2 seconds and an elongation time is estimated to be 2 – 3 minutes (average 140 seconds) with corresponding elongation rate of about 95 nucleotides/second for a human rDNA gene [312]. An average reinitiation interval of 1.4 seconds can be calculated when assuming the presence of ~100 polymerases per active gene. This reinitiation rate is consistent with the production of several million rRNA transcripts per cell cycle when assuming ~100 active rDNA genes per cell. Importantly, the rapid exchange of each protein between FC and the nucleoplasm suggests that RNAP I is not recycled after transcription termination, but rather reassembles at each round of transcription. The association of UBF and SL1 components with the promoter is also transient (from 3 to 5 seconds) but longer than of the RNAP I subunits, therefore allowing these factors to function in multiple rounds of initiation before dissociation from the DNA [317]. These calculations are well in accordance with previously estimated parameters of the rate of rRNA synthesis [326].
2.4 Regulatory mechanisms and pathways of RNAP I transcription

Synthesis of rRNA genes undergoes highly efficient regulation in response to metabolic as well as various external environmental changes and is critical for the adjustment of cell growth and proliferation according to growth factors, nutrient availability, differentiation and development stages. Most of the proteins required for RNAP I transcription can be potential targets for different regulatory pathways. This tight regulation is directly connected to synthesis of cellular proteins and therefore with cellular growth and proliferation. Production of rRNA is regulated via various mechanisms and pathways which are not well understood in details, and which can be modulated either via changes in transcription rate of the genes or/and via adjustment of the number of active rDNA genes. In mammalian cells the number of active rRNA genes varies between different cell types suggesting that the number of active genes is established during development and cellular differentiation [368]. In actively growing cells only about half of the rDNA genes are transcriptionally active and the other half is kept in an inactive state likely involving epigenetic mechanisms and the inactive state is stably maintained through cell divisions [356]. Upon tissue differentiation, the cellular metabolism is usually down regulated and rRNA synthesis reduced resulting in inactivation of a part of the rRNA genes [1069]. In response to regulatory stimuli triggered by e.g. cellular metabolism, the transcription rate of the active rDNA genes can increase [369]. Moreover, stimulation of differentiation can lead to an increase in the number of active genes if significant increase of rRNA is required. For example, in human peripheral blood lymphocytes, containing a single small nucleolus with one FC, large nucleoli appear with many FCs upon cell activation [368]. Regulatory mechanisms involving changes in transcription rate has been described in yeast. It was observed that the cells with only 42 copies of rRNA genes in their genome had similar amount of rRNA as in normal cells containing 143 copies of the genes of which 75 were active [326]. In addition, individual rRNA genes were found to have different number of elongating polymerases showing that the genes can be differently regulated. In most cases when short term regulation of the amount of rRNA is required, e.g. during cell cycle, regulation occurs via changing the transcription rate of the active genes or those genes which were not inactivated by repressive and stable epigenetic mechanisms [356, 378].

The regulatory pathways of rRNA transcription in response to growth-control factors, such as cell cycle arrest, nutrient starvation, protein synthesis inhibitors, occurs via dephosphorylation of TIF-IA at several sites [139, 142, 379, 382]. This leads to disruption of the interactions between TIF-IA with RNAP I and therefore makes polymerase unable to interact with SL1 and enter the preinitiation complex at the promoter. In addition, it induces translocation of TIF-IA into the cytoplasm. The phosphorylation state of TIF-IA is regulated via the mTOR kinase pathway which controls translation, ribosome biogenesis and many other growth-related processes in response to nutrients and environmental conditions [366], and the ERK-MAPK pathway controlling TIF-IA phosphorylation by ERK and RSK kinases in response to cell cycle activation [379]. Phosphorylation of UBF by the mTOR pathways upon growth induction facilitates the interaction between UBF and SL1 [404], and phosphorylation by ERK–MAPK pathway regulates its DNA binding [432]. The level of rRNA transcription is significantly changing during the cell cycle with a maximum in the S and G2 phases and little/suppressed transcription in metaphase which slowly recovers in G1 [873, 1063, 1069]. Phosphorylation of the TAF,110 subunit by CDK1/cyclin B at threonine 852 inhibits interaction of SL1 with UBF and results in transcriptional repression [426, 451] whereas acetylation of TAF,68 subunit by PCAF increases RNAP I transcription [169]. Activity of UBF is also regulated by phosphorylation [452], and via interaction with the cell-cycle regulating TAF1 subunit of TFIID complex [389, 404]. Acetylation of UBF by CBP in S-phase enhances its interaction with the PAF53
subunit of RNAP I and promotes its association with the rDNA promoter which then positively affects transcription [467]. Remarkably, SL1, UBF and RNA polymerase I remain associated with rDNA genes during mitosis therefore possibly marking the genes for reactivation [427, 428, 579, 580]. Stable inactivation of rDNA genes occurs via maintaining the DNA bound to nucleosomes in a transcriptionally silent chromatin state. The main factor required is NoRC (nucleolar remodeling complex), which belongs to the family of ATP-dependent ISWI/SNF2 chromatin remodeling complexes [987]. NoRC is associated with silent rRNA genes, it can position nucleosomes at promoter sites and can recruit histone deacetylase and DNA methyltransferase activities which results in a local repressive heterochromatic state and methylation of DNA at CpG sites at rRNA gene promoters [511, 521, 522]. Short RNA molecules derived from intergenic transcribed spacer within rRNA transcription unit are found to direct NoRC activity for rDNA gene silencing [528].

2.5 Spatial conformation of the transcribed rDNA gene
The length of the fir-tree like structure corresponding to the individual rDNA gene (13.3 kb transcribed region of about 4.5 micrometers in length) loaded with RNA polymerases largely exceeds the diameter of nucleoli suggesting that rDNA genes are not present in a linear form. Transcription fir-trees are prepared under non-physiological conditions and therefore lack most of the associated proteins. Consequently only a 2-dimentional picture is obtained and the spatial conformation is fully lost. Morphological studies clearly show that the transcription trees are highly compact RNP structures with a large amount of RNA [488, 499, 503] and therefore are unlikely part of electron transparent fibrillar centers but rather are located in DFC. However, visualization of the ‘transcription trees’ in the very compact DFC is impossible because the specific 3-dimentional organization cannot be preserved in electron microscopy samples. In addition, DFCs contain a large number of electron dense RNP which apparently are associated with nascent pre-rRNA transcripts [1062, 1075, 1076, 1079, 1084 - 1089] and are masking the fir-tree structure. High resolution visualization of different components within the nucleolus has revealed their specific location. Most of the RNAP I transcription factors are found in the FC but can also be detected in DFC [1063, 1074, 1079, 1079]. The active rRNA synthesis is localized in FC and FC/DFC border based on BrUTP labelling [549], but steady location of nascent pre-rRNA transcripts is mainly in DFC [1075, 1076, 1079, 1080, 1079] whereas the sites of transcription initiation are found mainly in FC [1074]. Transcribed rDNA regions as well as elongating RNA polymerase molecules are localized at the periphery of FC and in the area surrounding the FC/DFC border [1063, 1073, 1074, 1077, 1080, 1081, 1079]. A model based on this spatial location proposes that nascent rRNA is synthesized at the periphery of FC and enters the surrounding DFC where it interacts with processing RNP [1074, 1079, 577]. The identity and functions of extended DNA filaments without nucleosomes observed in electron microscopy in the central part of FC are still unknown, although it has been shown that this DNA does not correspond to the long non-transcribed intergenic spacers [577]. It is unlikely that this DNA corresponds to inactive rDNA genes which are present in the nucleosome-bound compact state of nucleolar chromatin [521]. Intriguingly, this intra-FC DNA is always present in the ‘open’ state and never subjected to compaction via association with nucleosomes, not even in metaphase when rRNA synthesis is suppressed [577, 554, 564]. This is in line with another remarkable feature that the transcription factors SL1 and UBF remain associated with FC during mitosis [427, 428, 579, 580]. Although the nucleolar structure has been extensively described in details, little is known about the spatial conformation of the transcribed rRNA gene. An electron tomography approach in combination
with highly efficient immunolabeling of RNA polymerases has recently been used to resolve this issue [1074]. The authors propose a model where the transcribed rDNA gene is coil-folded in several loops which are organized next to each other to form a cylinder-shaped structure. However, this model does not resolve the steric problem of how hundreds of transcripts [368] synthesized at extremely high rate can be efficiently disentangled from the complex network of the coil-folded DNA loops. It hardly seems possible that rRNA transcripts attached to large processing RNPs (such as U3 snoRNP) can follow RNA polymerase along DNA template being efficiently unwind from DNA and simultaneously exported to DFC. Also preparations of extended rDNA ‘fir-trees’ [241] show that all the transcripts can be easily separated from each other, which suggests that there is a unique conformation of the transcribed rDNA gene that allows individual transcripts to be spatially isolated.

3. RNA polymerase III transcription

RNA polymerase III transcribes several specific types of short (<400 bp) genes which do not encode proteins. The transcribed RNA from these genes functions as essential components in different cellular processes such as protein synthesis, metabolism, transcription, processing of pre-mRNA and pre-rRNA (transcribed by RNAP II and I respectively) as well as possible other functions. In comparison with RNAP II genes, the number of RNAP III transcribed genes in the genome is relatively small and diversity of transcription-regulatory mechanisms is not as high. RNAP III transcription factors and their functions are very well characterized. Nevertheless, as the genes are integral components of the central cellular processes, the specific regulatory mechanisms which synchronize expression of the RNAP III genes with other genes transcribed by RNAP I and II are not yet well understood.

3.1 Promoter types of the RNAP III transcribed genes

Known RNAP III genes can be sorted into four types based on their promoter structure [35]. Type 1 is assigned to 5S rRNA gene which is a component of ribosomes. The promoter of this gene is located within the transcribed region and consists of an A box, intermediate element (IE), and a C box [571, 573]. Transcription initiation from these promoters requires the general transcription factors TFIIIA, TFIIIB and TFIIIC.

The second type is assigned to tRNA genes whose promoters consist of two conservative A and B boxes also located inside of the transcribed region [585, 592]. These elements correspond to the functionally crucial D- and T-loops within the tRNA molecules. The tRNA genes require TFIIIB and TFIIIC.

The third type is assigned to several genes with distinct functions. The RNA encoded by U6 snRNA gene is a component of the spliceosome [594, 619], the human 7SK gene is involved in regulation of the RNAP II elongation factor CDK9 [392, 393, 600, 603], the human RNase P and RNase MRP genes encode RNA components of ribonucleoprotein complexes involved in processing of pre-tRNA and pre-rRNA [600, 601]. Promoters of these genes are external and located upstream of the 5’ end of a transcribed region. They consist of a TATA-box (in case of human U6 gene it is at position from -25 to -32), proximal sequence element (PSE) (from -68 to -65) and a distal sequence element (DSE) (from -215 to -240) [35, 726]. Type 3 genes require SNAPc and TFIIIB general initiation factors.

The fourth type is assigned to the 7SL gene in plants that encodes the RNA component of the cytoplasmic protein secretion system. Promoters of this gene are of a mixed type comprising...
intragenic A and B boxes as well as external elements [623, 726]. A number of unusual promoter structures have been described for some genes in different species [35]. RNAP III genes are present at different genomic locations: 5S rRNA and tRNA genes are grouped in clusters whereas U6 genes are dispersed as single genes over different chromosomes [726]. The cluster of the primate-specific miRNAs and Alu repeats interspersed on a ~100 kb region of chromosome 19 (C19MC) is recently described. These Alu-repeats function as RNAP III promoters and drive transcription of the adjacent microRNA genes [639, 641].

3.2 The RNAP III transcription factors

**RNA polymerase III** is composed of 17 subunits many of which have paralogues in RNAP I and -II [35]. The small and precise size of the RNAP III transcribed genes corresponds to the specific properties of the RNAP III such as exact positioning at the start of a gene via interactions with initiation factors, and termination precisely at the poly-T sequence at the end of a gene.

**TFIIIB** is a protein complex essential for transcription initiation of all types of RNAP III promoters. In yeast it consists of TBP and two other proteins, Brf1 (TFIIB-related factor 1) and Bdp1 (B double prime 1) [414 - 417, 692, 706]. Brf1 has structural similarity to TFIIIB (chapter 1.1.3) and contains similar N-terminal zinc-finger and core domains. The human TFIIIB complex is found in at least two diverse subunit compositions with apparent specificity for different types of RNAP III genes [665]. Similarly to yeast, promoters of tRNA and 5S rRNA genes (types 1 and 2) contain a TFIIIB which consists of TBP, Bdp1, Brf1 [670]. The U6 snRNA and other type 3 genes require a different TFIIIB complex consisting of TBP, Bdf1 and also Brf2 subunit that is translated from one of the truncated splice variants of the Brf1 pre-mRNA [664, 670, 673]. The mechanism of TFIIIB binding to a promoter is suggested to occur via a tight TBP-Brf1 heterodimer which can stably bind to the TATA-box and then recruit a Bdp1 subunit that weakly associates within a biochemically purified TFIIIB [664 - 668, 670]. When TFIIIB is stably bound to a promoter it can recruits RNAP III via extensive interactions of TBP and Brf1 with RPC39 and RPC9 subunits [679 - 682]. TFIIIB also assist in further positioning of the polymerase at the promoter and can direct multiple rounds of initiation. In addition, TFIIIB plays an essential role in formation of an open promoter structure (transcription bubble) which is an essential prerequisite for transcription initiation [731].

**TFIIIA** initiation factor is a DNA-binding protein (~49 kDa in human cells) containing of nine C2H2 zinc finger domains [736]. TFIIIA is weakly conserved between species and is required for transcription of 5S rRNA genes only [676, 739]. It recognizes the internal promoter of this gene and binds to the intermediate element, and A and C boxes via its zinc finger domains [35, 712, 713]. In addition, TFIIIA can bind to the 5S RNA to form a 7S storage ribonucleoprotein particle which is found in large amounts in X. laevis oocytes [676, 677].

**TFIIC** is a complex consisting of several subunits. The yeast complex consists of 6 subunits whereas human TFIIC contains of >10 subunits [684 - 686]. TFIIC is involved in transcription initiation of type 1 and 2 promoters. It recognizes TFIIBA bound at the promoters of the 5S rRNA genes or can directly bind DNA via recognition of the A and B boxes within tRNA genes [685, 702, 703]. It also interacts with TFIIIB and the RPC62 and RPC39 subunits of RNA polymerase III and contributes to their recruitment to promoters [711, 741 - 743]. It can enhance transcription on chromatin templates possibly via histone acetyltransferase (HAT) activity [710, 711]. An unusual role of TFIIC has been described in the general genome organization where it is recruited without RNAP III to the boundary elements preventing heterochromatin from spreading into neighboring euchromatic regions [746].
SNAPc complex is composed of 5 subunits (SNAP19, -43, -45, -50, -190 corresponding to their molecular masses) [34, 689, 690] which make specific interaction contacts with each other [697]. This complex is involved in transcription of type 3 genes only as well as via binding to PSE promoter element in near the U1 and U2 snRNA genes transcribed by RNAP II [697, 698]. SNAPc weakly binds to DNA due to the negative effect of the C-terminal part of SNAP190 and SNAP45, but strong binds to the U6 promoter when cooperating with TBP and transcription activator Oct-1 [559, 811 - 814].

Among other factors required for transcription of U6 snRNA gene are Oct-1 and STAF [32, 770]. These factors recognize their binding sites within the DSE element and cooperate with other factors in transcription initiation. They contain two types of domains which function as activators in transcription by RNAP II (in case of U1 and U2 snRNA genes) and by RNAP III (U6 snRNA gene). The Oct-1 contains a POU-domain consisting of two helix-turn-helix DNA-binding modules and functions in recruiting of SNAPc to the PSE site via interaction with the SNAP190 subunit inactivating its negative effect on the DNA binding [690, 891, 892].

4. Methodology for genome research

Sequencing of the genomes from many organisms opened up a novel field for comprehensive discovery of gene regulatory mechanisms towards the understanding of the molecular processes in living cells. It triggered the development of novel approaches in different fields such as genome-wide application of different molecular biochemical tools, identification of functional importance of genomic regions with yet unknown functions, new approaches in computational analysis and building of regulatory networks with integration of signalling pathways and gene regulatory mechanisms.

New biochemical approaches were developed for high-throughput measurements and analysis of multiple data points with extensive use of various computational tools that aim at a comprehensive investigation of the gene regulation and deciphering the underlying molecular mechanisms. One of the tasks that is critical for the understanding of molecular mechanisms is the identification of the gene regulatory elements. Among them are promoters, enhancers, LCRs and silencers (chapters 1.2 and 1.4) which are directly involved in gene regulation via interaction with transcription factors, as well as insulators and chromatin boundaries which provide additional regulatory levels. Among the recently developed methods for the finding of genomic binding sites are ChIP-cloning, ChIP-on-chip, ChIP-PET, ChIP-sequencing and CAGE which are used to identify the binding sites of DNA-interacting proteins and transcription start sites. During the last 5 – 7 years, with these methods a lot of significant information has been obtained and many novel insights were discovered.

4.1 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) is a central method used for identification of in vivo binding sites for transcription factors, histones and other chromatin-associated proteins. This method is based on covalent crosslinking of proteins to DNA at their natural binding sites in living cells by treatment of the cells with formaldehyde. The crosslinked chromatin can be then fragmented (mechanically or enzymatically) and then used for immunoprecipitation with antibodies against the specific protein [780]. This method is generally to cultured cells but also can be used with tissues [889, 890]. In spite of the obvious advantage to identify endogenous binding sites, this method has a number of limitation such as inability to precisely map the binding sites but rather pointing to a genomic region of about ~500 bp occupied with the transcription factor. This disadvantage can be overcome by combining
ChIP and DNase protection assay [784]. Another limitation is the inability to distinguish between the proteins which directly bind to DNA and those proteins which were linked to DNA via interaction with other proteins such as in multiprotein complexes. In addition, the specificity of antibodies and ability to efficiently recognize epitopes in a crosslinked protein is the major limiting factor. In spite of these disadvantages, ChIP remains the major method developed for chromatin research and was a basis for plethora of important observations obtained from a number of ChIP-related biochemical techniques (described below) and analysis tools.

4.2 ChIP-cloning
To identify the genomic binding regions for transcription factors, several studies combined ChIP with conventional cloning into plasmid vector and sequencing of the precipitated DNA fragments (ChIP-cloning). Among main disadvantages of this approach is the limited antibody specificity that results in cloning of non-specific DNA fragments. Therefore, laborious sequencing of a large number of clones and extensive validation of individual clones for binding using specific primers and quantitative PCR are required [717, 718]. Recently this method was used with a highly specific antibody against TBP and allowed the efficient identification of a large number of novel binding sites corresponding to promoters of known and novel genes and apparently also to enhancers [561]. The high-throughput validation in this study was optimized using microarrays (ChIP-on-chip) and revealed that the majority of the cloned fragments were significantly enriched in TBP binding.

4.3 ChIP-on-chip
The ChIP-on-chip approach was developed for an efficient screening of a large number of sites in order to identify the target genes/sites for various transcription factors as well as histone modifications [798 - 800, 1003]. With this technique the DNA after ChIP and total input DNA are labeled with different fluorescent dyes, mixed and hybridized on microarray slides. The ratio between signal intensity is a measure of the enrichment of certain targets. In this case sequencing is not required and only the DNA of targets of interest can be placed on the microarrays and further subjected to the analysis [796]. Different studies use distinct types of microarrays such as based on PCR products of different length [796], CpG-islands clones [877], large pieces of DNA such as BAC clones [886]. A very high intensification of the transcription factor binding studies became possible via application of ChIP-on-chip on high density microarrays. These arrays contain about 14 millions of 50mer oligonucleotides and designed to represent the entire non-repeated DNA in the human genome at about 100-bp resolution to identify a genome-wide map of active promoters [38].

The efficient identification of binding sites by ChIP-on-chip requires an extensive data analysis and setting up a certain threshold for the signal intensity to separate positive and negative targets. Calculation of this threshold includes the relative abundance of false-positive and false-negative targets and therefore requires extensive biochemical validation. The threshold can vary depending on antibody, microarrays types and experimental parameters. Several statistical approaches have been developed to assist data analysis [883]. Among them are median percentile rank [804], single array-error models [882], a sliding window analysis [883], variance stabilization [786], detection of enrichment [807], modeling-based methods [38] and others [803, 804]. Programs for the data analysis of the high density oligo-microarrays data include MPEAK [38] and PEAKFinder [834].

The identified sequences of the transcription factor binding sites can be used in follow-up analysis for the presence of a consensus sequence in the binding sites [284]. The exact binding site for the protein can not be precisely identified due to the method limitations, therefore algorithms have been
developed for the alignment of the sequences and extraction of the common sequence motifs. Among the tools for sequence analysis and motif discovery are MEME [948], AlignACE [980] and MDSce [982] which using different underlying models such as weight matrix model, Bayesian statistical score function and statistical classification methods. For example, estrogen receptor target sites were identified using classification and regression tree models [983]. In addition, integration of statistical algorithms with a pattern detection and phylogenetic footprinting allowed the detection of specific E2F1 and estrogen receptor target genes [1006, 1014]. Additional strategies can be used to increase the sensitivity of motif detection by building motif modules [1036, 1049, 1078].

4.4 ChIP-PET
The ChIP-PET method is a combination of the ChIP-cloning and the paired-end ditag (PET) massive sequencing strategy [908, 925]. In this case only 18 bp from 5’ and 3’ end of the cloned DNA fragments can be sequenced and then aligned to the genome to resolve the sequences of entire clones. The binding sites can be identified based on a large number of overlapping clones corresponding to a certain genomic loci, whereas clones representing background are expected to be randomly distributed throughout genome. The important advantage of ChIP-PET is that only short sequencing runs are required that allows to sequence tens of thousands clones and obtain a broad picture of the genomic locations for the binding sites. The main disadvantage of this method is the requirement of several rounds of blunt-end ligation and cloning during preparation of the PET library which can introduce a bias and decrease the efficiency and applicability of this method.

4.5 ChIP-sequencing
Although high-throughput sequencing can be a reliable method for identification of target sites, its broad application was limited due to requirement of extensive preliminary preparations of single DNA clones and relatively low output efficiency of conventional sequencing procedures. These limitations were largely resolved by the novel technical developments allowed to directly sequence millions of DNA fragments without additional purification steps – single DNA fragments can be attached to a solid surface of the carrier slides, amplified and used in especial sequencing procedure [1099]. Although the length of obtained sequences is limited to 30 - 40 bp, a large quantity of the parallel sequenced DNA fragments allows identification of over a billion nucleotides in a single run. Such an unprecedented efficiency opens up a lot of new perspectives and dramatically accelerates investigation and understanding of living processes and mechanisms. Broad application of this technique and its future modifications may result in new therapeutic and diagnostic developments and possibly will give a strong impulse for development of other research areas.

Among the first applications this method was immediately used for a finding of the genomic binding sites of different transcription factors and histone modifications over entire mouse and human genomes [1099 - 1101]. DNA samples after ChIP were directly amplified and used for sequencing. In this case DNA undergoes a minimal number of intermediate steps and, therefore, the identification of the binding sites is least biased in comparison with other techniques. Furthermore, this approach, called ChIP-sequencing, allows a reliable and precise identification of the target sites based on a large number of overlapping sequences corresponding to the same genomic positions. Similarly to ChIP-PET, background level can be defined from the genomic distribution of random sequences. Extensive bioinformatics analyses of the ChIP-sequencing data and deep data mining may uncover a novel molecular mechanisms and gene regulatory networks.
Chapter 1

4.6 CAGE
The cap analysis of gene expression (CAGE) approach has been used for the systematic analysis of 5'-end of the mouse and human transcripts and extensive mapping of transcription start sites over entire human and mouse genomes [1083]. This method is based on massive sequencing of concatemerized tags obtained from cleaved fragments (20 - 21 bp) of the biotin-capped 5'-ends of cDNA. These studies identified over 100,000 transcription start sites and promoters. Data analysis revealed an interconnected highly complex transcription network. Many promoters were found located within the protein coding genes and in the intergenic regions suggesting extensive transcriptional activity throughout the genome. The transcribed sequences often overlapped at the same or opposite DNA strands that is in line with recent observations [561, 947]. Furthermore, a large collection of sequences enabled an accomplished analysis of the core-promoter elements in vivo and determine the architecture for distinct types of promoters.

4.7 Transcription regulatory networks
The important application of the growing amount of data from ChIP-on-chip and transcriptome approaches together with protein–protein interactions maps is the modeling of transcriptional regulatory networks based on integrative analysis of the transcription factor binding sites and gene expression data [247]. These networks aim to model the multiple functional interactions at a global scale and to understand how they correspond to specific processes such as cell cycle, differentiation, cell growth and senescence, cellular responses to stress and metabolic signals [223, 1218]. Networks are built by recording the cascade of binding events of specific transcription factors to regulatory sites and the regulation of expression of the target genes in response to different nutrient stimuli in yeast [901 - 907] and in multicellular organisms [896 - 898]. These networks can be connected to the signal transduction pathways in gene regulation during development and differentiation [204, 767, 768]. Further improvements in experimental and computational approaches will provide an ability to build global regulatory networks of high complexity.

References
42. Simon MC, Fisch TM, Benecke BJ, Nevins JR, Heintz N. Definition of multiple, functionally distinct TATA elements, one of which is a target in the hsp70 promoter for E1A regulation. Cell. 1988. 52:723-729
51. Mitsiou DJ, Stunnenberg HG. TAC, a TBP-sans-TAFs complex containing the unprocessed TFII Aalphabet subunit. Mol Cell. 2000. 6:527-537
52. Tora L. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev. 2002. 16:673-675
63. Hanada K, Song CZ, Yamamoto K, Yano K, Maeda Y, Yamaguchi K, Muramatsu M. RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. EMBO J. 1996. 15:2217-2226
General introduction


73. Dikstein R, Ruppert S, Tjian R. TAF(II)125 is a bipartite protein kinase that phosphorylates the basal transcription factor RP74. Cell. 1996. 84:781-790


Chapter 1


118. Imbalzano AN, Zaret KS, Kingston RE. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. J Biol Chem. 1994. 269:8280-8286


124. Ozer J, Lezina LE, Ewing J, Audi S, Lieberman PM. Association of transcription factor IIA with TATA
binding protein is required for transcriptional activation of a subset of promoters and cell cycle progression in
125. Chi T, Carey M. Assembly of the isomerized TFIIA-TFIID-TATA ternary complex is necessary and
sufficient for gene activation. Genes Dev. 1996. 10:2540-2550
126. Kermekchiev M, Workman JL, Pikaard CS. Nucleosome binding by the polymerase I transactivator upstream
127. Lieberman PM, Berk AJ. A mechanism for TAFs in transcriptional activation: activation domain
IIA. Genetics. 1999. 153:1573-1581
130. Warfield L, Ranish JA, Hahn S. Positive and negative functions of the SAGA complex mediated through
interaction of Sp1p with TBP and the N-terminal domain of TFIIA. Genes Dev. 2004. 18:1022–1034
131. Chi T, Lieberman P, Ellwood K, Carey M. A general mechanism for transcriptional synergy by eukaryotic
133. Kobayashi N, Boyer TG, Berk AJ. A class of activation domains interacts directly with TFIIA and stimulates
9:1354-1365
135. Stargell LA, Moqtaderi Z, Dorris DR, Ogg RC, Struhl K. TFIIA Has Activator-dependent and Core Promoter
137. Deng W, Roberts SG. A core promoter element downstream of the TATA box that is recognized by TFIIIB.
139. Yuan X, Zhao J, Zentgraf H, Hoffmann-Rohrer U, Grummt I. Multiple interactions between RNA polymerase
I, TIF-IA and TAF(I) subunits regulate preinitiation complex assembly at the ribosomal gene promoter.
140. Tsai FT, Sigler PB. Structural basis of preinitiation complex assembly on human pol II promoters. EMBO J.
141. Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG, Burley SK. Crystal structure
143. Bushnell DA, Kornberg RD. Complete, 12-subunit RNA polymerase II at 4.1 Å resolution: implications for
144. Hahn S. Binding of TFIIIB to RNA polymerase II: mapping the binding site for the TFIIIB zinc
2004. 11:394-403
146. Chen HT, Hahn S. Mapping the Location of TFIIIB within the RNA Polymerase II Transcription Preinitiation
147. Bushnell DA, Westover KD, Davis RE, Kornberg RD. Structural basis of transcription: an RNA polymerase
149. Chen HT, Legault P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: an RNA polymerase
150. Chen HT, Legault P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: separation of RNA from DNA
by RNA polymerase II. Science. 2004. 303:1014–1016
151. Kaiser K, Stelzer G, Meisterernst M. The coactivator p15 (PC4) initiates transcriptional activation during
184. Chambers RS, Wang BQ, Burton ZF, Dahmus ME. The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. J Biol Chem. 1995. 270:14962-14969
201. Parvin JD, Sharp PA. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. Cell. 1993. 73:533-540
Chapter 1


224. Yankulov KY, Bentley DL. Regulation of CDK7 substrate specificity by MAT1 and TFIH. EMBO J. 1997. 16:1638-1646


230. Muller MT, Pfund WP, Mehta VB, Trask DK. Eukaryotic type I topoisomerase I topoisomerases is enriched in the nucleolus and catalytically active on ribosomal DNA. EMBO J. 1985. 4:1237-1243

Chapter 1

259. Wei W, Dorjsuren D, Lin Y, Qin W, Nomura T, Hayashi N, Murakami S. Direct interaction between subunit RAP30 of the transcription factor IIF (TFIIF) and RNA polymerase subunit 5, which contributes to the association between TFIIF and RNA polymerase II. J Biol Chem. 2000. 276:12266–12273


272. Kugel JF, Goodrich JA. Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIHH, TFIHH, and ATP on negatively supercoiled DNA. Proc Natl Acad Sci. 1998. 95:9232–9237

273. Kumar KP, Akoulitchev S, Reinberg D. Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event. Proc Natl Acad Sci. 1998. 95:9767–9772


General introduction


307. Holstege FCP, van der Vliet PC, Timmers HTM. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. EMBO J. 1996. 15:1666–1677


326. French SL, Osheim YN, Ciofi F, Nomura M, Beyer AL. In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Mol Cell Biol. 2003. 23:1558–1568
General introduction


346. Myers JK, Morris DP, Greenleaf AL, Oas TG. Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1. Biochemistry. 2001. 40:8479-8486


352. Woychik NA, McKune K, Lane WS, Young RA. Yeast RNA polymerase II subunit RPB11 is related to a subunit shared by RNA polymerase I and III. Gene Expr. 1993. 3:77–82


372. Yuryev A, Corden JL. Suppression analysis reveals a functional difference between the serines in positions two and five in the conserved sequence of the C-terminal domain of yeast RNA polymerase II. Genetics. 1996. 143:661-671
General introduction


Chapter 1

495. Wright KJ, Marr MT, Tjian R. TAF4 nucleates a core subcomplex of TFIIID and mediates activated transcription from a TATA-less promoter. Proc Natl Acad Sci. 2006. 103:12347-12352
General introduction


498. Dubrovskaya V, Lavigne AC, Davidson I, Acker J, Staub A, Tjarian R. Distinct domains of hTAFII100 are required for functional interaction with transcription factor TFIIF beta (RAP30) and incorporation into the TFIIID complex. EMBO J. 1996. 15:3702-3712


508. Lavigne AC, Mengus G, Gangloff YG, Wurtz JM, Davidson I. Human TAF(II)55 interacts with the vitamin D(3) and thyroid hormone receptors and with derivatives of the retinoid X receptor that have altered transactivation properties. Mol Cell Biol. 1999. 19:5486-5494


Chapter 1


535. Perletti L, Dantonel JC, Davidson I. The TATA-binding protein and its associated factors are differentially expressed in adult mouse tissues [In Process Citation]. J Biol Chem. 274:15301–15304


553. Mitsiou DJ, Stunnenberg HG. p300 is involved in formation of the TBP-TFIIA-containing basal transcription complex, TAC. EMBO J. 2003. 22:4501-4511


556. Guermah M, Ge K, Chiang CM, Roeder RG. The TBN protein, which is essential for early embryonic mouse development, is an inducible TAFII implicated in adipogenesis. Mol Cell. 2003. 12:991–1001


573. Yuan CX, Ito M, Fondell JD, Fu ZY, Roeder RG. The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. Proc Natl Acad Sci. 1998. 95:7939–7944


Chapter 1


602. Chong JA, Moran MM, Teichmann M, Kaczmarek JS, Roeder R, Clamph DE. TATA-Binding Protein (TBP)-Like Factor (TLF) is a Functional Regulator of Transcription: Reciprocal Regulation of the Neurofibromatosis Type 1 and c-fos Genes by TLF/TRF2 and TBP. Mol Cell Biol. 2005. 25:2632–2643

General introduction

619. Kunkel GR, Pederson T. Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used. Genes Dev. 1988. 2:196-204
622. Grant PA, Schieltz D, Pray-Grant MG, Steger DJ, Reese JC, Yates JR, Workman JL. A subset of TAFIIIs are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. Cell. 1998. 94:45–53
Chapter 1

677. Pelham HR, Brown DD. A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc Natl Acad Sci. 1980. 77:4170-4174
684. Wang Z, Roeder RG. DNA topoisomerase I and PC4 can interact with human TFIIC to promote both accurate termination and transcription reinitiation by RNA polymerase III. Mol Cell. 1998. 1:749-757
Chapter 1


703. Bartholomew B, Kassavetis GA, Geiduschek EP. Two components of Saccharomyces cerevisiae transcription factor IIIB (TFIIIB) are stereospecifically located upstream of a tRNA gene and interact with the second-largest subunit of TFIIIC. Mol Cell Biol. 1991. 11:5181-5189


709. Rougvie AE, Lis JT. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell. 1988. 54:795–804


General introduction


734. Gilmour DS, Lis JT. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol. 1986. 6:3984–3989


89
Chapter 1


762. Rosonina E, Kaneko S, Manley JL. Terminating the transcript: breaking up is hard to do. Gene Dev. 2006. 20:1050-1056


764. Ujvari A, Pal M, Luse DS. RNA polymerase II transcription complexes may become arrested if the nascent RNA is shortened to less than 50 nucleotides. J Biol Chem. 2002. 277:32527–32537


769. Whitelaw E, Proudfoot N. α-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3’-end processing in the human α2 globin gene. EMBO J. 1986. 5:2915–2922
General introduction

772. Luo W, Johnson AW, Bentley DL. The role of Rat1 in coupling mRNA 3’-end processing to transcription termination: Implications for a unified allosteric–torpedo model. Genes Dev. 2006. 20:954–965
777. Fong N, Bentley DL. Capping, splicing, and 3’ processing are independently stimulated by RNA polymerase II: Different functions for different segments of the CTD. Genes Dev. 2001. 15:1783–1795
782. Bucheli ME, Buratowski S. Npl3 is an antagonist of mRNA 3’ end formation by RNA polymerase II. EMBO J. 2005. 24:2150–2160
785. Orozco IJ, Kim SJ, Martinson HG. The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J Biol Chem. 2002. 277:42899–42911
Chapter 1


806. Ho CK, Shuman S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell. 1999. 3:405-411


816. Ares M, Grate J, Pauling M. A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA. 1999. 5:1138-1139


829. O'Mullane L, Eperon IC. The pre-mRNA 5′ cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5′ splice sites. Mol Cell Biol. 1998. 18:7510–7520
833. Tennyson CN, Klamut HJ, Worton RG. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet. 1995. 9:184–190
839. Misteli T, Spector DL. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol Cell. 1999. 3:697–705
Chapter 1

General introduction


894. McStay B, Sullivan GI, Cairns C. The Xenopus RNA polymerase I transcription factor, UBF, has a role in transcriptional enhancement distinct from that at the promoter. EMBO J. 1997. 16:396–405


Chapter 1


934. Tornaletti S. Transcription arrest at DNA damage sites. Mutat Res. 2005. 577:131–145


937. Marenduzzo D, Faro-Trindade I, Cook P. What are the molecular ties that maintain genomic loops? Trends Genet. 2007. 23:126-133


964. Molinari E, Gilman M, Natesan S. Proteasome-mediated degradation of transcriptional activators correlates with activation domain potency in vivo. EMBO J. 1999. 18:6439–6447
971. Walter J, Dever CA, Biggin MD. Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in Drosophila embryos. Genes Dev. 1994. 8:1678–1692
General introduction


1013. Qin S, Parthun MR. Recruitment of the type B histone acetyltransferase Hat1p to chromatin is linked to DNA double-strand breaks, Mol Cell Biol. 2006. 26:3649–3658


Chapter 1

1025. Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails, Genes Dev. 2006. 15:2343–2360


Chapter 1


1090. Schafer T, Strauss D, Petfalski E, Tollervey D, Hurt E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. EMBO J. 2003. 22:1370-1380


1092. Nissan TA, Bassler J, Petfalski E, Tollervey D, Hurt E. 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. EMBO J. 2002. 21:5539-5547


Chapter 2

Identification of novel functional TBP-binding sites and general factor repertoires

Sergey Denissov, Marc van Driel, Renate Voit, Maarten Hekkelman, Tim Hulsen, Nouria Hernandez, Ingrid Grummt, Ron Wehrens and Hendrik Stunnenberg

The EMBO Journal (2007) 26, 944–954
Transcription factor signatures of human promoters

Abstract
Our current knowledge of the general factor requirement in transcription by the three mammalian RNA polymerases is based on a small number of model promoters. Here, we present a comprehensive chromatin immunoprecipitation (ChIP)-on-chip analysis for 28 transcription factors on a large set of known and novel TATA-binding protein (TBP)-binding sites experimentally identified via ChIP cloning. A large fraction of identified TBP-binding sites is located in introns or lacks a gene/mRNA annotation and is found to direct transcription. Integrated analysis of the ChIP-on-chip data and functional studies revealed that TAF12 hitherto regarded as RNA polymerase II (RNAP II)-specific was found to be also involved in RNAP I transcription. Distinct profiles for general transcription factors and TAF-containing complexes were uncovered for RNAP II promoters located in CpG and non-CpG islands suggesting distinct transcription initiation pathways. Our study broadens the spectrum of general transcription factor function and uncovers a plethora of novel, functional TBP-binding sites in the human genome.

Introduction
The comprehensive mapping of transcription regulatory regions in the genome of higher eukaryotes and the analysis of transcription factors recruited to these sites are major challenges notwithstanding the availability of the entire sequence of many genomes. Current annotations are skewed towards protein-coding genes and the assignment of promoters towards CpG islands. Regulatory regions positioned far away from the transcription start site such as enhancers and locus control regions are difficult to identify. In silico prediction of regulatory regions remains difficult notwithstanding first successes (Xie et al, 2005; Hallikas et al, 2006). Our knowledge of the organization and factor composition of promoters in higher eukaryotes is based largely on reporter gene assays and in vitro transcription reconstitution studies involving a small number of model promoters. Collectively, these studies identified and characterized general transcription factors and provided valuable insights of the mechanisms of transcription (Lee and Young, 2000; Sims et al, 2004). It has remained unresolved whether general transcription factors are universally involved in transcription or whether they are truly specific for a given RNAP class. Experimental approaches to systematically identify regulatory regions and to characterize their organization and regulation are, therefore, of great importance.

The multitude of general (co)factors, sequence-specific DNA-binding factors, bridging complexes, chromatin modifying and remodeling complexes involved in transcription is staggering and has been estimated to involve up to 6% of the protein coding genes in mammalian genomes (Tupler et al, 2001). Chromatin immunoprecipitation (ChIP) has proven to be a valuable tool in establishing the involvement and chronology of the recruitment of transcription factors and cofactors to a gene or locus. Application of ChIP to large sets of genes, ChIP-on-chip, has added a new dimension to target site identification and transcription factor occupancy profiling. General patterns and principles of gene regulation are currently being uncovered (Ren et al, 2000, 2002; Cam et al, 2004; Kim et al, 2005; Boyer et al, 2006).

Here, we report the identification and annotation of genomic binding sites of the central transcription factor TBP (TATA-binding protein) using sequential ChIP and direct cloning of the DNA fragments. Annotation of the clones revealed unique genomic loci containing known or predicted sites and a surprisingly large proportion of TBP-binding sites in introns or in regions without gene annotation. An experimentally derived TBP target site microarray was used in ChIP-on-chip to obtain binding profiles 26 transcription factors and two histone marks. We show that some transcription factors
hitherto reported to regulate transcription by RNA polymerase II (RNAP II) are also recruited to rRNA promoters suggesting cross-regulation between these classes of genes. Furthermore, correlation analysis of ChIP-on-chip data revealed distinct profiles corresponding to CpG and non-CpG island promoters transcribed by RNAP II suggesting distinct mechanisms of transcription initiation.

Results

Identification of in vivo TBP-binding sites

To identify a broad selection of in vivo TBP-binding sites, we used sequential ChIP using the human U2OS cell line, a highly specific monoclonal antibody against the N-terminal part of the TBP (Ruppert et al, 1996) and cloning of the precipitated DNA fragments (Supplementary Figure 1A). We reasoned that targeting TBP, the central factor in transcription, should ensure that promoters of genes transcribed by all three RNA polymerases were obtained. Cloning of ChIP’ed DNA fragments without prior amplification yielded a library of >20K colonies. A representative number of colonies (2000) were randomly picked. The lengths of the cloned fragments ranged from 40 to 500 bp averaging about 160 bp (Supplementary Figure 1A and B).

Figure 1. Construction and annotation of the TBP-binding site library.

(A) Outline of the strategy for ChIP-cloning and filtering of sequences. 'Filter': short sequences (<40 bp), highly repetitive sequences and those with less than 90% identity to the genome were eliminated. 'Collapse': 497 overlapping sequences were collapsed into 177 contigs.

(B) Pie diagram of annotation. Transcription-linked features were obtained from UCSC genome browser (HG16) using a 1 kb window centered at the cloned DNA sequences. Annotation of RNAP II genes was based on SWISS-PROT, TrEMBL, RefSeq and mRNA GenBank databases. Identity to rRNA genes was obtained by NCBI BLAST alignment. The number of clones in different categories was determined using the non-collapsed set of 1361 clones.

Inserts larger than 40 bp were annotated using the UCSC genome browser and NCBI BLAST. Highly repetitive sequences and such with less than 90% identity to the genome were eliminated (Figure 1A). The sequence complexity and overlap of the remaining putative TBP-binding sites (1361 clones) were analyzed via genome alignment and sequence comparison using TIGR Assembler (Sutton et al, 1995); about 61% of the target sites (864 clones) were present only once. Overlapping sequences (497
clones) were collapsed into a total of 177 contigs and that mainly comprised promoters of high-copy-number genes such as tRNA and rRNA. TBP-binding sites were annotated and sorted on the basis of transcription-linked features such as the presence of known genes and mRNA. An annotation window of 1 kb centered on the cloned sequence was chosen based on the resolution of ChIP experiments. Annotation of the top-ranked hit for each sequence revealed that 29% overlapped with the first exon of annotated genes or with the 5' end of mRNA (Figure 1B) and mostly located to CpG islands. A remarkably large fraction of targets was located in introns of known genes or in regions lacking a gene or mRNA annotation (20 and 28%, respectively). Fragments corresponding to RNAP III genes (15%) comprised rRNA and other different small structural RNA genes. rDNA sequences accounted for 10% of the cloned fragments.

Validation of TBP-binding sites by ChIP-on-chip
To study the binding sites of TBP by ChIP-on-chip we PCR-amplified inserts from the 2000 randomly picked clones, printed them on glass slides and hybridized DNA from input chromatin and TBP ChIP. The ChIP/input ratios of a set of reference promoters printed on the array showed a highly significant correlation value \( r=0.83, P=10^{-7} \) with TBP occupancy as determined by single gene quantitative PCR (qPCR) (Supplementary Figure 2). This implies that the data obtained by ChIP-on-chip faithfully reflects TBP occupancy in vivo. To define a threshold value, we computed frequency histograms of the ChIP/input ratios for all targets as well as for annotated promoters of the RNAP I, II and III class. On the vast majority (>95%) of RNAP II promoters, TBP was enriched more than two-fold over negative controls (Figure 2A). RNAP I and III targets displayed a high TBP occupancy ranging from 6- to >30-fold. Applying an arbitrary two-fold cutoff value implies that 90% of the targets are significantly enriched for TBP.

ChIP-on-chip on the TBP-binding site microarray was also validated by profiling for binding sites of the transcription factors E2F1 and E2F4. We identified 22 targets that were selectively enriched with both E2F1 and E2F4 (Supplementary Table I); most of them corresponding to promoters of previously identified E2F target genes (Ren et al, 2002; Cam et al, 2004). Hence, the microarray can be used to reliably measure transcription factor occupancy.

Principal component analysis
For a comprehensive profiling of general transcription factors and assessment of factors occupancy, ChIP-on-chip experiments were performed with antibodies against 26 different RNAP I, II and III-linked transcription factors and two histone marks that correlate with transcription. The intrinsic structure and complexity of the data set was assessed by principal component analysis (PCA) of the ChIP/input ratios for the different transcription factors. PCA defines a small set of latent orthogonal variables (principal components, PCs) that describe maximal possible variance in the entire data set. Figure 2B shows that targets segregated into three spaces according to the highest variance in their factor profiles; color coding of the known TBP target sites belonging to either of the three gene classes visualized their good separation. A small number of known RNAP II promoters ended up in the RNAP III realm and vice versa; inspection of their genomic organization revealed that these targets contained closely positioned RNAP II and III promoters. Importantly, the majority of non-annotated, novel TBP-binding sites was found in the space assigned to RNAP II. We conclude that these TBP target sites are most likely regulatory regions directing RNAP II-dependent transcription.
Chapter 2

Figure 2. Analysis of ChIP-on-chip data for different classes of promoters.
(A) Frequency histograms of TBP ChIP/input ratios (non-collapsed set). Dashed line indicates two-fold threshold. Promoters of RNAP I, II and III genes are colored in green, red and blue, respectively. Normalization controls correspond to '0' value on the histograms.
(B) Projection of the ChIP-on-chip data set into the space of the second and third PCs. Intronic targets and those without gene/mRNA annotation are highlighted in light blue. The spaces containing 95% of targets are shown as ovals of the RNAP I, II and III targets. The fraction of variance comprised in individual PCs is indicated in brackets.

Novel TBP-binding sites direct transcription
To further characterize these novel TBP-binding sites, we compared their transcription factor occupancies with those of annotated promoters by computing the frequency histogram on the ChIP/input ratios. The novel TBP-binding sites showed a slightly lower distribution of TBP, TFIIB and RNAP II occupancy compared to annotated promoters (Supplementary Figure 3). To test whether the novel TBP-binding sites can direct transcription, we randomly picked 27 targets for further analysis. The majority of these targets (25/27) showed significant enrichment for TBP and RNAP II in single gene qPCR (data not shown); the qPCR values correlated well with ChIP/input ratios for TBP as determined by microarray analysis ($r=0.83$).

To assess the competence of these sites to direct transcription two approaches were used. First, the validated novel TBP-binding sites were PCR-amplified from the genome as 1 kb fragments and cloned into a promoter-less reporter along with positive and negative controls. The majority of the intronic sites (11 out of 15) activated unidirectional transcription of the reporter gene (Figure 3A). Remarkably high activation (250-fold) was found for a site (F11-3-46) located in the first intron of the EGFR gene 100 kb downstream of the transcription start site. Transcription activation was collinear with the direction of transcription of the EGFR gene suggesting that this novel site is an alternative promoter. Several intronic TBP target sites, such as F5-4-46 located in the 1st intron of TFIIA genes, displayed promoter activity in the opposite direction suggesting novel antisense transcripts. The majority of the novel TBP-binding sites without gene/mRNA annotation (8 out of 11) displayed significant activation of the promoter-less reporter (Figure 3B). Interestingly, one of the novel TBP target sites comprised the intronic enhancer of GADD45 gene. Consistent with its well-documented enhancer function, this target activated SV-40 promoter in both orientations (Figure 3D). A number of other TBP-binding sites tested in this assay displayed enhancer activity (Figure 3D) suggesting that a fraction of the cloned TBP-binding sites may comprise enhancers. Promoters of five housekeeping
Transcription factor signatures of human promoters

genes identified in our screen were used as positive controls in this assay and they displayed on average stronger activation potential than the novel TBP-binding sites (Figure 3C). Eight randomly chosen genomic regions displayed little to no transcription activation (Supplementary Figure 4).

Figure 3. Functional analysis of novel TBP-binding sites.
Genomic DNA fragments containing novel TBP-binding sites were cloned in both directions in front of promoter-less (A–C) or SV-40 promoter containing (D) reporter-gene plasmid vectors and transfected into U2OS cells; ratios of transcription activity of the reporter gene over empty vector are shown. (A) Novel TBP-binding sites located in introns of RNAP II genes. The '+' and '–' refer to the direction of transcription of the gene (sense and antisense, respectively). (B) Novel TBP-binding sites lacking gene/mRNA annotation. The '+' and '–' refer to the direction of the sequence (UCSC genome browser definition) with respect to the reporter gene. (C) Promoters of RNAP II-transcribed genes. (D) Enhancer assay: analysis of the targets in a reporter vector with SV-40 promoter. The '+' and '–' refer to the direction of the sequence (UCSC genome browser definition) with respect to the reporter gene.

To test the promoter activity of the novel sites in their genomic location in vivo, we used strand-specific RT qPCR (sts-RT qPCR) to identify transcripts originating from the TBP-binding sites. Primers were designed in close proximity (about 500 bp) around TBP-binding sites (Figure 4A). The ratio between relative RNA levels for two probes targeting the same strand (A/C and D/B, respectively) was used to assess the directionality of transcription: high A/C and D/B ratios suggest specific transcription started at novel sites in '–' and '+' direction, respectively. High ratios for both A/C and D/B would imply bidirectional transcription.

As presented in Figure 4B and C, about half of the targets (17/26) yielded transcripts originating around the novel TBP-binding sites (ratios >5-fold). A good correspondence to reporter assay was observed for 12 targets. For example, high D/B ratio was obtained at intronic EGFR site (F11-3-46) suggesting that transcription is initiated at the TBP-binding site in sense direction (collinear with the
Chapter 2

gene) (Figure 4D) corroborating and extending its assignment as an alternative promoter. Similarly, high A/C ratio obtained at the intronic site in TFIIA gene (F5-4-46) underscores the presence of antisense transcription (Figure 4D) as also deduced from the reporter assay. Interestingly, a novel site located in a centromeric satellite region (B10-10-39) displayed both high A/C and D/B ratios suggesting bidirectional transcription.
Collectively, these data provide strong evidence that the majority of the novel TBP-binding sites function as genuine promoters.

![Figure 4](image-url)

Figure 4. Analysis of strand-specific transcripts at novel TBP binding sites
(A) Schematic presentation of novel TBP binding site (TBS) and location of strand-specific RT qPCR probes. Dotted lines indicate putative transcripts initiated at the TBP binding site. The probes named A and C are complementary to transcripts in ‘–’ direction and probes B and D - to transcripts in ‘+’ direction. The A/C and D/B ratios between RNA levels were taken to assess transcription specifically started within the novel TBP binding sites in ‘–’ and ‘+’ directions, respectively.
(B) The ratios A/C (left part) and D/B (right part) for TBP binding sites in introns. Transcriptional directions indicated with ‘+’ and ‘–’ refers to sense and antisense direction.
(C) Same as (B) measured for the TBP binding sites loci lacking a gene annotation. The directions of transcription indicated with ‘+’ and ‘–’ refers to UCSC genome browser definition.
(D) Schematic presentation of transcripts from the EGFR and TFIIA αβ genes.

Correlation profiling analysis
To gain insight into the occupancy of the TBP-binding sites by general transcription factors in relation to their function in transcription, we performed correlation analyses determining the degree of linear relationship between variables, that is, between ChIP/input ratios for each of the different factors. The correlation values were calculated between every possible pair of factors on all the targets and were then color visualized (Figure 5A). To bring the multitude of values into an order, clustering algorithms were applied to calculate hierarchical dendrogram based on the difference
Transcription factor signatures of human promoters

between correlation values (Figure 5A and B); the length of the branches is used as measure of the degree of difference (similarity–dissimilarity). This type of analysis can be used to compare occupancy profiles: factors co-recruited to the same target sets will show a high correlation and will cluster together, whereas factors that do not co-occupy the same target sets will have a low correlation and will be placed more distant from each other.

Analysis of the entire ChIP-on-chip data set revealed four major clusters (Figure 5A and B). RNAP III-specific factors such as Bdp1 and Brf1 found in TFIIIB and the RNAP III subunit RPC1 clustered in one branch and showed a negative correlation with RNAP I and II factors. Another branch of the dendrogram consists of subunits of the SNAPc complex that are specifically involved in transcription from small nuclear RNA genes.

Figure 5. Correlation analyses of ChIP-on-chip data sets.

Pearson correlation values were calculated on entire ChIP-on-chip data set (25 antibodies against general transcription factors and two active histone marks) and structured by hierarchical clustering (Ward’s). The resulting dendrogram is represented as a cluster (A) and a rooted tree (B). The latter is combined with colorvisualized correlation values as depicted. The branches corresponding to the different clusters are color-coded. TBP was excluded from the analysis.

The third branch brings together the known RNAP II factors and the two histone marks correlated with active promoters; H3K9ac and H3K4me3 (Berger, 2002; Santos-Rosa et al, 2002). RNAP II closely co-clustered with these histone modifications in line with recent findings (Bernstein et al, 2005; Kim et al, 2005). The transcription coactivator CBP/p300 and the negative cofactor NC2 showed a high correlation with general factors such as TFIIB suggesting that these factors serve general roles in RNAP II transcription. TBP-associated factors (TAFs) were clustered in a distinct sub-branch suggesting that the RNAP II targets are heterogeneous with respect to TAF occupancy. The RNAP I branch displays short distances between factors (Figure 5A and B) and was the farthest separated and hence the most dissimilar from the other branches which is in good agreement with the PCA analysis (Figure 2B). Surprisingly, the histone acetylase PCAF hitherto known as a subunit of the STAGA/PCAF complex (Vassilev et al, 1998) and TAF12, known as a component of the PCAF and TFIID complexes (Ogryzko et al, 1998), co-clustered with RNAP I-specific factors. The
recruitment of these factors - hitherto described as RNAP II-specific - to rDNA units was confirmed by single gene qPCR analysis (Supplementary Figure 5).

Involvement of TAF12 in transcription of rRNA genes

The association of PCAF with rDNA is in accordance with our previous studies showing that PCAF acetylates TAFI68 and stimulates RNAP I transcription in a reconstituted in vitro system (Muth et al., 2001). The presence of an RNAP II-specific TAF at the rDNA promoter was surprising and suggested that TAF12 may play a role in RNAP I transcription. To examine whether TAF12 is associated with the RNAP I-specific TBP-TAF1-complex SL1, we performed GST pull-down assays and measured the interaction of TAF12 with individual subunits of SL1, for example, TBP, TAFI110, TAFI68 and TAFI48. Consistent with published data, TBP was found to associate with GST-TAF12 (Hoffmann and Roeder, 1996). Noteworthy, TAFI48 and TAFI110, but not TAFI68 and the RNAP I transcription factors TIF-IA and UBF, were specifically retained on GST-TAF12 beads, indicating a direct interaction of TAF12 with SL1 (Supplementary Figure 6). The interaction of SL1 and TAF12 was also shown by co-immunoprecipitation experiments. Partially purified SL1 was precipitated with antibodies against TAFI110 and coprecipitated TBP and TAF12 were identified on Western blots. A significant amount of TAF12 coprecipitated with TBP and TAFI110, showing that TAF12 is associated at least with a subpopulation of SL1 in vivo (Figure 6A). Notably, TAF10, another RNAP II-specific TAF, was not detected in the immunoprecipitation.

Figure 6. TAF12 associates with SL1 and stimulates rDNA transcription.

A) TAF12 is associated with SL1. HeLa nuclear extracts were fractionated by chromatography on phosphocellulose and SP resins, and SL1 was immunoprecipitated using anti-TAFI110 antibodies (lane 3) or rabbit IgG (lane 2) as a control. The immunoprecipitates were analyzed on Western blots for TBP, TAF12 and TAF10 as indicated. The input (lane 1) contains 50% of the material used for the IP. To monitor the efficiency of TAFI110 precipitation, 10% of the input fraction and 10% of the IP were separated by SDS–PAGE and probed with anti-TAFI110 antibodies (top panel). (B) U2OS cells were cotransfected with 2 g of the rDNA reporter plasmid pHrP2-BH and increasing amounts of pCMV-FLAG-hTAF12 (indicated on top) in a total amount of 8 g. Reporter transcripts and cytochrome oxidase 1 (cox 1) mRNA were detected using appropriate 32P-labeled riboprobes and quantified (NB). The expression of Flag-TAF12 was verified on Western blots with anti-FLAG antibodies (WB). The bar diagram represents the relative level of reporter transcripts from three independent experiments. (C) TAF12-containing SL1 fractions stimulate RNAP I transcription in vitro. TAF12 copurifies with transcriptionally active SL1 (left panel). HeLa nuclear extracts were chromatographed on phosphocellulose and S-Sepharose. Individual S-Sepharose fractions (20 l of fractions 2 and 6, respectively) were probed for the presence of TAFI110 and TAF12 on immunoblots. RNAP I transcription was assayed in a reconstituted system. The reactions were supplemented with SL1 fractions containing detectable amounts of TAF12 (fraction 2) or fractions with trace amounts of TAF12 (fraction 6). In lane 1, no SL1 fraction was added. The bar diagram represents the relative level of transcription from three different experiments.
Distinct factor profiles on CpG and non-CpG targets

To assess whether the DNA sequence composition such as CpG content specifies transcription factor occupancy or utilization, we filtered out RNAP I and III targets and sorted remaining targets enriched for TBP into two bins: overlapping or non-overlapping with CpG islands. A small number of closely positioned RNAP II/RNAP III promoters remained in the subsequent analyses. About half of the targets ended up in the CpG island bin in line with estimations of the number of genomic CpG island promoters (56%) (Antequera and Bird, 1994). The vast majority of known, annotated RNAP II promoters (84%) were found in the CpG islands bin (Figure 7A). Besides a small number of annotated RNAP II promoters, the non-CpG island bin contained the majority of TBP target sites located in introns or such lacking a gene annotation. Based on the PCA and functional analysis (Figures 2A, 3 and 4), these TBP-binding sites were classified as RNAP II regulatory regions.

Correlation analysis of the CpG-island bin revealed a dendrogram with four main branches (Figure 7B): two closely positioned branches containing the general RNAP II factors (marked in red) and the TAFs (marked in purple). The two other branches were placed opposite to the RNAP II factors and they contained clusters of SNAPc proteins and RNAP III factors. These branches were well structured because of the presence of snRNA genes as well as juxtaposed RNAP II and III promoters. The dendrogram calculated for targets in the non-CpG bin revealed two opposing branches: one branch was well structured and contained the general RNAP II factors (Figure 7C). Surprisingly, TAFs did not cosegregate with the RNAP II factors but were placed at a large distance in the opposing branch that was not well structured and contained RNAP III factors and SNAPc proteins. The opposite positioning of TAFs relative to the other RNAP II factors on the non-CpG TBP-binding sites suggests that TAFs are not efficiently recruited to the non-CpG targets.

Collectively, these data suggest different functional interactions of TAFs and other RNAP II factors on non-CpG versus CpG island targets pointing to distinct mechanisms of transcription initiation.

Discussion

In this study, we used ChIP followed by cloning of the precipitated genomic DNA fragments to identify in vivo TBP-binding sites. The vast majority (90%) of the cloned and filtered genomic fragments appear to be true in vivo TBP-binding sites (Figure 2A). Sequencing and annotation of these sites revealed that a remarkably large fraction (49%) is located in introns of known genes and in genomic locations lacking a gene annotation (Figure 1B). PCA placed these novel TBP-binding sites in the same space as annotated RNAP II targets.

Functional reporter assays revealed that the majority of these novel TBP-binding sites displayed unidirectional transcriptional activity (Figure 3A and B) providing evidence that these novel sites function are genuine promoters. stS-RT qPCR supported and extended this conclusion showing that transcripts could originate from the novel TBP-binding sites in vivo (Figure 4). Collectively, the data provide strong evidence that the majority of novel TBP-binding sites are true functional promoters. A number of the cloned TBP-binding sites displayed significant direction-independent activation of SV-40 promoter fulfilling the criteria of enhancers. The fact that the well-known GADD45 enhancer was also among our TBP-binding sites reinforces the notion that our approach also yielded enhancers. The presence of promoter-specific factors such as TBP and RNAP II on enhancers can be explained by DNA looping (Tolhuis et al, 2002) and crosslinking via protein–protein contacts. An alternative and very intriguing explanation is that a subset of general transcription factors may be directly recruited and assembled onto enhancers and subsequently handed over to the promoter or that some 'enhancers' act as promoters that may help to maintain an open chromatin structure (Szutorisz et al, 2005).
Figure 7. Distinct correlation profiles for CpG and non-CpG island RNAP II targets. (A) Distribution of CpG and non-CpG island targets in the different annotation groups. The CpG islands database was obtained from the UCSC genome browser. (B, C) Rooted trees represent Ward's hierarchical clustering of Pearson correlation values calculated on CpG (B) and non-CpG (C) targets. The branches of TAFs and other general transcription factors are colored in purple and red, respectively.

The remarkably large fraction of novel functional TBP-binding sites in our library indicates that the genome contains many more promoters that have not been identified experimentally or by current annotation algorithms. If this proportion holds true for the entire human genome (50%), the number of functional TBP-binding sites may exceed 80,000 which is roughly 2 more than the number of genes annotated to date. Taking into account the multitude of different tissues and developmental stages, the total number of promoters and enhancers is likely to be significantly larger. Our observations corroborate and extend recent transcriptome and ChIP-on-chip studies that reached similar conclusions (Kapranov et al, 2002; Bertone et al, 2004; Cawley et al, 2004; Cheng et al, 2005; Kim et al, 2005).
Integrated analysis of transcription factors binding profiles

We performed a comprehensive ChIP-on-chip study involving 26 general factors and two histone marks on 1000 experimentally derived TBP-binding sites. To uncover properties that cannot be extracted from individual subsets of data, we analyzed the ChIP-on-chip data in an integrated manner rather than as a collection (summation) of datasets for individual factors. PCA revealed a high intrinsic structure in the data set and segregated the TBP-binding sites into three distinct clusters. One of the advantages of PCA for ChIP-on-chip data analysis is that a 'true–false' threshold does not need to be established for each antibody. The presence of negatives in the data set does not obscure the analysis; on the contrary, it provides a higher level of overall variance that favors segregation of the most similar variables. The segregation of the three major gene classes transcribed by RNAP I, II and III indicates regulation by highly characteristic and distinct combinations of transcription factors.

We also used correlation profiling that calculates the degree of linear relationship between two multitudes of data points, in our case between ChIP/input ratios for different transcription factors. When applied to ChIP-on-chip data, it can be used to determine the degree of similarity/dissimilarity between transcription factors on the basis of their binding profiles on a large number of targets. Like in PCA, a 'true–false' threshold does not need to be established. To organize the multitude of correlation values of the entire data set, we used hierarchical clustering algorithms to calculate the differences between correlations and to convert them into distances so as to build a cluster dendrogram. Analysis of the entire data set revealed four major branches (Figure 5A and B) corresponding to RNAP I, II, and III and SNAPc target genes providing evidence for the involvement of distinct sets of factors in transcription by the three RNA polymerases in vivo. The branches had compact substructures with the exception of the RNAP II branch. The latter displayed a more open branch structure that likely reflects the broad assortment and heterogeneity of multiprotein complexes involved in transcription initiation by RNAP II (Lee and Young, 2000; Naar et al, 2001) as well as the temporally ordered recruitment of factors to heterogeneous RNAP II promoters (Cosma, 2002). High correlation values were obtained for proteins that simultaneously bind the same genomic locations and make long-lived contacts, such as in biochemically stable multiprotein complexes, because they can be co-crosslinked with high probability and efficiency. For example, RPA116 and PAF53 that are both subunits of RNAP I (Seither et al, 1997) or the Bdp1 and Brf1 subunits of the TFIIIB complex involved in RNAP III transcription (Schramm and Hernandez, 2002) have very high correlation values and are placed at short distances from each other in the dendrogram (Figure 5B). The distance between RPA116/PAF53 and Bdp1/Brf1 is, however, very far because the probability and efficiency of co-crosslinking is low or absent as the proteins are part of functionally unrelated biochemical complexes and their genomic binding site repertoires do not overlap. Extending the same logic to TAF12 and PCAF that tightly cluster in the RNAP I branch implies that they can be part of a stable complex that is distinct from the PCAF/STAGA/TFTC complexes. In line with these observations, PCAF has previously been shown by us to acetylate TAFI68 and stimulate transcription of rDNA gene in a reconstituted transcription system (Muth et al, 2001). Here, we provide evidence that TAF12 is also involved in RNAP I transcription (Figure 6). First, overexpression of TAF12 stimulated RNAP I transcription in a cell-based reporter assay. Second, RNAP I transcription was stimulated after supplementing a reconstituted transcription system with a TAF12-containing SL1 fraction. Finally, TAF12 was found in endogenous SL1 complex and physically bound at rDNA promoter. Note that the cluster analysis performed on the subset of CpG promoters resulted in TAF12 and PCAF cluster together with other RNAP II TAFs in line with their role in TFIID and SAGA...
Chapter 2

(Figure 7B). Thus, our data show that TAF12 and most likely also PCAF have dual functions in RNAP I and II transcription.

*Distinct clustering patterns on CpG and non-CpG RNAP II targets*

The primary DNA sequence of promoters plays an important role in recruitment of specific transcription factors. Multiple core promoter elements that are specifically bound by general transcription factors during pre-initiation complex formation have been described. Whether a particular factor is involved in transcription of a given gene class has not yet been addressed in a comprehensive manner in higher eukaryotes.

To assess whether the transcription factor occupancy on RNAP II genes involves distinct subsets of general transcription factors, we performed correlation analysis separately on non-CpG targets and on the targets located in CpG islands. Our correlation dendrograms showed a remarkable difference in the clustering and positioning of TAFs (Figure 7B and C); TAFs were placed at a larger distance from other basal RNAP II factors on non-CpG islands but clustered close on CpG island targets. Our data suggest that TAFs and the other general factors are not or very transiently co-recruited to non-CpG promoters and, therefore, are not efficiently co-crosslinked. TAFs appear to be (more) stably recruited to CpG island promoters, perhaps because these promoters are more active. This assumption is in line with our finding that many novel non-CpG sites show slightly reduced RNAP II occupancy. The virtually identical occupancy values for TBP and TFIIIB on CpG versus non-CpG targets suggest that these novel TBP-binding sites are occupied with the RNAP II machinery. Reporter assays show that most of the non-CpG targets comprise transcription-competent promoters. Thus, it is likely that in analogy to yeast (Basehoar et al, 2004; Huisinga and Pugh, 2004) at least two major pathways of transcription initiation by RNAP II exist in mammals. It will be interesting to extend these observation genomewide and to perform time-resolved ChIP-on-chip following gene activation to unravel the order of factor recruitment.

**Materials and methods**

*ChIP and ChIP cloning*

U2OS cells were crosslinked with 1% formaldehyde for 30 min at room temperature, quenched with 0.125 M glycine and washed at 4°C with three buffers: (i) PBS, (ii) buffer of composition 0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6 and (iii) 0.15 M NaCl in HEG buffer (1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6). Cells were then suspended in ChIP incubation buffer (0.15% SDS, 0.1% DOC, 1% Triton X-100, 150 mM NaCl, HEG) and sheared using a Branson-250 sonicator. Sonicated chromatin was centrifuged for 5 min and then incubated overnight with purified anti-TBP antibody (Diagenode) and protein A/G beads (Santa Cruz). Beads were washed six times with different buffers at 4°C: two times with solution of composition 0.1% SDS, 0.1% DOC, 1% Triton, 150 mM NaCl, HEG, one time with the solution same as before but with 500 mM NaCl, one time with solution of composition 0.25 M LiCl, 0.5% DOC, 0.5% NP-40, HEG and two times with HEG. Precipitated chromatin was eluted with 400 l of elution buffer (1% SDS, 0.1 M NaHCO3), incubated at 65°C for 4 h in the presence of 200 mM NaCl, phenol extracted and precipitated with 20 g of glycogen at -20°C overnight. For sequential ChIP, chromatin was eluted with a small volume of elution buffer, diluted to specific incubation conditions and processed same as that of the first IP with the same amount of antibody.
Transcription factor signatures of human promoters

For cloning, ChIP was performed with 108 cells and DNA obtained after the second ChIP was extracted and treated with T4 DNA polymerase to generate blunt ends, purified, ligated into a pBluescript vector and used for transformation of Escherichia coli.

qPCR
ChIP experiments were analyzed by qPCR with specific primers using a SYBR green kit (Applied Biosystems). Efficiency of ChIP was calculated as percentage of input and specificity - as folds over negative controls (transcriptionally silent genomic loci such as promoters and coding regions of -globin and myoglobin genes). Primers for qPCR were designed with Primer Express and verified by in silico PCR (genome.cse.ucsc.edu/cgi-bin/hgPcr) and by pPCR as amplifying a single specific amplicon. PCR efficiency of primers was calculated with series of 10-times dilutions and accepted when found to be reliable (20.15). Primer sequences are available as Supplementary Table II.

TBP-binding site microarray and ChIP-on-chip
Inserts from the clones obtained in TBP ChIP-cloning procedure were PCR-amplified, purified and used for sequencing and printing on glass slides. Every target was printed six times in different parts of the slide to ensure robustness of the microarray data.
For ChIP-on-chip experiments, ChIP'ed and input DNA was amplified by LM-PCR as described (Ren et al, 2000), labeled with Cy5 and Cy3 using random priming, purified and dissolved in hybridization buffer (33% formamide, 2.5 SSC, 6.6% dextran sulfate). Hybridization was performed overnight at 45°C. Slides were washed at room temperature for 20 min with 0.1 SSC, scanned and analyzed. Median values were calculated for six spots printed on array for each target and the ratios from two hybridizations were averaged. Targets with low intensity (below 2SD of local background) were filtered. The data is available from GEO under accession number GSE6738.

ChIP-on-chip data analysis
The ChIP/input ratios were normalized to the median of four reference controls (promoter and coding regions of myoglobin and -globin genes which were validated as negative by single gene qPCR for the antibodies). PCA and correlation analyses were performed using R software package (www.R-project.org) on data from non-redundant targets enriched for TBP >2-fold. In the final data matrix, all factors were rescaled to have zero mean and unit variance. Up to four PCs were considered in PCA. Pearson correlations were calculated for every pair of transcription factors and hierarchical clustering on these values was performed using Ward's clustering and average linkage. Stability of the clustering dendrograms was established in two ways. First, comparison of structures obtained with Ward's clustering and average linkage revealed significant similarity when calculated at level of 3–5 clusters. Second, leaving each factor out in turn revealed no structural changes in most cases, only for very few factors this resulted in minor changes of the clustering trees.

Promoter/enhancer gene-reporter assays
Genomic fragments of about 1 kb containing the validated TBP-binding sites were PCR-amplified and ligated in front of the reporter gene of pGL3-basic (promoter-less) or pGL3-promoter (SV-40 promoter) vectors. These constructs were transfected into U2OS cells together with pSV2-CAT by calcium phosphate method, gene-reporter activity was measured and normalized to CAT activity. The values were averaged from 2 to 6 replicates. The baseline of reporter gene expression was determined as average of eight transfections of the empty pGL3 vectors.
sts-RT qPCR

One microgram of total RNA isolated from U2OS cells with Trizol reagent (Invitrogen) was treated with DNaseI at 37°C for 20 min followed by inactivation at 80°C for 20 min. Ten picomoles of specific probe was added and denatured at 75°C for 10 min. To obtain high specificity, the reaction was not placed on ice but instead, the temperature was ramped 0.3°C/s down to 60°C and 8 l of reaction mix prewarmed at 60°C was added (3 l of 5 first strand buffer (Invitrogen), 2 l of 0.1 M DTT, 1 l of 10 M each dNTP, 2 l of water), mixed and incubated for 2 min. Then 1 l of heat-stable reverse transcriptase (SuperScript III, Invitrogen) was added, the samples were mixed and incubated at 60°C for 40 min. Then the samples were incubated at 95°C for 15 min to inactivate reverse transcriptase, treated with RNaseH+RNaseA at 37°C for 20 min, diluted 2 and 5 l from the samples were used for qPCR. The results were normalized (% of GAPDH mRNA). The analysis has been repeated twice with different RNA preparations and the results were averaged.

Functional analysis of TAF12

The cDNA encoding TAF12 was inserted into the plasmids pRe/CMV-Flag (Voit et al, 1999) and pGEX-4T3. For reporter assays, 3 105 U2OS cells were cotransfected with a total amount of 8 g of plasmid DNA including 2 g of the RNA reporter plasmid pHrP2-BH, 1 g of pEFGP, to monitor transfection efficiency at the same level, and different amounts of pRe/CMV-Flag-TAF12. RNA was isolated 40 h after transfection, and 5 g of total RNA were subjected to Northern blot analysis (Voit et al, 1999). To normalize for RNA loading, the Northern blots were re-hybridized with a riboprobe for cytochrome c oxidase 1 mRNA.

SL1 was partially purified from HeLa nuclear extracts by sequential chromatography on phosphocellulose P11 (1.5 M KCl fraction) and S-Sepharose (700 mM fraction). SL1 was immunoprecipitated with anti-TAFI110 antibodies for 4 h at 4°C in IP buffer (20 mM Tris–HCl, pH 7.9, 150 mM KCl, 5 mM MgCl2, 0.1 mM EDTA, 20% glycerol, 1 mM PMSF, 0.2% NP-40 and protease inhibitors). Mock IP was carried out in the presence of rabbit IgGs (Dianova). Immunoprecipitated proteins were bound to Protein A-Sepharose (1 h at 4°C), and subjected to Western blotting.

For GST pull-down assays, GST and GST-TAF12 were immobilized on GT-Sepharose and incubated with 20 l of reticulocyte lysates (TNT, Promega) containing in vitro synthesized 35S-labeled transcription factors and 35S-methionine. After incubation in buffer AM-150/0.2% NP-40 (substituted with protease inhibitors) for 4 h at 4°C, beads were washed and eluted proteins were separated by SDS–PAA PAGE and visualized by a PhosphorImager.

In vitro transcription reactions (25 l) contained 20 ng of linearized template pHrP2-BH, 2 l of TIF-IA/TIF-IC (Q-Sepharose-fraction), 10 ng of recombinant Flag-UBF1 purified from Sf9 cells (Voit et al, 1999), 0.5–2 l of SL1 (S-Sepharose fraction), 4 l of RNA polymerase I (H-400 fraction) and transcription buffer supplemented with ribonucleotides as described (Muth et al, 2001).

Antibodies

TAF4, -7 and -12 antibodies were kindly provided by Irwin Davidson; TAF10 and TRRAP by Laszlo Tora; PCAF by Yoshihiro Nakatani; NC2a by Michael Meisterernst. The other antibodies were purchased: TAF1 (sc-735), CBP/p300 (sc-369), TFIIeA (sc-237), RNP II N-terminus of RPB1 subunit (sc-899), E2F1 (sc-251), E2F4 (sc-1082) from Santa Cruz; RNPAP II CTD (8wg16) and
phosphorylated CTD at Ser5 (H14) from BABCO; H3K9ac from Upstate and H3K4me3 from Abcam.

Supplementary data
Supplementary data are available at The EMBO Journal Online (http://www.embojournal.org).

Acknowledgements
We are very grateful to our colleagues Irwin Davidson, Laszlo Tora, Yoshihiro Nakatani and Michael Meisterernst for antibodies. We thank Vera van Noort and Martijn Huynen for help in data analysis. We thank our colleagues for valuable suggestions and discussions. This work was supported by National Scientific Organization (NGI 050-71-016), National Cancer Foundation (KWF-KUN 2005-3347) and HEROIC, an Integrated Project funded by the European Union under the 6th Framework Programme (LSHG-CT-2005-018883).

References
Antequera F, Bird A (1994) Predicting the total number of human genes. Nat Genet 8: 114
Chapter 2


Chapter 3

ChIP-profiling implies differential TFIIA requirement for transcription of human genes

Sergey Denissov, Huiqing Zhou, Marc van Driel, Kees-Jan Francoijs and Hendrik Stunnenberg
Function of TFIIA in transcription

Abstract
The mechanisms of transcription initiation at the promoters of RNAP II transcribed genes have been investigated in a multitude of studies. Several proteins and multiprotein complexes have been characterized as general transcription factors due to their critical involvement in transcription of the majority of cellular genes. The role of the general factor TFIIA in transcription remained to be controversial. TFIIA appeared not to be essential for initiation in vitro, and transcription of the majority of yeast genes did not change upon depletion of cellular TFIIA in vivo. In this study we investigated the role of TFIIA in transcription of human genes via identification and analyses of in vivo genomic binding sites. Correlation analysis based on ChIP-on-chip data revealed that TFIIA and TAFs binding profiles are similar suggesting a tight functional link between these factors. High resolution mapping of TFIIA and TBP binding sites and hierarchical clustering revealed different groups of functionally related genes that may be regulated by TFIIA-dependent or -independent mechanisms. These results provide novel insight in the role of TFIIA in transcription of human genes in vivo.

Introduction
Transcription of eukaryotic genes is a highly complex process which requires a plethora of transcription factors. Transcription by RNAP II proofs to be exceptionally complex. RNAP II transcribes a large number of protein coding and non-coding genes required for various cellular processes. These genes undergo a tight regulation during initiation of transcription that involves the coordinated interplay of many transcription factors, various DNA regulatory elements and chromatin modulations [1-3]. A key step in transcription initiation by RNAP II is the coordinated assembly of the preinitiation complex (PIC) at the promoter involving many multiprotein complexes such as RNA polymerase II, general transcription factors and (co)regulators that can interact with various regulatory DNA elements. PIC is essential for efficient transcription regulation and an accurate positioning of the RNAP II at the correct start site prior to elongation.

Several general transcription factors named TFIIA, TFIID, TFIIB, TFIIE, TFIIF and TFIIH have been characterized which interact with each other and RNA polymerase II [1-3]. Most of them are stable multiprotein complexes that are essential for transcription initiation by RNAP II in reconstituted systems in vitro and also for all the cellular genes in vivo. TFIID is a central complex consisting of TATA-binding protein (TBP) and up to 14 different TBP-associated factors (TAFs) [4]. TBP and TAFs interacts with DNA and with different proteins required for both general and specific regulation of gene transcription [7] and TFIID is regarded as universal factor because it is essential for transcription initiation by all three cellular RNA polymerases on most of the genes [2, 3]. TFIID complex can bind to specific promoter elements and histone modifications, and interacts with large number of transcription factors, activators and RNA polymerase and conducts upstream regulatory signals from activator proteins [3-6]. Different TAF subunits have specialized functions such as transcriptional activation via protein-protein interactions, promoter selectivity via protein-DNA recognition and maintenance of structural integrity of TFIID. In addition, TAFs are also part of several other complexes, such as SAGA and TFTC, that do not contain TBP and are involved in regulation of transcription of specific genes [8].

TFIIA is essential transcription factor which in mammalian cells consists of three conserved subunits (alpha, beta and gamma) corresponding to the 2 subunits (TOA1 and TOA2) in yeast [9-13]. The alpha and beta subunits are post-translationally processed from a single polypeptide via proteolytic cleavage [14]. This process is apparently involved in regulation of stability of the TFIIA complex.
because alpha and beta subunits separately, but not their uncleaved form are substrates for proteasomal degradation [15].

Although TFIIA is referred to as a general factor its precise role in transcription is controversial. In highly purified in vitro transcription assays TFIIA, unlike other general factors, is not required for accurate initiation and basal level of transcription [16, 17]. In the presence of partially purified cell extracts, TFIIA appeared to be essential [17, 18]. It was suggested that TFIIA may act as an antirepressor antagonizing various transcription repressors via blocking or displacing them from the promoter [19-25].

The crystal structure of the TFIIA/TBP/DNA complex has been resolved to investigate the function of TFIIA in the preinitiation complex [26, 27]. The interaction of TFIIA with both TBP and a DNA region immediately upstream of the TATA-box appears to be conserved between yeast and human suggesting that TFIIA can stabilize the binding of TFIID to the promoter and consequently facilitate PIC assembly [19-22, 28], although the crucial requirement of TFIIA for transcription from TATA-less promoters suggests additional mechanisms [29].

TFIIA is directly involved in general transcription in vivo via interactions with TBP, TAF1, TAF4, TAF5, TAF6 and TAF11 subunits of TFIID and the general factors TFIIE and TFIIIF [11, 20, 30-36]. Interactions with several TBP-related factors (TRFs) suggest a general role of TFIIA in transcription at different promoters [37-40]. TFIIA also directly interacts with several activators such as AP-1 and SP-1 and also a number of cofactors suggesting that TFIIA may function as a coactivator, connecting activators and general transcription factors [23, 41-47]. This is supported by mutational analysis in yeast in which disruption of the interactions between TBP and TFIIA results in impairment of transcription in response to activators but has little effect on the overall expression level of cellular genes [51, 52]. Furthermore, upon reduction of endogenous TFIIA protein level in vivo, the majority of the genes do not reveal significant reduction in transcription level [20, 53], although specific genes might be affected because the depleted cells are specifically blocked in G2/M cell cycle phase [33, 53].

To further understand the role of TFIIA in transcription of eukaryotes, we identified the binding sites for TBP and subunits of TFIIA in human U2OS cells using ChIP-on-chip. Correlation analysis of the targets uncovered a high similarity between the binding profile for TFIIA and TAFs suggesting a functional relationship between these factors. We found distinct occupancy profiles for TFIIA and TBP binding sites obtained from a high resolution ChIP-on-chip mapping. These data suggest that there are TFIIA-dependent and -independent mechanisms of transcription initiation which may regulate specific classes of genes in human cells.

**Results**

**Identification of the binding sites for TFIIA and TBP**

Identification of genomic binding sites for transcription factors by ChIP is critically dependent on the performance of the specific antibodies. The antibody against TBP has been extensively characterized in ChIP and ChIP-on-chip in our previous study that led to identification of a large collection of known and novel functional promoters [54]. To investigate TFIIA in more details, we generated antibodies against individual subunits and affinity purified using specifically cleaned antigens (Materials and methods).

Next, we characterized the performance of the TFIIA antibodies in ChIP. We measured the binding of all three subunits to the promoters of a representative set of the RNAP II transcribed genes by quantitative PCR (qPCR). Significant enrichment (up to 25 folds) was found for the alpha and gamma
Function of TFIIA in transcription

subunits and the beta subunit showed lower values (Figure 1A). The enrichment values revealed a high correlation for all the three antibodies (an average $r = 0.94$) showing that TFIIA subunits bind to their target promoters with the same relative efficiency. The enrichment for TBP on the targets was about 2 folds higher then TFIIA alpha/gamma (Figure 1B). Correlation between TFIIA with TBP was lower (an average $r = 0.83$) because several some targets such as nucleolin or RPS19 promoters revealed a low TFIIA binding relative to TBP (3 - 5 times lower then TBP). Collectively, these data show that the TFIIA antibodies are efficient and specific in ChIP and suitable for the identification of genomic binding sites.

**Figure 1.** Binding of TFIIA and TBP at promoters of RNAP II transcribed genes.

Efficiency of ChIP was calculated as % of input for promoter regions of a representative set of genes as well as for non-specific genomic regions (background) such as the promoter and a second exon of the transcriptionally silent myoglobin gene. Enrichment factors were calculated as folds over mean background value.
(A) Enrichment of the three TFIIA subunits at a representative set of promoters. Targets are sorted on descending of the TFIIA alpha values.
(B) Enrichment of TBP and TFIIA alpha. Targets are sorted on descending of TBP values.
Correlation analysis of ChIP-on-chip profiles

To further investigate the role of TFIIA in transcription, we measured the binding of the subunits to a large number of genomic loci using ChIP-on-chip. In these experiments DNA from ChIP was amplified and hybridized on the TBP-binding sites microarrays containing a large number of experimentally derived promoters of RNAP II genes [54]. This TBP-binding sites microarray has been previously used for a comprehensive assessment of co-occupancies of 25 different RNAP I, -II and -III linked transcription factors.

To assess a possible functional relationship of TFIIA with other transcription factors, we performed correlation analyses that determine the degree of similarity-dissimilarity between binding profiles. Correlation values were calculated between every possible pair of factors and then were structured by clustering algorithms. The output of the clustering analysis is a hierarchical dendrogram where length of the branches is a measure of the difference between correlation values (similarity-dissimilarity). In this analysis, ChIP-on-chip occupancy profiles can be compared: transcription factors which are co-recruited to the same targets will have a high correlation and will be grouped together (short distance), whereas factors that do not co-occupy the same targets will have low correlation values and will be placed more distant from each other.

The hierarchical tree calculated on the entire ChIP-on-chip dataset together with TFIIA revealed four major clusters (Figure 2A). The RNAP I-specific factors were clustered at short distances from each other and farthest separated (the most dissimilar) from the other factors. The other branch combined two sub-branches with RNAP III-specific factors, such as Bdp1 and Brf1 and the RNAP III subunit RPC1, and subunits of the SNAPc complex which are specifically involved in transcription of snRNA genes.

The third branch grouped together the RNAP II transcription factors. General factors, such as TFIIB and TFIIE, clustered together, and RNAP II and two histone modifications (H3K9ac and H3K4me3) formed another sub-branch. TAFs clustered in another separate sub-branch which is in line with their presumed function as the subunits of the biochemically distinct complex TFIID. Surprisingly, all TFIIA subunits clustered at the same branch and at short distances from the TAFs (Figure 2A). Such a near positioning implies very similar binding profiles suggesting that TFIIA and TAFs may have a functional relationship and possibly participate in the same regulatory pathway during preinitiation complex assembly at human promoters.

To control whether the functional integrity of the TFIIA complex can be concluded from the ChIP-on-chip profiles, we left out TAFs datasets in order not to affect the clustering of TFIIA and repeated the analysis. We found that all the three TFIIA subunits were grouped in a distinct sub-branch (Figure 2B) showing that they have similar binding profiles, i.e. co-occupy their target promoters to the same relative degree.

High resolution map of TFIIA and TBP binding

To further investigate the involvement of TFIIA in transcription of human genes, we mapped by ChIP-on-chip the binding sites for alpha and gamma subunits as well as for TBP on a 80.3 Mb region on chromosome 6. Using tiling microarrays covering the genomic region with 100 bp density, we obtained detailed binding patterns for these transcription factors. At such a high resolution, the regions of enrichment appear as peaks of about 1 kb in width around the binding sites due to the variable size of DNA fragments generated by sonication (Figure 3A).
Figure 2. Correlation-clustering analysis of ChIP-on-chip datasets.

The ChIP-on-chip datasets for the TFIIA subunits were obtained on the microarray platform as described previously [54]. Pearson correlation values were calculated for every possible pair of factors based on the entire ChIP-on-chip dataset and structured by hierarchical clustering. The branch of RNAP II specific factors is colored in red. Alpha, beta and gamma subunits of TFIIA are indicated with symbols of the Greek alphabet. TBP was excluded from the analysis.

(A) The dendrogram is obtained from the analysis of ChIP-on-chip datasets for all the factors. The sub-branch of TAFs is colored in purple and the branches of TFIIA subunits are colored in blue.

(B) The dendrogram is calculated without ChIP-on-chip datasets for TAFs. The distinct sub-branch of TFIIA is colored in blue.
Chapter 3

Figure 3. Annotation of TBP and TFIIA targets obtained by high resolution mapping. The binding sites for TBP and TFIIA were identified at 100 bp resolution and the distances to the closest transcription start sites of the RNAP II transcribed genes (Ensemble database) were determined. (A) A graphic representation of the high resolution microarray data shows the chromosome region containing histone genes. The ChIP/input ratios of the enriched probes are represented as bars; they form peaks around the binding sites. Y-axis represents a log2 scale of the normalized ChIP/input ratio. (B) The scatter plot represents the distribution of targets over the range of TBP ChIP/input ratios (X-axis, log2 scale) in relation to their distance to the closest TSS (Y-axis, in kilo base pairs, kb). (C) The histogram shows the relative distribution of the targets for TFIIA alpha and gamma subunits (blue and green lines) and TBP (grey line) around transcription start sites. X-axis shows the distance in kilo base pairs (kb).
To characterize the binding sites, we determined their distance to the closest transcription start site of annotated genes. From 707 detected TBP binding sites, 19% (137 sites) represented promoters of tRNA genes, and 43% (304 sites) were located near transcription start sites (around 2 kb) of the protein coding genes and revealed variable ChIP/input ratios (Figure 3B). For 41% of these TBP targets the binding of both TFIIA subunits could be detected and showed distribution of ChIP/input ratios similar to TBP (data not shown). Low intensity peaks of either alpha or gamma subunits alone were detected for an additional 26% of targets suggesting inefficient or incomplete binding of the TFIIA complex or its individual subunits. The majority of the overlapping TBP and TFIIA binding sites were positioned in close proximity to the transcription start sites of the protein coding genes (Figure 3C) implying their co-binding to the promoters of the genes. For 33% of the TBP binding sites located near transcription start sites, significant peaks of either alpha and/or gamma subunits could not be detected.

**Differential binding of TFIIA and TBP**

Mapping of the binding sites showed that a substantial number of the TBP targets had no significant TFIIA binding. To obtain a general overview of relative intensity of all TBP targets we plotted the TFIIA ChIP/input ratios against descending TBP values (Figure 4A). The values for alpha and gamma subunits showed a similar distribution and were scattered around the same regression trend. TBP values followed a different trend that revealed a number of targets in a high range which showed low TFIIA ratios (Figure 4A, B). To validate these targets, we verified the binding of TBP and TFIIA on a number of these sites by qPCR measurement. All targets displayed low enrichment factors for TFIIA (on average 3 folds over background) (Figure 4C). At about 50% of the sites the enrichment of TBP largely exceeded the TFIIA values (from 3 up to 8 times). For the other targets, TBP enrichment was about 2 folds higher similarly as detected on a representative set of targets (Figure 1B).

To further assess the differential binding of TFIIA and TBP, we performed hierarchical clustering to group targets based on the similarity between the ChIP/input ratios. The targets segregated mainly in two groups (clusters) with distant branching in the hierarchical tree (Figure 5A). One branch (marked in red) contained 79 targets of which the majority displayed high ratios for TBP and TFIIA. The other branch (marked in blue) grouped together all the remaining targets (491) with variable enrichment for TBP and low/absent for TFIIA. Thus, the clustering analysis segregated the targets into two main groups that significantly differed in TFIIA and TBP (co)binding.

To investigate a possible functional relationship between the targets within these two groups, we compared the biological function of the corresponding genes based on their gene ontology (GO) annotation. The cluster of targets with high TBP and high TFIIA binding appeared to be highly enriched for genes involved in establishment and maintenance of chromatin architecture, namely the histone genes (Figure 5B). In addition, it contained a number of heat shock protein genes as well as several metabolic enzymes and transcription factors. The second cluster of targets with low/no TFIIA binding was significantly enriched for promoters of transcription factors most of which were zinc finger proteins, and various enzymes, such as kinases, involved in various signal transduction pathways. Thus, these results suggest that the requirement for TFIIA in transcription may correlate with specific gene functions and possibly is linked to different regulatory mechanisms.
Figure 4. Differential binding of TBP and TFIIA.

(A) The graph shows co-distribution of the ChIP/input ratios for TBP and TFIIA subunits. The targets are sorted on descending TBP values. Dash line underlines the targets on the graph with significantly different values for TBP and TFIIA. ChIP/input ratios are shown on Y-axis in log2 scale.

(B) A graphic representation of the high resolution microarray data shows the examples of the TBP binding sites corresponding to low ChIP/input ratios for TFIIA. Chromosome locations are indicated underneath of the peaks. Y-axis represents a log2 scale of the normalized ChIP/input ratio.

(C) Enrichment of the gene promoters with TBP and TFIIA subunits was measured by ChIP. For novel targets the chromosomal positions are indicated (genome assembly hg17).
Function of TFIIA in transcription

**Figure 5.** Clustering analysis of the TBP and TFIIA ChIP-on-chip profiles.

(A) Hierarchical clustering of the TBP and TFIIA targets obtained from the high resolution profiling. The gradient color scale reflects the ChIP/input ratios ranging from zero (blue color) via median value (white color) up to maximum (red color). The median and maximal values were determined individually for the datasets of TFIIA subunits and TBP. Hierarchical tree is shown above the color range graph. The most distant branches are marked in red and blue colors. Absence of TFIIA binding is indicated in grey. The targets corresponding to tRNA genes were filtered out.

(B) GO annotation analysis of the two groups of targets from the most distant branches (as calculated by hierarchical clustering, Figure 5A). In the graph these groups are referred to as low and high TFIIA promoters. Relative abundance of the functionally related genes involved in a certain cellular process is indicated as % of genes taken for analysis.

**Discussion**

In this study we investigated the binding of TFIIA to the target genes to understand the function of this general factor in transcription in human cells. Using specific antibodies against the three subunits of TFIIA complex (alpha, beta and gamma) we found significant and highly correlating enrichment values on a large number of reference promoters of RNAP II transcribed genes (Figure 1A). To assess the binding of these subunits to a large collection of targets we used ChIP-on-chip profiling in combination with correlation and clustering analyses. The correlation approach calculates the degree of linear relationship between the datasets of ChIP/input ratios, i.e. the degree of similarity/dissimilarity between transcription factors on the basis of their binding profiles. The
clustering algorithms bring the multitude of correlation values between different transcription factors into a hierarchical order based on the differences between them. In a cluster dendrogram these differences are converted into distances between the factors and represented as branches of variable length. The correlation-clustering approach allows to assess the functional relationship between different factors in vivo. High correlation values will be obtained for proteins that simultaneously bind the same genomic sites and make stable contacts with each other, such as in multiprotein complexes, because they can be co-crosslinked with high probability and efficiency on DNA. The biological relevance of this approach was extensively discussed previously [54].

The correlation analysis revealed significant similarity between the TFIIA subunits based on the binding profiles to a large number of targets (Figure 2B), indicating that they occupy their target promoters to the same relative extent and that all the subunits are efficiently co-crosslinked to the same DNA locations. Surprisingly, TFIIA co-clustered with TAFs on a distinct sub-branch and not with other general transcription factors (Figure 2A) implying a high similarity between the binding profiles of TFIIA and TAFs. This result suggests that they likely form a stable biochemical complex on promoter DNA and bind to the same sets of targets in vivo. Such a close functional relationship is strongly supported by reported interactions between TFIIA and TAF1, TAF4, TAF5, TAF6 and TAF11 [11, 20, 30 - 35]. It is possible that TFIIA functionally cooperates with TAFs during transcription initiation and associates with TFIIID and/or SAGA complexes [46]. Furthermore, TFIIA and TAFs interact with general transcription factors such as TFIIE and TFIIF and also different activators [3 - 6, 36] suggesting a very similar coactivator functions for TFIIA and TAFs.

**Differential binding of TFIIA and TBP**

As a further step in our assessment of TFIIA function in transcription, we profiled at high resolution the binding sites for alpha and gamma subunits as well as TBP on a portion of chromosome 6. A large fraction (43%) of the TBP binding site was located near annotated transcription start sites and therefore marked promoters (Figure 3), 38% were positioned distantly from annotated promoters, and 19% of the sites comprised tRNA genes. Remarkably, for a significant number of TBP targets TFIIA binding could not be detected or only found for one of the subunits suggesting that TFIIA complex binds inefficiently at these sites, and/or that the ChIP/input ratios are below detection threshold. The significant differences between TBP and TFIIA binding could be validated by qPCR measurement (Figure 4C). Furthermore, the enrichment values at a representative set of promoters revealed that at some sites, such as nucleolin or RPS19 promoters, the binding of TFIIA is significantly lower in comparison with TBP then on others, like histone genes (Figure 1B). The differential binding of TFIIA and TBP suggests that certain genes are little, or not at all, dependent on TFIIA. Because the binding of TBP correlates very well with the binding of RNA polymerase II and likely reflects transcription activity of genes [54], it can be assumed that also certain highly transcriptionally active genes may be TFIIA-independent. Assuming a function link between TFIIA and TAFs, the TFIIA-independent genes may also be transcribed via TAF-independent mechanisms. This is in line with the finding that transcription of a large number of genes in yeast is not dependent on TAFs or shows a variable dependence on individual TAFs [49].

Interestingly, this hypothesis would imply that at some genes TBP functions in transcription initiation independently of the TFIIA/TAFs unlike as in the TBP-TAFs complex TFIIID. Such a mechanism was suggested to take place in human cells [54] and in yeast where a TAF-free form of TBP binds at the target promoters and may cooperate with TFIIA and the TBP-free TAF complex SAGA [46, 50].
Two major subclasses of targets were revealed by hierarchical clustering (the most distant clusters/branches). These two subclusters corresponded to the targets with high and low/absent TFIIA binding (Figure 5A). Remarkably, the genes in these groups revealed a specific involvement in certain biological processes. Particularly, the histone genes appeared to have a high TFIIA binding, whereas the promoters of a significant number of transcription factors and signal-transduction proteins reveal low TFIIA binding. This result may explain the observations obtained in yeast upon depletion of endogenous TFIIA: the surprisingly small decrease in overall mRNA levels [20, 51-53] can possibly be due to unchanged cellular abundance of transcription regulators whose gene may be TFIIA-independent. In addition, unchanged expression of signalling enzymes can explain cell viability upon significant TFIIA depletion, whereas significant reduction in mRNA levels of the TFIIA-dependent histone genes can lead to the shortage of histone proteins and disbalanced chromatin assembly resulting in G2/M cell cycle block [33, 53].

Thus, our investigation of the in vivo binding profiles for TFIIA and other transcription factors substantiated a functional similarity between TFIIA and TAFs and suggested a TFIIA/TAFs-independent transcription initiation mechanism. These observations have to be further potentiated by detailed analysis of expression levels as well as ChIP-on-chip profiling upon stable/inducible TFIIA knockdown. Conclusive data may possibly be obtained from conditional TFIIA knockout in different cell types and/or tissues.

Materials and methods

Antibodies

Polyclonal antibodies against TFIIA subunits were generated in rabbits using the specific parts of the proteins expressed in E.coli: the first 63 amino acids of the alpha and the entire beta and gamma subunits. To obtain a high specificity of these antibodies antigen-affinity purification was performed using the corresponding antigenic proteins which were purified to very high homogeneity via PrepCell system (Bio-Rad) and via reverse-phase column (Pharmacia). The purified proteins were freeze-dried and then covalently bound to the reactive beads (Affigel-10/15, Bio-Rad) under non-aqueous conditions (DMSO). Then the beads were blocked with ethanolamine and stored in PBS buffer. A fraction of the beads containing 100 micrograms of antigenic proteins was incubated for 4 hours at room temperature with 2 ml of rabbit antiserum. The antiserum was preliminarily cleaned from proteins by caprylic acid precipitation and dissolved in PBS supplemented with 0.01% triton X-100. Then the beads were washed with cold buffers: 2 times with PBS, 4 times with RIPA and 2 times with PBS. The antibodies were eluted 2 times during 5 minutes with 600 ul of 0.1M glycine pH 2.4 and the low pH was immediately neutralized with Tris-base. Assessment of the antibody specificity by western blot revealed major bands of the corresponding size for all the subunits.

The TBP antibody (SL30) was obtained from Diagenode.

ChIP, quantitative PCR (qPCR) and ChIP-on-chip with TBP-binding site microarray

The experiments were performed with U2OS cell line and procedures and data analyses have been performed as described in our previous study [54]. Primer sequences are available in Supplementary Table S1.

ChIP-on-chip with high resolution DNA microarrays and target annotation

DNA from ChIP and input samples was amplified by T7-linear amplification method and submitted for labelling and hybridizations on DNA microarrays at NimbleGen Systems. ChIP/input ratios were
normalized and the binding sites were detected by the Mpeak program [48]. The list of the peaks is shown in Supplementary Table S2. For annotation of the targets the Ensemble collection of genes and TSSs was used. The targets corresponding to tRNA genes were filtered out. GO annotation was performed with FatiGO+ tools.

Acknowledgments
We thank our colleagues for valuable suggestions and discussions. This work was supported by National Scientific Organization (NGI 050-71-016), National Cancer Foundation (KWF-KUN 2005-3347) and HEROIC, an Integrated Project funded by the European Union under the 6th Framework Programme (LSHG-CT-2005-018883).

References
11. Yokomori K, Admon A, Goodrich JA, Chen JI, Tjian R. Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. Genes Dev. 1993. 7: 2235-2245
19. Imbalzano AN, Zaret KS, Kingston RE. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. J Biol Chem. 1994. 269: 8280-8286
43. Chi T, Carey M. Assembly of the isomerized TFIIA-TFIID-TATA ternary complex is necessary and sufficient for gene activation. Genes Dev. 1996. 10:2540-2550
44. Lieberman PM, Berk AJ. A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIIA-TFIID-promoter DNA complex formation. Genes Dev. 1994. 8:995-1006
Chapter 4

The binding profiles of RNAP I transcription factors suggest a unique three-dimensional structure of the active rDNA gene

Sergey Denissov, Marc van Driel, Christine Mayer, Ingrid Grummt and Hendrik Stunnenberg
Abstract
Transcription of the rDNA gene by RNAP I is a highly efficient process. A single gene can be transcribed by more than a hundred of RNA polymerases simultaneously and the rRNA transcripts are efficiently assembled into ribosomes. Although a wealth of studies describe the structure and organization of the nucleolus, the topology of an active rDNA gene within the nucleolar environment remains elusive. In this study we present the binding profiles of different transcription factors throughout the rDNA locus. We find that TBP and the TAF I110 subunits of human SL1 complex are efficiently crosslinked to the enhancer and R3 terminator region in addition to their association with the promoter. Efficient cross-linking is also detected at different sites within the transcribed region of the rDNA gene and is, surprisingly, increased upon H2O2 treatment. Binding of TBP is detected in different phases throughout the cell cycle providing molecular evidence for the association of SL1 with nuclear organizer regions in metaphase. We propose a model for the spatial organization of an active rDNA gene in the nucleolus that integrates our binding profiles and accumulated knowledge about rRNA transcription and provides solutions for many unresolved questions about nucleolar functioning.

Introduction
Actively growing cells require thousands of ribosomes per minute, or up to several millions per cell cycle [1-4]. A proportional number of building components, i.e. the more than 80 different ribosomal proteins and four ribosomal RNA (rRNA) species, are required for the assembly of such a large number of ribosomes. Thus, it is not surprising that the rRNA comprise about 50% of the cellular transcriptional output.

Transcription of the rRNA gene by RNA polymerase I (RNAP I) and processing of the 45S precursor rRNA transcript into the 28S, 18S and 5.8S rRNA molecules takes place in the nucleolus, a non-membrane subnuclear compartment [4]. A diploid human genome contains about 400 copies of the rRNA gene which are grouped in tandem repeats on five acrocentric chromosomes [5]. These groups are named nucleolus organizer regions (NORs) and serve as the sites of nucleoli assembly. The nucleolus can be divided into three sub-compartments visible as morphologically distinct areas in electron microscopy (EM). The central part is called fibrillar center (FC) and appears as a transparent circular area of a variable diameter [6-10]; it contains extended DNA fibrils of the rRNA gene but few proteins and little RNA [11, 12]. RNAP I transcription factors and other proteins such as topoisomerase I are mapped in situ to FC suggesting that FC is the place of an active transcription [4, 11, 13]. The electron dense narrow area at the periphery of the FC is named dense fibrillar components (DFC). It is composed of tightly packed ribonucleoprotein structures (RNPs) that are involved in the processing of the newly synthesized pre-rRNA [4, 13-16]. The processed rRNA molecules are transported to the third nucleolar compartment named granular components (GC), an extended area around the DFC that consists of granules of about 15 nm in diameter being pre-ribosomal particles at different stages of assembly [15, 17-19].

The human 43 kb rRNA repeat unit consists of two parts: a transcribed unit of about 13 kb and an intergenic non-transcribed spacer (~30 kb) [20, GeneBank U13369]. The transcribed unit contains a 5’ and 3’ external and two internal transcribed spacers that separate the sequences of 18S, 5.8S and 28S rRNA molecules. The promoter of the rRNA gene consists of a core element at position ~45 to +18 relative to the start site and an upstream control element (UCE) at ~156 to ~107 [2, 21]. In addition, there is an upstream enhancer region that activates rDNA gene transcription [22].
Termination of the RNAP I transcription occurs at the 3′ end of the rRNA gene at a specific region consisting of the three clusters of specific elements (R1, R2 and R3) called Sal boxes [23, 24]. Transcription initiation requires the assembly of a pre-initiation complex comprising of RNAP I and a number of other transcription factors at the promoter. Promoter selectivity factor 1 (SL1 in human cells) is a protein complex consisting of TATA-binding protein (TBP) and several TBP-associated factors (TAFs) such as TAF$_{110}$, TAF$_{63}$ and TAF$_{48}$ [25] that specifically recognize the core promoter [26, 27]. Upstream binding factor (UBF) binds to the UCE as a dimer via high-mobility-group (HMG) domains and facilitates formation of a pre-initiation complex by interacting with RNAP I and SL1 [28-34]. UBF is thought to functions as an architectural protein at the promoter as well as throughout the entire rDNA transcribed unit maintaining the structural integrity of the NORs [29-38]. The regulatory factor TIF-IA functions at the promoter via interaction with initiation-competent forms of RNAP I and TAF subunits of SL1 [30, 39-41]. When RNAP I enters into the elongation phase, TIF-IA dissociates from the polymerase while UBF and SL1 remain at the promoter to support multiple rounds of transcription initiation.

Elongation of RNAP I is a highly efficient and dynamic process [4, 42]. Cytological studies revealed that a large number of polymerases (50 - 100) can be simultaneously engaged in active transcription of a rRNA gene. In the EM, this can be observed as fir-tree like structures, also called Miller spreads, that represent an extended rDNA genes associated with elongating polymerases connected to nascent pre-rRNA transcripts of increasing length radiating away [4, 43]. The transcribed region of active rDNA genes does not appear to contain nucleosomes and is not subjected to compaction in metaphase [10]. How these structures that largely exceed the diameter of nucleoli are packed remains unresolved. Furthermore, the three-dimensional topology of an active rRNA gene and where is it positioned within the nucleoli remain unclear. Morphological studies showed that the ‘fir-trees’ exist as highly compact RNP structures which are rather located in the DFC as this compartment contains a large number of electron dense RNPs associated with nascent pre-rRNA transcripts [13-15, 44-49]. However, the dense structure of the DFC masks the fir-tree preventing its visualization. Resolving the topology of an active rRNA gene is critical for understanding the mechanisms of the highly efficient, simultaneous synthesis and export of hundreds of the rRNA transcripts.

In this study we present chromatin immunoprecipitation (ChIP) data revealing the profiles of different RNAP I transcription factors throughout the human rDNA gene. The SL1 complex strongly crosslinks to the promoter, as well as to the enhancer and terminator regions. We show that this crosslinking is maintained during the cell cycle. Furthermore, occupancy of the SL1 complex is detected throughout the rDNA coding region and surprisingly, is increased upon suppression of RNAP I transcription by peroxide treatment. Based on our data we propose a three-dimensional structure of the actively transcribed rDNA gene. Our model integrates available data and implies a simple and efficient way for the organization of the highly productive rRNA transcription process.

Results

TBP is crosslinked throughout the rDNA gene

In a previous study we identified in vivo binding sites for TBP in human genome using an optimized ChIP procedure followed by direct cloning and sequencing of the precipitated DNA fragments [50]. We found a large number of functional promoters of novel genes in addition to annotated genes transcribed by RNAP II and -III. About 10% of the total number of clones contained sequences corresponding to rRNA gene as identified by NCBI BLAST.
Here we performed a detailed characterization of the rDNA clones by aligning their DNA sequences to the entire rDNA unit. Surprisingly, these cloned DNA fragments aligned not only with the promoter region but also with the transcribed rDNA locus (Figure 1A). Analysis of the overlapping sequences showed that about 60% of all the clones segregated into three major contigs corresponding to the promoter, the enhancer and the terminator R3 regions containing 25, 34 and 21 clones, respectively. The remaining sequences were distributed throughout the transcribed region with a slightly increased abundance at sites within 18S rRNA sequence. We did not obtain clones that correspond to the non-transcribed intergenic spacer.

**Figure 1.** Analysis of the TBP-binding sites obtained from ChIP-cloning. (A) The cloned sequences were aligned with rDNA repeat unit to identify their coordinate on the rDNA gene. The frequencies are presented as a number of aligned clones per every 500 bp. The +1 coordinate corresponds to the transcription start site (indicated as an arrow). The rDNA gene and its structural elements are schematically represented underneath of the frequency plot. (B) The diagram represents a frequency of the cloned TBP-binding sites (light grey bars) at nine regions throughout the rDNA gene. Occupancies of TBP at these sites (dark grey bars) were measured by ChIP and represented on independent scale.

These data suggest that TBP is efficiently crosslinked throughout the active rDNA gene. To validate this hypothesis we designed specific primers for several sites in the transcribed part of the rRNA locus and performed targeted ChIP and quantitative PCR (qPCR) to determine TBP occupancies at these sites. Significantly, TBP occupancy was detected at all tested positions within the transcribed region with the highest occupancy values at the promoter, enhancer and terminator R3. Occupancy of
TBP within the rDNA transcribed region was not evenly distributed but displayed higher values at 18S rRNA region (Figure 1B). The relative occupancy values correlated very well with the numbers of cloned sequences at the corresponding sites (Figure 1B). Thus, our results show that TBP can be efficiently crosslinked throughout the transcribed region and with very high efficiency at the promoter, enhancer and terminator R3.

It was possible that the observed TBP crosslinking outside of the promoter/enhancer locations was an artifact of the relatively long crosslinking time used to increase the yield in the ChIP-cloning procedure. To control for this possibility we also determined the occupancy of TBP at the various rDNA sites using very short times of formaldehyde treatment (2 and 5 min). As expected, the recoveries decreased proportionally as a result of less extensive crosslinking (Figure 2A). A reduced crosslinking was also observed at other genomic regions such as 2nd exon of myoglobin and intergenic spacer of the rRNA gene that served as the reference baseline (background) (Figure 2A). However, the relative occupancy normalized to the background remained high (Figure 2B). This result indicates that TBP can be efficiently crosslinked over the transcribed rDNA gene independently of the crosslinking time.

**Figure 2.** Binding of TBP and TAF110 at different sites throughout the rDNA gene. (A) Cells were treated with formaldehyde for 30, 5 and 2 minutes and used for chromatin immunoprecipitation with TBP antibody. Corresponding time points are indicated with the shades of grey. Efficiency of TBP ChIP is presented as recovery values (% of input). Recovery values for three genomic loci used as reference controls (negative background) are presented on a graph with separate scale. (B) Relative occupancies of TBP were calculated by normalization of the recovery values (shown on Figure 2A) for the rDNA sites versus background values averaged over the three sites. (C) Relative occupancy of TAF110 at the rDNA sites.
In our previous study, we showed that the proteins that are an integral part of one biochemical complex can be efficiently crosslinked to the same DNA sites [50 and references therein]. Extending this finding to rRNA transcription we now tested whether the factors that stably interact with TBP displayed also a similar crosslinking profile. As shown on Figure 2C, the promoter, enhancer and terminator R3 displayed higher occupancy for TAF$_{110}$, one of the subunits of TBP-containing SL1 complex, than other sites. These data correlate well with occupancy of TBP. The highest occupancy of TAF$_{110}$ was observed at promoter in line with their direct interaction [26]. Collectively, our results show that the promoter-specific transcription complex SL1 can be efficiently crosslinked at different regions throughout the rRNA gene and not only the promoter. The nearly equal occupancy of TBP and TAF$_{110}$ at the promoter, enhancer and terminator R3 suggest stable contacts of SL1 complex with these DNA sites. Low but very significant occupancy at other sites is probably caused by intermediate contacts via network of protein-protein interactions within the three-dimensional structure of the rDNA unit.

**Binding of TBP during the cell cycle**

The level of rRNA transcription changes significantly during the cell cycle with a maximum in the S and G2 phases and suppression in metaphase [4, 29, 30]. A remarkable feature of rDNA genes is that their DNA is not subjected to compaction during chromosome condensation [10, 52, 53]. Furthermore, the transcription factors SL1 and UBF, and also RNA polymerase I remain associated with NORs during mitosis [54-57].

![Figure 3](image)

**Figure 3.** Occupancy of TBP on the rDNA gene in different phases during cell cycle. Cells were synchronized in four cell cycle phases (M, G1, S and G2) and used for ChIP. The occupancy values for TBP in these phases are indicated with the shades of grey.

To investigate these unique features, we assessed whether the binding of TBP at the rDNA locus is altered during the cell cycle by performing ChIP on synchronized populations of cells representing the different phases M, G1, S and G2. In metaphase cells, a high occupancy of TBP was detected at both the promoter and terminator, whereas the occupancy at the enhancer was significantly lower and comparable with the occupancy measured at the transcribed region of the rDNA locus (Figure 3). The occupancy of TBP gradually increased at the promoter and terminator during G1, S and G2, and at enhancer region TBP occupancy reached similar values as at the promoter and terminator. This result correlates with the data describing transcriptional activity of the rDNA gene during the cell cycle [4,
The occupancy at the rDNA transcribed region remained largely unchanged during the entire cell cycle. Collectively, these data provide molecular evidence for the stable association of TBP with active rDNA genes during mitosis.

**Occupancy profiles of RNAP I transcription factors**

The binding of SL1 to regions outside the rDNA promoter hints at a particular spatial conformation of the transcribed rDNA locus. To gain further insight, we measured the occupancy of other RNAP I transcription factors at the rDNA sites. We found that TIF-IA binds predominantly at the promoter (Figure 4A). This is in line with its reported function in the initiation process i.e. TIF-IA transiently interacts with initiation-competent RNAP I and the promoter-specific SL1 complex [39-41]. Lower but significant occupancy could be also detected at the other rDNA sites.

![Figure 4](image)

**Figure 4.** Occupancies of different RNAP I transcription factors on the rDNA gene. Occupancy values of TIF-IA, RPA116 and PAF53, and UBF are presented on figures A, B and C, respectively.

The two subunits of RNA polymerase, RPA116 and PAF53 [34], displayed a high occupancy at all sites within the transcribed region (Figure 4B) in agreement with the observed high density of elongating polymerases on the rRNA gene [42, 43]. High occupancy was also found at the promoter indicating that RNAP I is a part of the assembled preinitiation complex. The binding was observed also at the enhancer region in line with the previously suggested architecture of the rDNA promoter [36]. Low occupancy values were detected at the terminator region R3 indicating that the RNA polymerase was mostly dissociated before it reaches this region of the rDNA gene.
High occupancy of UBF was detected at all the positions (Figure 4C) in line with previous data suggesting that UBF has a structural role in establishing NORs via binding to multiple sites at rDNA [37, 38]. The promoter region revealed the highest occupancy which corresponds well to the reported function of UBF as a promoter-specific transcription factor interacting with TIF-IA and SL1 as well as with its proposed architectural role in promoter DNA bending [35, 36]. Similarly high occupancies were found at the 5’ external transcribed spacer.

**Peroxide treatment alters the binding of SL1 to rDNA gene**

As a next step we monitored the occupancy of transcription factors at rRNA gene upon treatment of the cells with peroxide that is a protein denaturing agent that efficiently reduces RNAP I transcription [58]. The oxidative stress upon peroxide treatment activates JNK kinase pathway resulting amongst others in TIF-IA phosphorylation and its transcriptional inactivation [58]. We performed ChIP on cells treated for 60 min with 0.2 mM H$_2$O$_2$ resulting in a 75% reduction of the level of rRNA (data not shown). The reduction of the occupancy of elongating RNAP I at the rDNA transcribed region but not at promoter/enhancer (Figure 5A) indicated that this is likely due to decrease in initiation caused by TIF-IA inactivation. In contrast, binding of UBF at rDNA was not affected by the treatment (Figure 5B). The occupancy of TBP on the rDNA promoter, enhancer and terminator was also not significantly affected but, surprisingly, at the transcribed region TBP occupancy was very significantly increased (5 - 10 times) upon treatment and became almost similar to that at promoter (Figure 5C). Similar result was also observed for TAF$_{110}$ (Figure 5D) and TIF-IA (Figure 5E).

To control whether this effect of peroxide resulted from transcription inhibition itself or was mediated via other mechanisms (e.g. protein denaturing), we performed ChIP with cells treated with anisomycin which also inhibits RNAP I transcription via JNK kinase pathway [58]. After treatment for 60 min with 10 µM anisomycin the level of rRNA was reduced by about 40% (data not shown). In line with this, the occupancy of RNAP I was also decreased at promoter, enhancer and at transcribed region (Figure 6D). Similarly, occupancy of TIF-IA and TBP decreased at promoter and enhancer and slightly decreased at the transcribed region (Figure 6D and C).

Collectively, these data suggest that peroxide may cause certain structural/organizational changes in the rDNA locus causing an increased binding of SL1 to the rDNA coding region and a reduced transcriptional activity.

**Discussion**

In this study we present the analysis of the binding patterns of TBP and other RNAP I transcription factors at the rDNA gene. The majority of rDNA sequences obtained from ChIP-cloning of TBP-binding sites [50] segregated into three contigs corresponding to the promoter, enhancer and terminator R3 regions; the remaining cloned fragments were scattered throughout the coding region of the rDNA gene (Figure 1). We show here that this result correlates very well with the occupancy of RNAP I transcription factors at these different sites: a high occupancy of TBP and TAF$_{110}$ subunits of the SL1 complex was observed at the promoter, enhancer and terminator region R3. In addition, significant occupancy was detected throughout the transcribed region of the rDNA (Figure 2). The binding profiles of RNAP I subunits, TIF-IA and UBF (Figure 4) were found to be in accordance with published results and their functions in transcription of the rRNA gene.
Chapter 4

Figure 5. Occupancy of transcription factors on the rDNA gene upon treatment with H$_2$O$_2$.
Normally growing cells and cells treated with peroxide were used for ChIP. The occupancy values for the corresponding samples are indicated by shades of grey for RPA116 (A), UBF (B), TBP (C), TAF1110 (D), and TIF-IA (E).

The interpretation of these data requires a careful consideration of the principle of formaldehyde crosslinking. Formaldehyde reacts with amino- and imino-groups at proteins and nucleic acids generating covalent bonds between them [59], i.e. fixes protein-protein and protein-nucleic acid interactions. In ChIP, the efficiency of crosslinking correlates with the occupancy and is dependent on the stability of the interactions as well as the proximity of the crosslinking targets in time and space [50]. Stable interactions, such as those found in stable DNA-bound multiprotein complexes, will be fixed with high efficiency and revealed in ChIP with high occupancy values. Proteins located far from DNA or connected to DNA via instable transient contacts with other proteins, will be crosslinked with low efficacy resulting in low occupancy values. For example, the similar binding profiles of RPA116...
and PAF53 at the rRNA gene reflect their interaction as subunits of the stable RNAP I complex (Figure 4B). Similarly, both TBP and TAF110 as subunits of the stable SL1 complex display concordantly high occupancy values at the promoter (Figure 2).

Extending this reasoning, the high occupancies of TBP and TAF110 at sites other than the promoter, such as the enhancer and terminator R3 regions, suggest stable and close contacts of these regulatory regions with SL1 complex and that during active rRNA synthesis these regions remain in close proximity to each other and to SL1 complex facilitating their efficient crosslinking. The intimate contact between promoter and enhancer may be not so surprising because UBF is reported to mediate bending of promoter DNA and may facilitate the positioning of enhancer and promoter in close proximity to each other [36]. The close connection of SL1 with the R3 region of the terminator is a novel and unexpected finding hinting at architectural interactions in the active rRNA gene. It is possible that SL1 serves as a binding core that interacts with the promoter, enhancer and terminator sites holding together the ‘triad’ loci.

Lower occupancy values of TBP and TAF110 can be detected throughout the rDNA transcribed region (Figure 2) possibly as a result of a distant positioning of SL1 from this DNA and crosslinking via chains of intermediate interactions with other proteins such as UBF and RNAP I [30, 35]. Noticeably higher occupancy at the region of the 18S rRNA (Figure 2), correlating with the high

**Figure 6.** Occupancy of transcription factors on the rDNA gene upon treatment with anisomycin. Normally growing cells and cells treated with anisomycin were used for ChIP. The occupancy values for the corresponding samples are indicated by shades of grey for RPA116 (A), TIF-IA (B), TBP (C).
frequency of the 18S rDNA fragments in ChIP-cloning (Figure 1), suggests the spatially closer positioning of the SL1 subunits to this region then to the surrounding sites. Interestingly, we found that the interaction of TBP with rDNA is stably maintained during the cell cycle. In metaphase, we observed high occupancy values at the promoter and terminator but low occupancy at the enhancer (Figure 3). This indicates that the enhancer region is not very stably associated in M phase and hence can not be efficiently crosslinked to the SL1. Throughout the G1, S and G2 phases the occupancy of TBP on the enhancer remains high in line with the reported activation of rDNA transcription [4, 29, 30]. These results suggest that the disruption of the interactions between the enhancer and the SL1 complex may function as a mechanism of transcription suppression in metaphase. Reestablishing the interactions in G1 allows to resume transcription of the rDNA gene.

Additionally, we observed a surprising effect on occupancy of the SL1 subunits upon H2O2 treatment. The treatment induced a strong increase, rather then decrease of TBP binding throughout the rDNA transcribed region, whereas the occupancy at the promoter, enhancer and terminator R3 region remained virtually unchanged (Figure 5). Similar result was obtained for TAF110 and the promoter-specific factor TIF-IA. This effect may be explained by the appearance of long-lasting contacts between the promoter-bound SL1 complex and the rDNA transcribed region. Peroxide is likely to induce topological changes in the nucleolar compartments so that the SL1 and rDNA coding region are brought in closely proximity from each other. This effect of peroxide is likely mediated by its protein denaturing effects rather then the transcriptional block itself, because in the control experiment the inhibition of RNAP I transcription by anisomycin resulted in overall decrease of factor occupancies (Figure 6).

A spatial model of the actively transcribed rDNA unit

The spatial organization of the transcribed rRNA gene has been a matter of debate for many years and essentially has remained unresolved. Based on our observations we propose a model which (i) fits our novel finding of SL1 occupancy throughout rDNA, (ii) describes a steric positioning for numerous rRNA transcripts allowing their co-transcriptional release for processing, (iii) integrates our current knowledge of the organization and functions of the nucleolar sub-compartments and the localization of transcription factors, DNA and RNA within the nucleolus. To model a spatial conformation of the rDNA gene we addressed several questions in a logical order. First, based on our results from TBP ChIP-cloning and SL1 occupancy ChIP measurements we conclude that the promoter, enhancer and terminator R3 regions are spatially located in very close proximity with SL1 complex and may form a specific structure which we named ‘central tetrad’. Next, we aimed to resolve how the SL1 complex may be crosslinked to the transcribed region of the rDNA gene. Obviously, this implies a certain degree of spatial proximity, but faced the problem how SL1 can be closely located to the entire rRNA gene at the same time? We reasoned that this could only be possible by invoking a compact and rather uniform positioning of the rDNA around the central tetrad. A sphere or cylindrical forms could fulfill the requirement and we considered the cylindrical shape as more preferable because it would allow easy access for transcription factors to the promoter and enhancer. Moreover, the occupancy of SL1 is not uniform over the rDNA transcription unit as would be expected in case of a sphere, but higher at the 18S rRNA region. Furthermore, cylindrical shapes have been observed in recent electron microscopy study [11]. Thus, we propose that the rDNA transcription region is wrapped as a cylinder around the central tetrad located inside of the cylinder. This organization allows contacts between the centrally located SL1
complex and different regions of the rDNA as well as an unrestricted access of transcription factors to the central tetrad.

Next, we aimed to model the mode in which the rDNA may be organized within the cylindrical shape. The simplest way would be a spring-like structure in which the DNA fibril is lined up into a consecutive row of non-intersecting rings (Figure 7A). In this way, two DNA fibrils should enter into the rDNA spring on both sides to place the promoter, enhancer and terminator regions in close proximity to each other and to allow formation of the central tetrad. The model provides an explanation of the increased cross-linkability of SL1 to the rDNA coding region upon peroxide treatment: it seems likely that H₂O₂ destabilizes a protein structure holding DNA in a spring-like structure resulting in a collapse of the rDNA onto the central tetrad located in the core of the spring.

**The spring-core rDNA structure during transcription**

The spring-core model resolves several long-standing issues concerning the spatial organization of the rRNA transcription. For example, it is yet unclear how the long rRNA transcripts synthesized simultaneously can be very efficiently disentangled from the crowd of about 100 other nascent rRNA transcripts and rDNA fibrils. Another question is how so many RNA polymerases can travel together along the rDNA dragging the long pre-RNA transcripts associated with the large processing RNPs. Based on EM studies, these RNPs are located in the protein-dense DFC and unlikely can move at high rate. Additionally, the apparent simplicity in the transcription ‘fir-trees’, in which after deproteinization all the nascent transcripts are easily separated from each other [4, 43], suggests a unique spatial organization in which all the individual transcripts are isolated from each other and are not intertwined.

Our spring-core model proposes a simple and efficient way to keep the rRNA transcript separated and prevented from intertwining. We suggest that the elongating RNAP I molecules do not need to travel over entire rDNA unit and pull the nascent transcripts. Instead, the RNA polymerases can stay at one place and synchronously pull the rDNA over. The nascent rRNA transcripts in this case are spatially isolated during synthesis and can radiate away and associate with processing RNPs. Such a spatial organization requires no especial mechanism for disentangling the RNA and DNA and allows the simultaneous transcription by many polymerases in a highly organized manner.

This pulling force of the RNA polymerases results in rotation of the entire spring-core structure (Figure 7B). In this model, the RNA polymerase starts from the promoter inside of the rDNA spring, continues elongation at the outside of the spring moving along one face, and finally drops off at the end of the spring at the terminator R1 (Figure 7B). The elongating RNAP I molecules can be aligned into rows as it was observed recently [11]. The tracks or rows can be formed by nuclear actin; the interaction of RNAP I with actin has been shown to be essential during rRNA transcription elongation [51, 60]. The entire spring-core structure may be stabilized by UBF that was found to be as a major factor in organization and maintaining of a NOR structure [38].

The rotation of the rDNA spring by RNA polymerases obviously will result in torsional stress on the DNA fibril around the transcribing RNAP I as well as in the intergenic spacer being positive on one side of the spring and negative on the other. These effects can be eliminated by DNA Topoisomerase I which can relax both negative and positive supercoils and also release topological changes in the rDNA during transcription. In line with this assumption, it is known for a long time that the majority of the cellular Topoisomerase I is located in the nucleolus [62, 63].
Figure 7. The spring-core model of a spatial conformation of the active rDNA gene.

(A) Schematic representation of the spring-core structure. SL1 complex (the blue oval) bound to the promoter, enhancer and terminator R3 regions (indicated as colored boxes and marked by P, E and T, respectively) comprises a central tetrad. Spring structure is formed by the rDNA transcribed region that is wrapped around the central tetrad located in a core of the spring. The circles of the spring that comprise 18S rRNA gene are positioned as the closest to the SL1 complex. Two DNA fibrils enter into the core of the spring from both sides.

(B) The scheme describing a functioning of the spring-core model during rDNA transcription. RNA polymerases (green spheres) start elongation in the core of the spring at the central tetrad (SL1 complex is shown as blue oval). Then they move out and continue transcription at the surface of the spring. During elongation polymerases pull over the rDNA resulting in rotation of the spring-core structure. The polymerases while transcribing change relative position along the side of the spring as indicated with arrow underneath. Nascent pre-rRNA transcripts (zigzag black line) of increasing length are attached to RNA polymerases. These transcripts are assembled into processing RNPs (light brown spheres). After termination, RNAP I dissociates from the rDNA (the group of green
pieces) and completed rRNA transcripts follow the further processing steps as RNPs complexes. The dark-red spheres represent Topoisomerase I. This enzyme functions to remove torsional strain obtained in the DNA of the non-transcribed intergenic spacer due to the rotation of the rDNA spring structure. The dark-red spheres represent Topoisomerase I.

(C) Schematic representation of the nucleolar sub-compartment and their correspondence to the spring-core structure (frontal projection). Central core area of the rDNA spring structure corresponds to the fibrillar components (FC). SL1 transcription initiation complex and DNA of the central tetrad (dashed circle) are located in the center of the FC. rDNA transcribed region (circle black line) and elongating RNA polymerases (green spheres) are situated at the periphery of the FC around the border between FC and DFC. Blue spheres represent ubiquitous location of UBF molecules which associate with entire rDNA gene. Nascent pre-rRNA transcripts attached to the polymerases enter into dense fibrillar components (DFC) where they associate with processing RNP (orange spheres). Grey circle around DFC represents granular components (GC).

The spring-core model within the nucleolus structure

The spring-core model allows integration of the nucleolar morphological structure and functions of its sub-compartment. We suggest that the core of the spring is formed by one rDNA gene and corresponds to the fibrillar component (FC) sub-compartment (Figure 7C). In support of this hypothesis, the FC was found to contain only one rDNA gene [11, 12]. Second, using a BrUTP run-on labelling approach, the sites of transcription initiation were mapped inside the FC [11, 13] strongly suggesting that the rDNA promoter, or a central tetrad in our model, is also located within the FC. Third, high resolution in situ studies showed that the RNAP I transcription factors including topoisomerase I are predominately localized within the FC [11-14, 61-63], that is again in line with our data which places the central tetrad inside the FC. Fourth, our model can explain the nature of the extended nucleosome free DNA fibrils positioned in the central part of the FC [10]. This nucleosome free DNA does not correspond to non-transcribed intergenic spacer or inactive rDNA gene [64] but its precise identity and functions were unresolved yet. We suggest that this DNA corresponds to the central tetrad and to the DNA fibrils connected to the tetrad. Indeed, the SL1 complex is localized inside the FC in situ [11-13], and by our ChIP it was shown to bind to the promoter, enhancer and terminator (Figure 1). Collectively, these data strongly suggest that the intra-FC DNA corresponds to the rDNA regions of the central tetrad. Furthermore, it is known that in metaphase, when rRNA synthesis is suppressed, these DNA fibrils are not subjected to compaction via association with nucleosomes [10, 52, 53] and the SL1 and UBF transcription factors remain associated within the FC of the NORs [54-57]. These facts, in combination with our ChIP data showing the association of TBP with rDNA in metaphase, reinforce our notion that the DNA filaments in the center of the FC correspond to the place of SL1 binding, i.e. central tetrad. Collectively, these data support the assignment of the FC and the tetrad to the core of the rDNA spring structure as proposed in the model.

If the spring-core structure indeed corresponds to the FC, then the area around it must be the DFC. This implies that the rDNA in the shape of the spring together with associated elongating RNAP I should be located at the border between FC and DFC (Figure 7C). In this case, the nascent RNA transcripts should radiate away around the spring into the DFC where they associate with processing RNPs (Figure 7B and C). In line with this model, the rDNA transcribed region as well as elongating RNA polymerase molecules were mapped at the periphery of the FC in the area around the FC/DFC border [11, 61, 65-67]. Active rRNA synthesis was mapped to the same location by BrUTP run-on labelling [10] whereas the nascent pre-rRNA transcripts were invariably mapped inside of the DFC [13-15, 61, 66, 66]. These data suggest that the nascent rRNA transcripts while synthesized at the periphery of the FC enter into DFC where they associate with processing RNPs [11-13].
Chapter 4

Our model is also in accordance with the dynamics of RNAP I transcription. A recent study showed that the SL1 is a long-live complex with maximal residence time at the rDNA [42]. This is in line with our data that suggest a stable binding of the SL1 to the promoter, enhancer and terminator as a part of central tetrad. Additionally, it was found that the RNAP I molecules do not participate in the reinitiation but are assembled de novo at every initiation round [42] in agreement with the low occupancy of RNAP I at terminator R1 (Figure 4B) suggesting that the RNA polymerase drops off at the end of the spring and is prevented from the reaching of the central tetrad. In this scenario, the RNAP I cannot enter a next round of transcription by jumping over to the enhancer/promoter but falls off the rDNA at terminator.

Thus, based on our data and spatial modeling we propose the unique three-dimensional spring-core structure of the actively transcribed rRNA gene where promoter, enhancer and terminator are bound together to SL1 complex and are surrounded by the cylindrically shaped transcribed region of the rDNA gene. This model correlates very well with nucleolar structure and can predict physical localization of DNA, RNA and transcriptional factors in the nucleolar compartments that corresponds to the available data from in situ EM studies. Apart from that, the model describes the mechanistic principle of RNAP I transcription and resolves the spatial basis of the highly efficient rRNA synthesis.

Materials and methods

Chromatin immunoprecipitation (ChIP)
U2OS cells were crosslinked with 1% formaldehyde for 30 min at room temperature, quenched with 0.125 M glycine and washed at 4°C with three buffers: PBS, buffer B (0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6) and buffer C (0.15 M NaCl in HEG buffer (1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.6)). Cells were then suspended in ChIP-incubation buffer (0.15% SDS, 1% Triton X-100, 150 mM NaCl, HEG) and sheared using a Branson-250 sonicator. Sonicated chromatin was centrifuged for 5 min and then the amount equivalent to 10^8 cells was incubated overnight with 2 μg of an antibody and 10 μl of protein A/G beads (Santa Cruz). Beads were washed 6 times with different buffers at 4°C: 2 times with [0.1% SDS, 0.1% DOC, 1% triton, 150 mM NaCl, HEG], 1 time with [same as before but with 500 mM NaCl], 1 time with [0.25 M LiCl, 0.5% DOC, 0.5% NP-40, HEG] and 2 times with HEG. Precipitated chromatin was eluted with 400 μl of elution buffer [1% SDS, 0.1 M NaHCO_3], incubated at 65°C for 4 hours in the presence of 200 mM NaCl, phenol extracted and precipitated with 20 μg of glycogen at -20°C overnight.

The procedure for ChIP-cloning is described earlier [50].

Quantitative PCR (qPCR)

ChIP experiments were analyzed by qPCR with specific primers using a SYBR green kit (Applied Biosystems) and results from two experiments were averaged. Efficiency of ChIP was calculated as a percentage of input, and specificity - as folds over average of negative controls (the transcriptionally silent genomic loci such as 2nd exon of myoglobin and the sites in the intergenic non-transcribed spacer). Primers for qPCR were designed with Primer Express and verified as amplifying a single PCR fragment. PCR efficiencies of the primers were calculated with series of 10-times dilutions and accepted when found to be reliable (2 +/- 0.15). Primer sequences are listed in Supplementary Table I.
Cell cycle synchronization

U2OS cells were grown at about 80% confluence and were arrested in metaphase by incubation with 40 ng/ml nocodazole in DMEM medium supplemented with 10% serum. After 18 hours the round shaped cells were collected by gentle flow of the medium and either directly crosslinked (the M-phase), or plated on a separate dish in the medium without nocodazole and crosslinked after 8 hours (the G1-phase). In addition, about 50% confluent U2OS cells were incubated two times with 2.5 mM thymidine in DMEM for 18 hours with 12 hours interval of incubation in normal DMEM. After treatment the cells were washed and incubated for 4 hours (for S phase) and 7½ hours (for G2 phase).

Acknowledgments

This work was supported by National Scientific Organization (NGI 050-71-016).

References

Chapter 4


33. Hanada K, Song CZ, Yamamoto K, Yano K, Maeda Y, Yamaguchi K, Muramatsu M. RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. EMBO J. 1996. 15:2217–2226


44. Fakan S, Hughes ME. Fine structural ribonucleoprotein components of the cell nucleus visualized after spreading and high resolution autoradiography. Chromosoma. 1989. 98:242–249


Architecture of the rDNA gene

Chapter 5

E2F transcriptional repressor complexes are critical downstream targets of p19$^{\text{ARF}}$/p53-induced proliferative arrest

Benjamin Rowland, Sergey Denissov, Sirith Douma, Hendrik Stunnenberg, René Bernards and Daniel Peeper

Cancer Cell (2002) 2, 55 - 65
Abstract
The p16<sup>INK4a</sup>/pRB/E2F and p19<sup>ARF</sup>/p53 tumor suppressor pathways are disrupted in most human cancers. Both p19<sup>ARF</sup> and p53 are required for the induction of senescence in primary mouse embryonic fibroblasts (MEFs), but little is known about their downstream targets. Disruption of E2F-mediated transcriptional repression in MEFs caused a general increase in the expression of E2F target genes, including p19<sup>ARF</sup>. We detected no contribution of E2F-mediated transactivation in this setting, indicating that a predominant role of endogenous E2F in asynchronously growing primary MEFs is to repress its target genes. Moreover, relief of transcriptional repression by E2F rendered MEFs resistant to senescence induced by either p19<sup>ARF</sup>, p53, or RAS<sup>V12</sup>. Thus, E2F transcriptional repressor complexes are critical downstream targets of antiproliferative p19<sup>ARF</sup>/p53 signaling.

Significance
E2F transcriptional repressor complexes are considered to be important downstream components of the p16<sup>INK4a</sup>/pRB tumor suppressor pathway. We find that E2F repressors are also critical targets for the ARF and p53 tumor suppressors during induction of replicative senescence and cell cycle arrest. Hence, our finding suggests that p16<sup>INK4a</sup> and ARF/p53 converge at the level of E2F repressor complexes. This is unexpected, because the pRB and p53 pathways were thought to communicate to different downstream targets. As E2F repressors are controlled by these two major tumor suppressor pathways, our model predicts that E2F-dependent transcriptional repression is deregulated not only by a mutant p16<sup>INK4a</sup>/pRB pathway, but also upon mutation of the ARF/p53 pathway, i.e., in the vast majority of human tumors.

Introduction
Upon explantation, cultured primary murine embryonic fibroblasts (MEFs) divide only a limited number of times before they undergo replicative senescence (Hayflick, 1965; reviewed in Sherr and DePinho, 2000). This cell cycle arrest is accompanied by increased levels of the p16<sup>INK4a</sup>, p19<sup>ARF</sup> (both encoded by the INK4a locus), p53, and p21<sup>CIP1</sup> proteins. Both p16<sup>INK4a</sup> and p19<sup>ARF</sup> are induced during in vitro propagation of primary MEFs. p16<sup>INK4a</sup> inhibits cellular proliferation in a manner that requires the function of either the retinoblastoma protein pRB or, as shown recently, both pRB-related “pocket” proteins, p107 and p130 (Bruce et al. 2000; Koh et al. 1995; Lukas et al. 1995 and Medema et al. 1995). p19<sup>ARF</sup>, too, is a tumor suppressor gene that has been proposed to protect cells against excessive mitogenic or oncogenic signaling (Sherr, 1998). p19<sup>ARF</sup> neutralizes the E3-ubiquitin ligase for p53, MDM-2, and thereby stabilizes p53 (Pomerantz et al. 1998 and Zhang et al. 1998). As a result of p19<sup>ARF</sup> expression, p53 transcriptionally induces its target genes, including MDM-2 and the cell cycle inhibitor p21<sup>CIP1</sup> (Vogelstein et al. 2000 and Vousden 2000). Although p53 is a regulatory target of p19<sup>ARF</sup>, the latter can interfere with cell cycle progression in both p53-dependent and p53-independent manners (Carnero et al. 2000; Martelli et al. 2001 and Weber et al. 2000).

When a constitutively active mutant allele of the Ha-RAS gene (encoding RAS<sup>V12</sup>) is introduced into primary MEFs, they also enter replicative senescence, but now prematurely (Serrano et al., 1997). Both spontaneous and RASV12-induced senescence depend on the presence of functional p19<sup>ARF</sup> and p53, as primary fibroblasts deficient for either of these genes can be cultured indefinitely, irrespective of RAS<sup>V12</sup> expression (Kamijo et al. 1997; Serrano et al. 1997 and Tanaka et al. 1994). However, how activation of the p19<sup>ARF</sup>/p53 pathway by (RASV12-induced) senescence signaling eventually results in an irreversible cell cycle arrest is not well understood. As MEFs deficient for p21<sup>CIP1</sup> are not immortal and still undergo RAS<sup>V12</sup>- or p19<sup>ARF</sup>-induced arrest, p21<sup>CIP1</sup> likely is not a critical...

When primary cells enter either RAS^{V12}-induced or spontaneous senescence, pRB accumulates in its active, hypophosphorylated form (Serrano et al. 1997 and Stein et al. 1990). We and others recently reported that both spontaneous and RAS^{V12}-induced senescence are dependent on the retinoblastoma gene family. Fibroblasts deficient for all three pocket proteins (pRB, p107, and p130) were shown to be immortal and failed to senesce upon expression of either RAS^{V12} or p19^{ARF} (Dannenberg et al. 2000; Peep er et al. 2001 and Sage et al. 2000). pRB interacts with a number of proteins involved in cell cycle regulation, including MDM-2 (Xiao et al., 1995), PML (Alcalay et al., 1998), the tyrosine kinase c-ABL (Welch and Wang, 1993), and E2F transcription factors (Helin et al. 1992 and Kaelin et al. 1992), but which of these act in the senescence response is unclear.

The E2F transcription factor family consists of six structurally related members, five of which (E2F-1 through E2F-5) contain a transactivation domain that is inhibited by binding to a pocket protein. E2F-1, -2, and -3 preferentially associate with pRB, E2F-4 with p107 or p130, and E2F-5 with p130 (Muller and Helin, 2000). The pocket proteins not only interfere with transactivation, but form, upon association with E2F and histone deacetylases (HDACs), active transcriptional repressor complexes (Harbour and Dean 2000 and Zhang et al. 2000). E2F-6 is unique in that it has a pocket-protein-independent repression motif (Cartwright et al. 1998; Gaubatz et al. 1998 and Trimarchi et al. 1998). The pRB family controls cell cycle progression by transiently associating with E2Fs. During G1, D type CYCLINS activate CDK4 and CDK6, which in turn phosphorylate pRB (Weinberg, 1995). pRB and HDAC subsequently dissociate, which results in derepression of E2F target genes (Zhang et al., 2000). Further phosphorylation of pRB by CYCLIN E-CDK2 causes pRB to dissociate from E2F, thereby actively inducing E2F-dependent transactivation and stimulating the cell to enter S phase (Harbour et al., 1999).

Pocket proteins are not the only communicative link between senescence signaling and the E2F transcription factors. A number of additional observations suggest a possible role for E2F in the regulation of senescence: the levels of various E2Fs decrease during the onset of senescence (Dimri et al. 1994 and Haddad et al. 1999), and overexpressed E2F-1 induces both ARF (Bates et al. 1998; DeGregori et al. 1997 and Dimri et al. 2000) and premature senescence (Dimri et al., 2000). To analyze in more detail its role in senescence, we disrupted the transcription-regulating function of E2F and analyzed the (p19^{ARF}/p53-dependent) senescence response in primary murine fibroblasts. As E2Fs have the ability to mediate active transcriptional repression, as well as activation, we also addressed in primary MEFs which of these functions is required for the transcriptional regulation of endogenous E2F target genes, including p19^{ARF}, and for an adequate response to antiproliferative p19^{ARF}/p53 signaling.

Results

Ectopic expression of an E2F-1 C-terminal deletion mutant displaces endogenous E2Fs from DNA

Overexpression of E2F-1 has been reported to induce ARF transcription (Bates et al. 1998; DeGregori et al. 1997 and Dimri et al. 2000). To address whether p19^{ARF} is also subjected to transcriptional repression by endogenous E2Fs, we interfered with E2F signaling by the use of a mutant of E2F-1 (E2F-DB; Figure 1A) in primary MEFs. E2F-DB lacks both the C-terminal transactivation and the pRB binding domains, but can still bind to DNA in heterodimeric complex with DP-1. E2F-DB has been used previously to rescue pocket protein-mediated transcriptional repression (Johnson 1995 and
Qin et al. 1995) and to show that cell cycle arrest induced by either p16\textsuperscript{INK4a}, TGFβ, or contact inhibition requires active E2F-mediated transcriptional repression (Zhang et al., 1999).

**Figure 1.** Ectopic expression of an E2F-1 C-terminal deletion mutant displaces endogenous E2Fs from DNA. 
A: Schematic representation of the E2F-1 mutants used in this study. B: E2F-DB displaces endogenous E2F from the DNA. Nuclear extracts were prepared from primary p53\textsuperscript{−/−} MEFs infected with retroviruses encoding E2F-DB or control viruses and analyzed by EMSA for E2F DNA binding activity in the absence or presence of antibodies for E2F-DB or E2F-4. C: E2F-DB binds to E2F-responsive promoters in vivo. ChIP assays performed in wild-type primary MEFs for E2F-DB on the p19\textit{ARF}, CYCLIN A, and CYCLIN E promoters. Data are represented as real-time PCR signals from p19\textit{ARF}, CYCLIN A, or CYCLIN E, relative to a γ-ACTIN control PCR performed on the same ChIP.

The mechanism of action of E2F-DB has previously been shown to involve binding to E2F sites and subsequent displacement of endogenous E2Fs, as demonstrated by electrophoretic mobility shift assays (EMSA) in both fibroblasts and epithelial cells (Krek et al. 1995 and Zhang et al. 1999). First, we wished to establish whether E2F-DB displaces endogenous E2F from E2F sites also in MEFs. Indeed, upon infection of MEFs with E2F-DB-encoding retroviruses, we observed that the DNA binding activity of endogenous E2F was almost completely abolished and replaced by an E2F-DB/DNA complex (Figure 1B). This suggests that E2F-DB occupies the E2F sites in E2F-responsive promoters in vivo. To address this, we performed chromatin immune precipitation (ChIP) on E2F-DB, followed by real-time PCR on three different E2F-responsive promoters. Consistent with the results obtained by EMSA, we observed that in E2F-DB-expressing MEFs, significant amounts of E2F-DB occupy the p19\textit{ARF}, CYCLIN A, and CYCLIN E promoters (Figure 1C). Together, these data provide strong evidence that E2F-DB acts to displace endogenous E2Fs from E2F-responsive promoters.
Endogenous E2Fs act as repressors in cycling primary murine fibroblasts

The E2F transcription factors have the ability to mediate either transcriptional repression or activation. In order to determine the relative contribution of these two functions on the regulation of endogenous target genes of E2F, we compared wild-type E2F-1 to mutants of E2F-1, which lack either the repression function only (E2F-1 (Y411C)) or both the repression and transactivation functions (E2F-DB). Upon retroviral expression in primary MEFs, wild-type E2F-1 differentially induced its transcriptional targets p19ARF, PCNA, p107, MCM3, and CYCLINS E1 and A (Figure 2A). E2F-DB, on the other hand, markedly induced all of these targets (Figure 2A). This was not a global effect of E2F-DB, as the expression of p16INK4a, CYCLIN D1, and CDK4 remained unaffected (Figure 2E). As expected, a DNA binding-deficient mutant of E2F-DB (E2F-DB (E132)) (Hsieh et al., 1997) failed to induce the expression of the E2F target genes (Figure 2B).

If endogenous E2Fs act predominantly to activate transcriptional targets, one would expect that E2F-DB would interfere with this activation, something we clearly did not observe. We therefore conclude that in cycling primary MEFs, endogenous E2F controls target gene expression predominantly by means of active repression. Restoration of specifically E2F's transactivation function (i.e., expression of E2F-1 (Y411C)), which transactivates but does not repress (Helin et al., 1993), failed to give rise to additional induction on top of that achieved by E2F-DB, indicating that the transactivating function of E2F-1 is dispensable for the induction of at least this set of E2F target genes. As E2F-1 controls its own transcription (Hsiao et al. 1994; Johnson et al. 1994 and Neuman et al. 1994), we wished to exclude that E2F-DB induces the levels of transcriptionally competent endogenous E2F-1, which in turn switches on expression of E2F target genes. Indeed, the levels of endogenous E2F-1 remained unaltered in the presence of E2F-DB (Figure 2E), excluding this possibility.

E2F-mediated transcriptional repression occurs through association with pocket proteins. Hence, it seemed likely that the E2F-DB-mediated derepression is pocket protein dependent. To address this directly, we infected MEFs deficient for all three pocket proteins (TKO MEFs) (Dannenberg et al., 2000) with E2F-DB-encoding retrovirus. Whereas E2F-DB derepressed p19ARF, PCNA, CDK1, MCM3, and CYCLINS E1 and A in wild-type MEFs, this regulation was completely absent in TKO MEFs (although E2F-DB was expressed slightly less in TKO MEFs than in wild-type MEFs; Figure 2C). These results demonstrate that in MEFs, the observed E2F-dependent repression of target genes is mediated mainly, if not exclusively, by pocket proteins. Importantly, in TKO MEFs, we also failed to observe an E2F-DB-mediated decrease in the levels of E2F targets, which, as E2Fs in these cells are thought to be free and transactivation competent, supports our notion that in proliferating MEFs, E2Fs serve mainly as repressors of transcription.

Recently, p19ARF was shown to bind to E2F-1 and target it for degradation (Martelli et al., 2001). Possibly, E2F-DB sequesters p19ARF and thereby protects endogenous E2F from degradation, leading to the induction of E2F target genes. However, E2F-DB-dependent derepression of all E2F targets tested occurred similarly (albeit to different extents) in wild-type and p19ARF−/− MEFs, excluding an important role for p19ARF in this respect (Figure 2D).

Expression of p19ARF, an E2F target gene, is induced by RASV12 and required for RASV12-induced senescence (Palmero et al., 1998), raising the possibility that p19ARF induction by RASV12 requires E2F activity. However, E2F-DB did not inhibit p19ARF induction by RASV12, but rather enforced it (Figure 2E). The RASV12-dependent induction of p19ARF in the presence of E2F-DB was functional, as it led to an increase in p53 levels (Figure 2E). This result strongly suggests that RASV12-dependent induction of p19ARF and p53 does not require E2F-mediated transactivation.
Transcription repression by E2F complexes

Figure 2. Endogenous E2Fs act as repressors in primary murine fibroblasts.

A: Disruption of E2F function results in derepression of various E2F targets. Young primary wild-type MEFs were infected with retroviruses encoding E2F-DB, E2F-1, or E2F-1 (Y411C). At 5 days after infection, cell extracts were prepared and subjected to Western blotting with the indicated antibodies.

B: E2F-DB-mediated derepression requires its DNA binding function. MEFs were infected with E2F-DB, E2F-DB (E132), or control viruses and processed as in A.

C: E2F-dependent repression is mediated by pocket proteins. Wild-type and Rb−/−/p107−/−/p130−/− (TKO) MEFs were infected with E2F-DB or control viruses and processed as in A.

D: E2F-dependent repression is independent of p19ARF status. Wild-type and p19ARF−/− MEFs were infected and processed as in A.

E: RASV12 induces p19ARF independent of E2F transactivation. Wild-type MEFs were infected with retroviruses encoding E2F-DB, RASV12, or both. Cell extracts were prepared at 6 days after infection. Similar observations were made for extracts prepared at 3, 12, and 24 days postinfection (not shown). In all panels, CDK4 served as a loading control.

Disruption of E2F-dependent repression results in immortalization

Increased p19ARF levels cause cell cycle arrest or senescence (Kamijo et al. 1997 and Quelle et al. 1995). As E2F-DB derepressed p19ARF expression in MEFs, we expected that E2F-DB would cause premature senescence. Surprisingly, whereas control-infected MEFs lost their replicative potential, at least initially (see below), cells expressing E2F-DB completely failed to undergo senescence (Figure 3A). This effect of E2F-DB was independent of its CYCLIN A binding domain (Figure 3B), which is required to appropriately inactivate the DNA binding function of E2F in S phase (Krek et al., 1994). Moreover, this bypass of senescence was not observed for wild-type E2F-1, which caused both cell death and premature senescence (consistent with previous reports; DeGregori et al. 1997; Dimri et al. 2000; Qin et al. 1994; Shan et al. 1994 and Wu and Levine 1994; Figures 3A and 3D), nor was it
observed for E2F-1 (Δ24) lacking the CYCLIN A binding domain (Figure 3B). The pRB binding-deficient E2F-1 (Y411C) mutant caused cell death (Figure 3B), although especially at higher cell density, a substantial amount of the cells did continue to proliferate (Figure 3D). As expected, the DNA binding-deficient point mutant of E2F-DB failed to yield any proliferative advantage (Figure 3C).

**Figure 3.** Disruption of E2F-dependent repression results in immortalization. 
**A:** E2F-DB immortalizes primary MEFs. Wild-type, primary MEFs at passage 3 were infected with pBABE-puro retroviruses encoding wild-type or mutant E2F-1, as indicated. At 2 days postinfection, infected cells were selected for expression of the puromycin-selectable marker for 5 days and used in a proliferation curve performed in duplicate. 
**B:** Immortalization by E2F-DB does not require a CYCLIN A binding domain. MEFs were infected with retroviruses encoding E2F-1 (Δ24), E2F-DB (Δ24), or E2F-1 (Y411C). Analysis was performed as in A. 
**C:** Immortalization by E2F-DB requires its DNA binding function. MEFs were infected with retroviruses encoding E2F-DB, E2F-DB (E132), or control viruses. Analysis was performed as in A. 
**D:** E2F-1 (Y411C) can stimulate MEF proliferation. Low-passage MEFs were infected with either E2F-DB, E2F-1 (wild-type), or E2F-1 (Y411C) and cultured under subconfluent conditions. Photographs were taken at passage 7. 
**E:** Immortalization by E2F-DB occurs in the presence of p19ARF and p16INK4a. Western blot analysis for the indicated proteins of immortal E2F-DB-expressing MEF clones at passage 20. 3T3 indicates a spontaneously immortalized MEF clone (Vector I from the growth curve in A) that apparently had lost p19ARF expression. E2F-DB lines 3 and 4 represent cell extracts taken from E2F-DB lines I and II, respectively, also taken from the growth curve in A. E2F-DB lines 1 and 2 are independently propagated MEFs. MEF indicates passage 8 primary MEFs. 
**F:** Immortalization by E2F-DB occurs in the presence of functional p53. Western blot analysis of the indicated proteins of immortal E2F-DB-expressing MEF clones at passage 20, after 16 hr treatment with cis-platin (50 µM). CDK4 served as a loading control.

We next determined whether the E2F-DB-mediated immortalization requires mutations in the genes that are frequently mutated during spontaneous immortalization, namely p19ARF and p53. We
Transcription repression by E2F complexes

propagated four independent E2F-DB-expressing MEF populations for 20 passages and then analyzed p16\(^{INK4a}\), p19\(^{ARF}\), p53, and p21\(^{CIP1}\) levels, as well as p53 function. Compared to control MEFs at passage 8, E2F-DB MEFs at passage 20 continued to abundantly express both p19\(^{ARF}\) and p16\(^{INK4a}\) (Figure 3E). By contrast, spontaneously immortalized (“3T3") mouse fibroblasts had lost expression of p19\(^{ARF}\). We then used the DNA damage-induced stabilization of p53 and concomitant transcriptional activation of p21\(^{CIP1}\) to determine whether p53 was still functional in the E2F-DB-immortalized MEFs. cis-platin treatment led to stabilization of p53 in all four late-passage E2F-DB-expressing lines (Figure 3F). Moreover, p53 was functional, as DNA damage increased the levels of its transcriptional target p21\(^{CIP1}\) (Figure 3F).

Disruption of E2F-dependent repression results in bypass of RAS\(^{V12}\)-induced senescence

RAS\(^{V12}\)-induced premature senescence much resembles spontaneous senescence of primary cells. Therefore, we tested whether disruption of E2F transcriptional repression not only leads to immortalization of primary MEFs, but also rescues RAS\(^{V12}\)-induced premature senescence. MEFs expressing only RAS\(^{V12}\) prematurely senesced, although after prolonged culturing one of these populations (“RAS\(^{V12}\) II”) regained proliferative potential (Figure 4A; see below). By contrast, MEFs expressing both RAS\(^{V12}\) and E2F-DB, in spite of overexpressing RAS\(^{V12}\) (not shown), efficiently bypassed RAS\(^{V12}\)-induced senescence (Figure 4A).

Like E2F-DB-immortalized MEFs, both RAS\(^{V12}/\)E2F-DB MEF lines had retained normal expression of both p19\(^{ARF}\) and p53, as well as of p16\(^{INK4a}\) (Figure 4B). This was in contrast to a spontaneously immortalized RAS\(^{V12}\)-expressing cell line (RAS\(^{V12}\) II) that had lost expression of both p16\(^{INK4a}\) and p19\(^{ARF}\) (Figure 4B). Thus, disruption of E2F function results in derepression of p19\(^{ARF}\), but in spite of this, it also concomitantly causes a bypass of both spontaneous and RAS\(^{V12}\)-induced senescence, thereby allowing cells to proliferate indefinitely while ignoring the sharp elevation in endogenous p19\(^{ARF}\) levels.

Figure 4. Disruption of E2F-dependent repression results in bypass of RAS\(^{V12}\)-induced senescence.

A: E2F-DB expression allows primary MEFs to bypass RAS\(^{V12}\)-induced senescence. Wild-type, primary MEFs at passage 3 were (co-) infected with retroviruses encoding RAS\(^{V12}\) and either control or E2F-DB-encoding retroviruses, selected for puromycin and hygromycin, and processed as described for Figure 3A. B: Bypass of RAS\(^{V12}\)-induced senescence by E2F-DB occurs in the presence of p19\(^{ARF}\), p16\(^{INK4a}\), and p53. Western blot analysis of passage 15 clones, after completion of the growth curves described in A. p53/\(^{3T3}\) (lane 4) indicates a spontaneously immortalized MEF clone expressing high levels of mutant p53, which failed to transcriptionally activate p21\(^{CIP1}\). CDK4 served as a loading control.
Disruption of E2F-dependent repression rescues cell cycle arrest induced by ectopically expressed p19ARF and p53

The results shown above strongly argue that endogenous E2F repressor function is required for cells to respond to senescence-associated induction of p19ARF. To test whether E2F-DB can bypass also an ectopic p19ARF-induced cell cycle arrest, we generated NIH 3T3 cell lines stably expressing E2F-DB or empty vector. These cell lines retained functional p53, as judged by its DNA damage-induced stabilization and concomitant induction of p21^{CIP1} (data not shown). We infected both cell populations with a p19ARF-encoding retrovirus. As expected, p19ARF induced a robust G1 arrest in the control cells (Figure 5A). By contrast, this arrest was reduced by half in E2F-DB-expressing cells. This observation was supported by a colony formation assay: upon infection with a retrovirus producing a p19ARF-RFP chimeric protein, the control population showed only few proliferating, p19ARF-positive cells (Figure 5B). In fact, the small number of cells that did express p19ARF (localized in the nucleoli, in agreement with previous observations; Weber et al. 1999 and Zhang and Xiong 1999) had a large and flattened morphology, typical of senescent cells. Apparently, strong selection occurred against maintaining p19ARF expression. By contrast, the vast majority of the E2F-DB-expressing cells produced clearly detectable levels of p19ARF without displaying a senescent morphology (Figure 5B), reinforcing our notion that ectopically expressed p19ARF blocks cellular proliferation in an E2F-dependent manner.

Figure 5. Disruption of E2F-dependent repression rescues cell cycle arrest induced by ectopically expressed p19ARF and p53.  
A: E2F-DB interferes with cell cycle arrest induced by overexpression of p19ARF. Parental or E2F-DB-expressing NIH-3T3 cells were infected with either control or p19ARF-ires-GFP-encoding retroviruses (either undiluted or diluted into medium, as indicated) and treated with nocodazole (50 ng/ml) to specifically analyze the cell fraction in G1. At least 90% of the cells were infected, as judged by the number of GFP-positive cells. At 48 hr after infection, the cell cycle profile was determined by FACS analysis. B: E2F-DB interferes with inhibition of colony formation by overexpression of p19ARF. Parental or E2F-DB-expressing NIH-3T3 cells were infected with either control or LZRS-p19ARF-RFP-ires-zeo retrovirus. At 6 days after infection, fluorescence microscopy photographs were taken. Representative examples are shown. C: E2F-DB interferes with cell cycle arrest induced by overexpression of p53. Parental or E2F-DB-expressing p53^{-/-} cells were infected with either control or LZRS-p53-RFP-ires-zeo retrovirus (either undiluted or diluted into medium, as indicated). BrdU incorporation was
measured 48 hr after infection. D: E2F-DB interferes with inhibition of colony formation by overexpression of p53. Parental or E2F-DB-expressing p53−/− cells were infected as in C. At 6 days after infection, fluorescence microscopy photographs were taken. Representative examples are shown.

E2F-DB rescued p19ARF-induced cell cycle arrest, a p53-dependent event (at least in part), raising the possibility that E2F-DB acts downstream of p53. To test this, we established p53−/− MEFs stably expressing E2F-DB, which we subsequently infected with a p53-encoding retrovirus. As expected, control cells ceased to undergo DNA replication almost completely upon expression of p53 (Figure 5C). By contrast, in the presence of E2F-DB, p53 decreased the number of cells undergoing DNA replication only by roughly half. In support of this finding, we observed that in a colony formation assay, E2F-DB-expressing MEFs continued to proliferate despite clearly detectable levels of retrovirally expressed p53-RFP fusion protein, which was localized in the nucleus (Figure 5D). In control cells, neither abundant expression nor nuclear localization of p53-RFP was compatible with proliferation, and cells that did express p53-RFP again displayed a “flat-cell” phenotype (Figure 5D). Thus, E2F-DB not only rescues spontaneous senescence (a p19ARF- and p53-dependent process), but also significantly interferes with a cell cycle arrest imposed by ectopic overexpression of p19ARF and p53. These results further support our notion that E2F acts downstream of both p19ARF- and p53-dependent antiproliferative signaling.

**E2F activity is required for cell cycle reentry from quiescence**

The data above show that E2F’s transactivation function is dispensable for the induction of (at least a number of) E2F target genes. Moreover, they strongly suggest that the absence of E2F-mediated transactivation is compatible with proliferation of primary murine cells, something that has been reported previously for other cell types (Sellers et al. 1995; Zhang et al. 1999 and He et al. 2000). However, most E2Fs possess a transactivation domain, arguing that specific situations should exist where transactivation is required. Indeed, ectopic E2F-1-mediated transactivation, but not repression, is sufficient to induce S phase entry in the absence of mitogens (Johnson et al. 1993; Kowalik et al. 1995; Qin et al. 1994 and Shan et al. 1996). Moreover, E2F1−/− fibroblasts display a delayed G0-S transition in response to mitogen stimulation (Wang et al., 1998). We therefore investigated whether endogenous E2F transactivation function is not only sufficient, but also required for mitogen-induced cell cycle reentry.

![Figure 6](image.png)

**Figure 6.** E2F activity is required for cell cycle reentry from quiescenceNIH 3T3 cell lines stably expressing either E2F-DB or empty vector were propagated in 10% serum and subsequently deprived of serum for 72 hr. Then, the cells were refed with 10% serum for 24 hr. Represented is the proportion of cells incorporating BrdU, as a measure of DNA synthesis, in each situation. Average and standard deviations are based on three independent experiments.
We deprived E2F-DB expressing cells of serum, which caused them to exit the cell cycle as efficiently as control cells (Figure 6). Apparently, in immortal murine fibroblasts, no E2F function is required for G0 entry, consistent with previous observations in other cell types (Zhang et al., 1999). By contrast, upon serum refeeding, the control cells readily reentered the cell cycle, whereas the E2F-DB-expressing cells were significantly impaired in their DNA replication (Figure 6). Taken together, this suggests that loss of E2F transcriptional activity prevents mitogen-induced cell cycle reentry, indicating that in this specific situation, E2F-mediated transactivation is required.

**Discussion**

**E2F-mediated transcriptional repression and p19ARF-dependent regulation of senescence**

We demonstrate that in primary murine fibroblasts, various E2F targets, including p19ARF, are repressed rather than transactivated by endogenous E2F. More importantly, disruption of E2F-mediated repression allowed MEFs to bypass both spontaneous and RASV12-induced senescence and to proliferate indefinitely, in the face of high levels of p19ARF. We therefore conclude that p19ARF promotes proliferative arrest in an E2F transcriptional repression-dependent manner, a notion that is reinforced by our observation that disruption of E2F repression interfered also with the induction of cell cycle arrest by ectopically expressed p19ARF or p53.

Expression of E2F-DB led to induction (i.e., derepression), as opposed to reduction (i.e., loss of transactivation), of various E2F targets, including p19ARF. Moreover, similar to E2F-DB-expressing MEFs, pRB family-deficient MEFs, which lack E2F-repressor complexes, are immortal despite high levels of p19ARF. These observations indicate that in primary MEFs, it is specifically E2F's transcriptional repression function, but not its transactivation function, which is required for the senescence checkpoint. What, then, is the function of E2F's transactivation domain? We observed neither an effect of E2F-DB on the levels of p19ARF (or other E2F targets) in pRB family-deficient cells, nor a contribution of E2F's transactivation domain on top of the induction of E2F target genes by E2F-DB in wild-type MEFs. Although we cannot exclude that in proliferating cells some residual E2F transactivator function remains despite the presence of E2F-DB, our results argue that in cycling primary MEFs, E2F-dependent transactivation is dispensable. By contrast, E2F-DB did efficiently prevent mitogen-stimulated cell cycle reentry. Our data therefore suggest that E2F-dependent transactivation is required only in specific circumstances, like in the presence of apoptosis-inducing signals or during cell cycle reentry (as shown schematically in Figure 7). Furthermore, we show that it is specifically E2F-dependent repression that is required for cell cycle exit in response to activation of the p19ARF/p53 pathway.

Although various E2F target genes were derepressed both in E2F-DB-expressing wild-type MEFs and in pRB family-deficient MEFs, the extent of derepression was greater (with the exception of CYCLIN E) in the latter cell type. This can be explained by the fact that E2F-DB acts in a dominant-negative manner and displaces most, but not all, endogenous E2F complexes from the DNA (see Figure 1B). By contrast, pRB family-deficient MEFs completely lack E2F/pocket protein complexes and as a result show maximal derepression of E2F targets. Importantly, ectopic expression of E2F-DB and genetic ablation of pocket proteins do result in a similar biological phenotype: both interfere with the senescence checkpoint and lead to immortalization.

Recent data show that the simultaneous absence of E2Fs 1, 2, and 3 causes defects in G1/S target gene activation and inhibition of proliferation (Wu et al., 2001). This may seem at odds with our finding that MEFs proliferate well in the presence of E2F-DB. However, whereas in E2F-1-3 triple knockout cells E2F-4 and E2F-5 are free to repress transcription and thereby may inhibit
Transcription repression by E2F complexes

proliferation, this is compromised by E2F-DB, as this displaces endogenous E2Fs from E2F sites (Krek et al. 1995 and Zhang et al. 1999).

Figure 7. Hypothetical model for E2F function in cell cycle regulation. See text for details.

Previous studies (Bates et al. 1998 and Dimri et al. 2000) showed that wild-type E2F-1 induced the human p14ARF gene, whereas E2F-DB did not. This seemingly is in contrast with our present observations with E2F-DB. However, Bates et al. made their observations in SAOS-2 cells, which lack functional pRB and therefore contain only a limited amount of E2F repressor complexes. On the other hand, Dimri et al. used primary human WI-38 fibroblasts, which suggests that E2F-dependent ARF regulation varies depending on species or cell type, possibly as a result of differences in the composition and/or quantity of E2F complexes. For some targets we did observe induction by wild-type E2F-1, consistent with previous reports (Muller and Helin, 2000). However, we propose that some effects of overexpressed wild-type E2F may be caused by titrating cellular factors like pocket proteins away from E2F sites in promoters. In this case, the net effect would be similar to what we describe here for the transcription-deficient E2F-DB mutant, namely an increase in the levels of E2F targets, like ARF.

Primary MEFs expressing E2F-DB produce elevated levels of CYCLIN E1, which can contribute to immortalization (Peeper et al., 2002). In fact, overexpression of CYCLIN E1 has been observed in, and most likely contributes to the emergence of, many human tumors (Keyomarsi and Herliczek, 1997). However, it clearly is not the sole target in this respect, as also Rb-deficient MEFs produce high levels of CYCLIN E1 (Herrera et al., 1996), yet undergo senescence normally. Indeed, numerous genes are regulated by the pRB/E2F pathway, strongly suggesting that immortalization requires derepression of a number of E2F targets. An important question that remains to be addressed is how p19^{ARF}/p53 signaling leads to a requirement for E2F repressor complexes in a pathway leading to proliferative arrest. In this respect, we note that our data do not discriminate between a linear pathway between p19^{ARF}/p53 and E2F and a more indirect mechanism of communication with intermediary signals.

Communication between p19^{ARF}, p16^{INK4a}, p53, and E2F
It has been suggested previously that p19^{ARF} requires functional p16^{INK4a} to induce cell cycle arrest (Carnero et al., 2000). As E2F-DB rescues not only p19^{ARF}-induced arrest, as we show here, but also a p16^{INK4a}-induced arrest (Zhang et al., 1999), and since E2F-DB-expressing MEFs proliferate in the presence of normal p16^{INK4a} levels, it is formally possible that E2F-DB immortalizes by acting
downstream of p16\(^{\text{INK4a}}\). Recent data, however, reveal that p16\(^{\text{INK4a}}\) is dispensable both for spontaneous and RAS\(^{V12}\)-induced senescence in MEFs (Krimpenfort et al. 2001 and Sharpless et al. 2001). This makes it highly unlikely that disruption of E2F-dependent repression renders MEFs immortal through interference with p16\(^{\text{INK4a}}\) signaling and argues in favor of our model that in senescence signaling, the p19\(^{\text{ARF}}\)/p53 pathway acts in an E2F-dependent manner.

In conclusion, we show that in primary MEFs, E2F-mediated repression is linked to antiproliferative signaling by p19\(^{\text{ARF}}\) in at least two ways (Figure 7). Upstream, E2F represses p19\(^{\text{ARF}}\), in a pocket protein-dependent manner. Downstream, functional E2F is required for both p19\(^{\text{ARF}}\) and p53 to induce cell cycle arrest and for the appropriate execution of the senescence program. Interestingly, the INK4a locus encodes two unrelated proteins, p16\(^{\text{INK4a}}\) and p19\(^{\text{ARF}}\), proposed to act in independent pathways, namely pRB- and p53-dependent, respectively. The Dean and Livingston laboratories have previously demonstrated that E2F repressor complexes are required for an appropriate cellular response to p16\(^{\text{INK4a}}\) (Gaubatz et al. 2000 and Zhang et al. 1999). Quite unexpectedly, however, we find that E2F repressor complexes are also essential for cell cycle exit in response to the other INK4a product, p19\(^{\text{ARF}}\). Thus, together these findings suggest that p16\(^{\text{INK4a}}\) and p19\(^{\text{ARF}}\)/p53 do not operate independently, but converge on E2F repressor complexes. This model would predict that tumors with mutations in the INK4a locus (as well as in p53) share deregulation of E2F-controlled target genes, irrespective of which INK4a gene is affected. As the E2F repressor complex is controlled by these two major tumor suppressor pathways, our model predicts that E2F-dependent transcriptional repression is deregulated in the vast majority of human tumors.

**Experimental procedures**

**Cell culture**

MEFs of OLA and FVB origin were isolated as described (Peep et al., 2001) and cultured in DMEM medium supplemented with 10% FBS, L-Glutamine, and penicillin/streptomycin (all GIBCO) and 0.1 mM β-mercaptoethanol. NIH-3T3 cells were cultured as MEFs, but with 10% NCS (GIBCO) and without β-mercaptoethanol. The Phoenix packaging cell line was used for the generation of ecotropic retroviruses (Serrano et al., 1997). MEFs were infected for 8 hr with filtered (0.45 µm) virus supernatant supplemented with 8 µg/ml polybrene.

Proliferation curves were performed by infection of low-passage MEFs with pBABE-puro retroviral vectors containing HA-E2F-1 (wild-type or mutant, as described in Figure 1) and pBABE-hygro retroviral vectors containing RAS\(^{V12}\). Cells were subsequently selected for puromycin and hygromycin resistance until all uninfected cells had died. Then, 1.5 \times 10^5 cells were seeded in a 6 cm dish and counted and split every 3.5 days. For cis-platin treatment, E2F-DB immortalized MEFs were cultured for 20 passages and incubated in 50 µM cis-platin for 16 hr.

**Plasmids**

pBABE-puro vectors containing HA-tagged E2F-1 wild-type, E2F-1 (1–368), E2F-1 (Δ24), and E2F-1 (1–368 Δ24) were kind gifts from W. Krek (Krek et al., 1995). E2F-1 mutants (1–374) (Helin and Harlow, 1994), E2F-1 (1–374, E132) (Hsieh et al., 1997), and E2F-1 (Y411C) (Helin et al., 1993) were subcloned into pBABE-puro. LZRS-p19\(^{\text{ARF}}\)-RFP-ires-zeo and LZRS-p53-RFP-ires-zeo retroviral vectors were kindly provided by T. Brummelkamp. MSCV-p19\(^{\text{ARF}}\)-ires-GFP and MSCV were kind gifts from C. Sherr.
Western blot analysis and EMSA
Preparation of protein extracts and Western blotting were performed as described (Peeper et al., 2001). Primary antibodies were M-156 for p16\(^{INK4a}\), C19 for p21\(^{CIP1}\), C22 for CDK4, PC-10 for PCNA, H-295 for Cyclin D1, M-20 for Cyclin E, C-19 for Cyclin A, C-19 for CDK1, G-19 for MCM3, C-18 for p107, KH95 for E2F-1 and E2F-DB (all Santa Cruz), KH-20 for E2F-DB (used in Figure 2E) (Neomarkers), R02120 for RAS (Transduction Laboratories), R562 for p19\(^{ARF}\) (Abcam), and Ab-7 for p53 (Calbiochem). Mobility shift assays were performed as described (Beijersbergen et al., 1995). Antibodies used for supershifts were KH-95 for E2F-DB and C-20 for E2F-4 (both Santa Cruz).

ChIP and real-time PCR
Primary MEFs were infected at passage 2 with pBABE-puro-HA-E2F-DB or parental pBABE-puro retroviruses and selected for puromycin resistance. At passage 6, chromatin was isolated from approximately 5 \(\times\) 10\(^7\) E2F-DB or control MEFs in essence as described (Botquin et al. 1998 and Orlando et al. 1997), but excluding CsCl-gradient purification. Chromatin was sonicated to an average size of 1000 bp and subsequently precleared with protein A/G sepharose beads (Santa Cruz) in incubation buffer (0.1% SDS, 1% Triton X-100, 0.1 M NaCl, 1 mM EDTA, 1 mM Tris [pH 8.0], 0.5 mM EGTA, 1 mg/ml BSA, and protease inhibitors). Antibody incubation (using 2 µg of KH95 for E2F-DB [Santa Cruz]) of precleared chromatin was performed overnight in incubation buffer at 4°C, and immunocomplexes were recovered with protein A/G-Sepharose beads. Immunoprecipitates were washed sequentially with 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 0.15 M NaCl, 1 mM EDTA, 10 mM Tris (pH 8.0), 0.5 mM EGTA; 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 0.5 M NaCl, 1 mM EDTA, 10 mM Tris (pH 8.0), 0.5 mM EGTA; 0.25 M LiCl, 0.5% deoxycholate, 0.5% NP-40, 1 mM EDTA, 10 mM Tris (pH 8.0), 0.5 mM EGTA, and 1 mM EDTA, 10 mM Tris (pH 8.0), 0.5 mM EGTA. Immunocomplexes were eluted twice from the beads in 1% SDS, 0.1 M NaHCO\(_3\) at room temperature for 15 min. Protein-DNA crosslinks were reversed in 0.2 M NaCl at 65°C for 4 hr, after which DNA was isolated by phenol-chloroform extraction. Real-time PCR was performed using the GeneAmp 5700 Sequence Detection System (PE Biosystems) using the SYBR Green I kit (PE Biosystems). Primers used for real-time PCR were for p19\(^{ARF}\) (E2F binding sites region), (forward) TTTTATAGATGGACTCGGAGCAA and (reverse) GTCCCGAAACTTTCGTCTATG; for CYCLIN A (E2F binding sites region), (forward) CCGGGCCTTCTCGTGAC and (reverse) CAAGTAGCCCGGACTATTGA; for CYCLIN E (E2F binding sites region), (forward) GGGCGTGTTTCTTTACGGG and (reverse) GCCCTGACATCTAGCCCCA; and for γ-ACTIN (exon 5), (forward) TCCGCAAAGACCTGTATG and (reverse) CTCTCTTGTGACATCGTGAC.

Cell cycle analysis and fluorescence microscopy
NIH-3T3 cells were infected with pBABE-puro-HA-E2F-DB or parental pBABE-puro retroviruses, and polyclonal pools were selected for puromycin resistance. For FACS analysis, cells were infected with MSCV-p19\(^{ARF}\)-ires-GFP or MSCV control virus. After 32 hr, 50 ng/ml nocodazole was added to the medium for 16 hr. Then, cells were permeabilized and stained with propidium iodide, and cell cycle profiles were determined by FACS analysis and analyzed using CellQuest software. For fluorescence microscopy, cells were infected with LZRS-p19\(^{ARF}\}-RFP-ires-zeo or parental pLZRS-ires-zeo retroviruses. Cells were plated at 10\(^5\) cells/10 cm dish. Photographs were taken after 6 days of zeocin selection using fluorescence microscopy, at 60× magnification.
p53\(^{-/-}\) MEFs were infected with pBABE-puro-HA-E2F-DB or parental pBABE-puro retroviruses, and polyclonal pools were selected for puromycin resistance. Subsequently, cells were infected with pLZRS-p53-RFP-ires-zeo or parental pLZRS-ires-zeo retroviruses. At 48 hr postinfection, 7.5 µg/ml bromo-deoxyuridine (BrdU) was added to the medium for 1 hr. Then, cells were processed for staining with anti-BrdU antibodies. FACS analysis and fluorescence microscopy were performed as described above. For serum starvation experiments, NIH-3T3 cells expressing E2F-DB or control cells were propagated in medium containing 10% NCS. Hereafter, cells were starved in 0.1% serum for 72 hr and then refed with 10% serum for 24 hr in the presence of 7.5 µg/ml BrdU.

Acknowledgements
We thank C. Sherr for generously providing a p19\(^{ARF}\) retroviral vector, K. Helin, W. Krek, and X. Lu for (mutant) E2F-1 vectors, T. Brummelkamp for p19\(^{ARF}\) and p53-RFP retroviral vectors, J. Jonkers for p53\(^{-/-}\) MEFs, and J.-H. Dannenberg and H. te Riele for wild-type and triple knockout MEFs. We thank our colleagues for helpful discussions and A. Berns for critically reading the manuscript. This work was supported by grants from the Netherlands Organization for Scientific Research and the Dutch Cancer Society.

References
Transcription repression by E2F complexes


Chapter 6

General discussion
General discussion

Coordinated regulation of gene expression is a fundamental basis for the realization of all cellular processes. Extensive investigations of the molecular mechanisms involved in gene regulation revealed a broad and highly complex network of various factors and processes. Among the most intensively investigated aspects of gene regulation are transcription factors, regulatory DNA elements and modifications of chromatin. Recently a research line focused on novel non-coding RNA (ncRNA) molecules revealed their broad regulatory functions.

In this thesis we describe a detailed investigation of the molecular basis for gene regulatory mechanisms. In Chapter 2 we aimed to identify functional DNA regions, such as promoters and enhancers, based on the binding of the universal transcription factor TBP (TATA-binding protein) and using ChIP-cloning and ChIP-on-chip methods. Furthermore, we obtained the binding profiles for general transcription factors at human promoters and performed a comprehensive data analysis of these to obtain novel insight in gene regulation. In Chapter 3 we investigated the molecular function of the general transcription factor TFIIA based on its binding profile at a large number of genomic loci. Finally, in Chapter 4 we presented an extensive analysis of the binding profiles of different RNAP I transcription factors and proposed a unique three-dimensional structure of an active rDNA gene. A detailed discussion on the results of this thesis in connection with current knowledge is presented below.

Genomic regulatory elements and molecular mechanisms of gene regulation

The integral parts of a gene, such as promoter, transcribed region with exons and introns, and termination sites, as well as distant gene regulatory elements, such as enhancers and LCRs, are determined by their specific DNA sequence and higher order chromatin structure [1, 2]. The identification of specific genomic regions as genes or gene regulators is one of the major challenges which is required for a reliable gene annotation and for the understanding of regulatory mechanisms.

In Chapter 2 we used ChIP-cloning to identify TBP binding sites in vivo. A large number of the sites were located in introns of protein-coding genes or at a large distance from annotated and predicted promoters. Using ChIP-on-chip, strand-specific RT-qPCR and gene-reporter assays, we showed that the majority of these TBP binding sites were functional promoters of hitherto unidentified genes, and possibly also enhancers and regions with yet unknown functions. Based on these results, we suggested that in addition to protein-coding genes there is at least an equal number of non-coding genes which had not been recognized by current gene-prediction algorithms. This observation is in line with a number of recent studies which reached similar conclusions [3 - 9]. Furthermore, the data in these studies suggest that nearly the entire human genome is transcribed. In the recent work from the ENCODE project, the annotation of genes and gene-regulatory elements in different tissues was addressed with unprecedented detail [8, 9]. In addition to transcript mapping and identification of a large number of novel non-coding genes, a large number of binding sites for different transcription factors were identified. Collectively, the results from our work and other recent studies provide novel insight into genome organization and gene regulation in the context of the complex network of cellular processes.

Although the functions of the novel non-coding RNA species are mostly unknown, recent studies revealed that they may be directly involved in transcription regulation. Abundant species of small RNAs, such as snoRNAs and microRNA, generated by processing of the spliced introns and large non-coding RNA genes-precursors, are involved in many different regulatory pathways of transcription, RNA processing and protein synthesis [10, 11]. Regulation of gene expression and heterochromatin formation via epigenetic mechanisms was established for several non-coding RNA
species such as Xist and Air. Furthermore, as a novel function of non-coding RNA it was shown that some of the several hundred novel RNA transcripts within the HOX-locus is directly involved in gene-specific transcription repression via regulation of a chromatin modifying enzyme [7]. Assuming that most of the different non-coding RNA transcripts found in the HOX loci may have a specific function, and considering that these non-coding genes can be specifically regulated in response to different stimuli, the gene-regulatory complexity can be tremendous, explaining processes such as the highly regulated expression of HOX genes in time and space during development. Extending such a regulatory network to the entire genome illustrates an enormous complexity of gene-regulatory mechanisms, which are responsible for a highly coordinated expression of every individual gene in every distinct cell in an organism.

Following the identification of the novel promoters, we determined the binding profiles for 25 general transcription factors measured by ChIP-on-chip on a large number of known and novel promoters using TBP binding sites microarray. We performed correlation-clustering analysis to measure the degree of difference (similarity-dissimilarity) between the binding profiles of these factors in relation to their function in transcription. This approach can be used to assess co-binding of the transcription factors at target sites: high correlation values between the binding profiles indicates that the factors are co-recruited to the loci, whereas factors that do not co-bind to the same targets will have a low correlation value. Using this method we found that the transcription factors could segregated into different groups corresponding to their involvement in RNAP I, -II, and -III specific transcription. This result showed that correlation-clustering analysis of ChIP-on-chip data can resolve functional interactions between transcription factors.

To address the specific involvement of the transcription factors in regulation of different classes of promoters, we used the correlation approach to analyze two distinct groups of targets: CpG-islands and non-CpG genomic regions. They represent promoters which are very different in their sequence composition suggesting different mechanisms of transcription initiation at these loci. CpG-type promoters usually lack core-promoter elements such as TATA-box, DPE or Inr, but contain multiple transcription start sites [18]. They usually contain a number of binding sites for transcription factors Sp1 and Sp3 which possibly direct the assembly of the pre-initiation complex [19]. Despite the fact that the promoters of about 80% of human protein-coding genes are located in CpG islands, their detailed functioning in transcription initiation remain not well defined. We found distinct clustering patterns for two groups of targets: TAFs had remarkably lower correlation with general transcription factors at the non-CpG targets, whereas the correlation at the CpG-targets was higher. This result suggests that TAFs and the other general factors are not (efficiently) co-recruited to non-CpG promoters but in contrast appeared to be stably recruited to CpG-island promoters. It is possible that the TAFs subunits of TFIIID recognize specific DNA elements in the CpG islands and therefore can stabilize and/or position preinitiation complex at these promoters. This suggests distinct functional interactions of TAFs and other RNAP II factors on non-CpG versus CpG-island targets pointing to distinct mechanisms of transcription initiation. Our data are in line with those obtained in yeast in which two different transcription initiation pathways were assigned to the two types of promoters [20, 21]. The TATA-less promoters corresponded to about 80% of yeast genes, mainly of a ‘housekeeping’ type, and most of them were dependent on the TFIIID complex. About 20% of the yeast genes have TATA-box containing promoters and were regulated by SAGA, a TBP-free TAF containing complex which can function in connection with the TAF-free form of TBP [21, 22]. Thus, our results suggest that the transcription initiation process at the core promoters of human genes may also be regulated via distinct mechanisms.
An unexpected finding of the clustering analysis was that RNAP II-specific factors such as the histone acetylase, PCAF, and a component of TFIID complex, TAF12, co-clustered also with RNAP I-specific factors. To confirm this observation, we showed that these factors bind to rDNA gene in vivo. In addition, TAF12 and PCAF stimulated RNAP I-transcription in a cell-based reporter assay and in a reconstituted transcription system. Furthermore, we found that TAF12 is a component of the endogenous SL1 complex. Thus, we provided evidence that TAF12 and PCAF are also involved in transcription by RNAP I. Collectively, our data suggest a molecular regulatory link between RNAP I and II.

A regulatory link between RNAP I, II and III transcription was previously suggested based on the highly coordinated transcriptional output of ribosomal components, such as 45S rRNA precursor, 5S rRNA and mRNA of over 70 ribosomal proteins that comprise over 50% of cellular RNA [12]. The molecular basis of this co-regulation remains unknown. The synthesis of the ribosomal components is linked with cellular processes and tightly controlled via different regulatory pathways in response to various cellular metabolic stimuli as well as environmental changes (chapter 1, part 2). For example, ribosome biosynthesis is co-regulated with cell growth and proliferation so that the number of cellular ribosomes is adjusted to the level of nutrients and growth factors as well as differentiation and development stages [13]. RNA polymerase I activity has been shown to play a central role in the regulation of synthesis of ribosomal components at the level of transcription because increased of RNAP I transcription led to a concomitant increase of RNAP II and III transcription [14]. In addition to these data, several studies described that different RNAP II-specific factors are required for transcription of RNA polymerase III [15, 16, 17]. Thus, the study presented in Chapter 2 together with accumulated data in the literature suggest that a number of (general) transcription factors required for different types of RNA polymerases provide a molecular link for a tight cross-regulation between transcription of different classes of genes.

A role of general factor TFIIA in transcription of human genes

The coordinated assembly of a preinitiation complex is an essential event in transcription initiation. Preinitiation complexes at promoters of the RNAP II transcribed genes consists of many multiprotein complexes such as RNA polymerase II, general transcription factors and (co)regulators [1, 2]. The general transcription factors are essential for transcription in vitro and for all the RNAP II genes in vivo. Nevertheless, the role of one of the general factors, namely TFIIA, remains controversial. This factor is essential for cell viability, but in highly purified in vitro transcription assays it is not required for accurate initiation [23] and cells remain viable upon significant reduction of endogenous TFIIA protein level in vivo [24]. In Chapter 3 we investigated the role of TFIIA by determining its binding profiles at a large number of loci in the human genome.

First, we measured the binding of the three TFIIA subunits (alpha, beta and gamma) on a set of known and novel promoters on the microarray platform which we had used for the profiling of the binding sites for 25 different transcription factors (described in Chapter 2). This allows the combined analysis of TFIIA with the other factors. To assess a possible functional relationship between TFIIA and these transcription factors, we performed correlation-clustering analyses (described in Chapter 2) on the entire dataset. Interestingly, we found that all three TFIIA subunits clustered at the same branch and at short distances from the TAFs and not with other general transcription factors. This result implied very similar binding profiles, i.e. these factors co-occupy their target promoters to the same relative extent and they are efficiently co-crosslinked to the same DNA sites. It suggested that TFIIA and TAFs form a stable complex on promoters and bind to the same sets of genomic loci. A
close functional relationship between these factors is strongly supported by numerous interactions of TFIIA with TAF1, TAF4, TAF5, TAF6 and TAF11 reported in the literature [25 - 28]. In addition, TFIIA and TAFs are similar with respect to their interactions with general transcription factors such as TFIIE and TFIIF and different activators suggesting a similar coactivator function [29]. Thus, these results suggested that TFIIA and TAFs may have a functional relationship and cooperate during preinitiation complex assembly at a subset of human promoters.

As a next step we mapped with high resolution the binding sites for TFIIA and TBP along a 80.3 Mb region on human chromosome 6. Remarkably, we found that a substantial number of genomic sites had insignificant TFIIA binding although high values for TBP occupancy were obtained. This result suggested that certain genes are little, or not at all, dependent on TFIIA although they are actively transcribed. Considering a close functional link between TAFs and TFIIA, our result may imply that the TFIIA-independent genes also do not require TAFs. In line with this suggestion, transcription of many genes in yeast is not independent on TAFs [30]. Furthermore, a TFIIA/TAFs independent mechanism would imply that TBP can function in transcription initiation independently of the TFIID complex. This hypothesis is also in line with our data described in Chapter 2 showing that TAFs may not be required for transcription of non-CpG promoters in human cells (Chapter 2), and also with the data from yeast where a TAF-free form of TBP can function in transcription initiation of a subset of genes [31].

Remarkably, we found that TFIIA binding correlated with a particular biological function of the genes. Gene ontology (GO) annotation revealed that histone genes, heat shock protein genes and several transcription factors and a number of metabolic enzymes appeared to have a high TFIIA binding, whereas the promoters of many other transcription factors (mostly zinc finger proteins) and various signal-transduction proteins (such as kinases), displayed low TFIIA binding. This suggested that these genes will not be significantly affected upon depletion of cellular TFIIA. Hence, a small decrease in overall mRNA levels obtained in yeast upon TFIIA reduction can be explained by unchanged transcription of the TFIIA-independent genes encoding transcription regulators [24]. The unchanged expression of signalling enzymes can explain the reported cell viability upon extensive depletion of TFIIA. In contrast, transcription of the TFIIA-dependent histone genes in yeast may be significantly reduced upon TFIIA reduction and lead to a shortage of histone proteins and disbalanced chromatin assembly and the observed G2/M cell cycle block [24]. Collectively, our and other results suggest that the requirement of general factor TFIIA in transcription may correlate with specific gene functions and possibly is linked to different transcription regulatory mechanisms of human genes.

**Resolving the topology of an active rDNA gene**

Actively growing cells require several millions of ribosomes per cell cycle. A proportional number of building components must be synthesized and the ribosomal RNA (rRNA) therefore comprises about 50% of the total cellular transcriptional output [32-35]. Transcription of rDNA genes by RNAP I occurs at a very high rate and a single rDNA gene is transcribed by hundreds of RNA polymerases simultaneously. This can be visualized at the cytological level by the fir-tree like structures corresponding to an extended rDNA genes associated with elongating polymerases connected to the radiating away nascent pre-rRNA transcripts of increasing length [35]. rRNA transcription is localized in the nucleolus, a non-membrane subnuclear compartment which further accommodates rRNA processing and ribosome assembly. Although the morphological organization of the nucleolus is well described, the spatial topology and positioning of rDNA gene and the large number of nascent transcripts within the nucleolar environment remains elusive.
In Chapter 4 we described the binding profiles of different transcription factors throughout the rDNA locus. We found that TBP and the TAF\textsubscript{110} subunits of SL1 complex could be efficiently crosslinked to the enhancer and terminator region R3 in addition to their crosslinking to the promoter. In addition, crosslinking of the SL1 subunits was detected also at different sites within the transcribed region of the rDNA gene. These specific binding patterns suggested stable contacts (direct or indirect via protein-protein contacts) of the SL1 complex with these DNA sites and implied a unique three-dimensional conformation of the rDNA gene during active transcription. To resolve this conformation we performed a spatial modeling of the rDNA topology in order to explain the specific transcription factors binding patterns. Considering the mechanism of formaldehyde crosslinking we postulated that the promoter, enhancer and terminator R3 regions should be spatially located in very close proximity to the SL1 complex and may form a specific structure which we named ‘central tetrad’. The crosslinking of SL1 to the transcribed region suggested that the rDNA must be compactly and uniformly positioned around the central tetrad. We proposed a spring-core model where the rDNA transcription region is wrapped as a cylinder around the central tetrad located inside of this cylinder. This positioning provides solutions for many unresolved questions about nucleolar functioning. In our model, the transcribing RNA polymerases are located on the outside of the spring-core structure and during RNA synthesis synchronously pull along rDNA resulting in a rotation of the entire spring-core structure. This organization suggests that the nascent transcripts radiate away around the cylinder and are isolated from each other preventing intertwining. This architecture would facilitate co-transcriptional access of the rRNA transcripts to snoRNPs during early processing events. Such a spatial organization provides a simple and efficient mechanism for disentangling RNA and DNA and allows simultaneous transcription by many polymerases in a highly organized manner.

The proposed three-dimensional structure of the rDNA gene presented in Chapter 4 integrates our current knowledge on the organization and functions of the nucleolar sub-compartments and the localization of transcription factors, DNA and RNA within the nucleolus. Our suggestion that the central nucleolar compartment, fibrillar component (FC), corresponds to the center of the spring-core structure found numerous supporting lines in previous studies. The position of the central tetrad in our model is supported by the facts that the inside part of the FC is the place for transcription initiation [36] and that RNAP I transcription factors including topoisomerase I are predominately localized within the FC [36–40]. Furthermore, rDNA regions connected to the central tetrad and located in the center of the spring-core structure can explain the nature of the extended nucleosome-free DNA fibrils positioned in the central part of the FC and maintained in a non-compacted state throughout the cell cycle [41]. In addition, our model implied that the transcribed region of rDNA genes and elongating RNA polymerases must be located at the periphery of the spring-core structure. This assumption is in agreement with high resolution in situ data that mapped the rDNA as well as active rRNA synthesis mainly at the outer border of the FC, whereas nascent pre-rRNA transcripts were invariably mapped inside of the electron dense narrow area surrounding the FC (the dense fibrillar components, DFC) [35, 36, 39]. This suggests that nascent transcripts while synthesized at the periphery of the FC enter into DFC where they associate with processing RNPs.

Thus, based on the transcription factors binding profiles at the rDNA gene we proposed the unique three-dimensional spring-core structure of an actively transcribed rRNA gene. Our model described a possible mechanistic principle of RNAP I transcription and provided a spatial basis for highly efficient rRNA synthesis. The model correlates very well with nucleolar morphology and the available data from in situ electron microscopy studies describing physical localization of DNA, RNA
Chapter 6

and transcriptional factors in the nucleolar compartments. Our study contributes to the further understanding of topological aspects of rDNA gene transcription.

E2F-dependent transcription repression in cellular senescence

Following a limited number of cell cycle divisions in cultured primary murine embryonic fibroblasts (MEFs) end in replicative senescence [42]. This can also be induced by introduction of the RAS

$^{V12}$ oncogene [43]. At the molecular level, cell cycle block and senescence are accompanied by increased levels of the p16

$^{INK4a}$, p19

$^{ARF}$, p53 and p21

$^{CIP1}$ proliferation suppressor proteins. These proteins mainly function in blocking of phosphorylation activity of cyclin dependent kinases (CDKs), which are critical positive regulators in cell cycle progression. The cell cycle inhibitory role of p16

$^{INK4a}$ can be mediated via co-factors such as the retinoblastoma proteins pRB, p107 and p130 [44, 45], whereas p19

$^{ARF}$ functions either via inactivation of E3-ubiquitin ligase MDM-2 [46], which results in stabilization of p53 and induction of its target genes including the cell cycle inhibitor p21

$^{CIP1}$ [47], or via p53-independent pathways [48]. MEFs without functional p19

$^{ARF}$ and p53 can be cultured indefinitely and do not undergo cellular senescence [43], but the mechanism of the p19

$^{ARF}$/p53 pathway critical in senescence signaling is not fully understood. It appears to depend on retinoblastoma proteins [43, 49] accumulating during senescence, which interact with a number of proteins involved in cell cycle regulation, such as MDM-2, PML, and E2F transcription factors [50].

The pRB proteins bind to the transactivation domain of 5 different E2F factors (E2F1 – E2F5) resulting in inhibition of their function [51] and formation of active transcriptional repressor complexes via association of pRB with HDACs [52]. This binding is transient during the cell cycle: pRB proteins are phosphorylated in G1 phase by cyclin D-CDK4, cyclin D -CDK6 and cyclin E-CDK2 resulting in disruption of E2F repressor complexes and activation of E2F-dependent transactivation which is required for the progression to the S phase [52, 53, 54]. In addition to pRB proteins, the role of E2F transcription factors in senescence was suggested by the ability of E2F-1 to activate ARF and by decreased levels of E2Fs in senescent cells [55, 56].

In Chapter 5 the role of E2Fs in senescence in MEFs was investigated in more detail. As an experimental system, the function of E2F in transcription signalling was disrupted via overexpression of mutated proteins followed by analysis of senescence response. We showed that overexpression of E2F-DB protein, lacking both the C-terminal transactivation and the pRB binding domains, resulted in displacement of endogenous E2Fs from their binding sites on chromatin. This resulted in significant activation of all the tested E2F target genes, such as p19

$^{ARF}$, PCNA, p107, and cyclins E1 and A, suggesting that in MEFs the endogenous E2F controls gene expression predominantly by means of active repression. Expression of E2F-DB in cells lacking all the three proteins (BR, p107 and p130) had no effects on the target genes indicating that the mechanism of repression is mainly regulated via BR proteins.

Surprisingly, the increased level of p19

$^{ARF}$ upon E2F-DB expression did not result in cell cycle block or senescence as would have been expected. Moreover, the cells completely failed to undergo senescence (both spontaneous and RAS

$^{V12}$-induced) in spite of expression of the cell cycle suppressors p19

$^{ARF}$, p16

$^{INK4a}$, p53 and p21

$^{CIP1}$. Furthermore, an antiproliferative effect of high level p19

$^{ARF}$ and p53 overexpressed in MEFs was significantly reduced upon expression E2F-DB suggesting that E2F functions downstream of both p19

$^{ARF}$- and p53-dependent signaling. Collectively, these results strongly suggest that, at least in certain cellular systems, repressor function of endogenous E2F is required for cells to respond to senescence-associated induction of p19

$^{ARF}$ and p53. The data show that E2F repression is linked to antiproliferative signaling by p19

$^{ARF}$. 

184
References

9. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res. 2006. 16:1–10
14. Laférté A, Favry E, Sentenac A, Riva M, Carles C, Chédin S. The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components Genes Dev. 2006. 20:2030-2040
Chapter 6


Addendum

Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4

Michiel Vermeulen, Klaas Mulder, Sergei Denissov, Pim Pijnappel, Frederik van Schaik, Radhika Varier, Marijke Baltissen, Henk Stunnenberg, Matthias Mann and Marc Timmers

Cell (2007) 131, 58 - 69
Selective binding of TFIID to histone H3K4me3

Summary
Trimethylation of histone H3 at lysine 4 (H3K4me3) is regarded as a hallmark of active human promoters, but it remains unclear how this posttranslational modification links to transcriptional activation. Using a stable isotope labeling by amino acids in cell culture (SILAC)-based proteomic screening we show that the basal transcription factor TFIID directly binds to the H3K4me3 mark via the plant homeodomain (PHD) finger of TAF3. Selective loss of H3K4me3 reduces transcription from and TFIID binding to a subset of promoters in vivo. Equilibrium binding assays and competition experiments show that the TAF3 PHD finger is highly selective for H3K4me3. In transient assays, TAF3 can act as a transcriptional coactivator in a PHD finger-dependent manner. Interestingly, asymmetric dimethylation of H3R2 selectively inhibits TFIID binding to H3K4me3, whereas acetylation of H3K9 and H3K14 potentiates TFIID interaction. Our experiments reveal crosstalk between histone modifications and the transcription factor TFIID. This has important implications for regulation of RNA polymerase II-mediated transcription in higher eukaryotes.

Introduction
The organization and functional state of chromatin is closely linked to posttranslational modifications of histones (Jenuwein and Allis, 2001). Histone modification patterns are involved in processes determining cell fate, development, and carcinogenesis ([Carrozza et al., 2003], [Margueron et al., 2005], [Santos-Rosa et al., 2002] and [Torres-Padilla et al., 2007]). Modified histones are thought to serve as binding scaffolds for regulatory proteins that translate these modifications into physiological responses. Genome-wide localization studies show that trimethylation of histone H3K4 (H3K4me3) marks active promoters in human cells in a pattern very similar to H3K9 and K14 acetylation ([Bernstein et al., 2005] and [Heintzman et al., 2007]). Recently, several interactors of methylated H3K4 have been identified. These interactors bind to methylated H3K4 via different domains, such as WD-40, Tudor, MBT, and the plant homeodomain (PHD) ([Kim et al., 2006] and [Ruthenburg et al., 2007]).

The genome-wide distributions of H3K4me3 and H3K9,14Ac are highly similar to that of the human TAF1 protein ([Bernstein et al., 2005] and [Heintzman et al., 2007]), which represents the largest subunit of TFIID. Promoter recruitment of this basal transcription factor is stimulated by genespecific transcription factors and transcriptional cofactors. Besides TAF1, TFIID harbors TATA-binding protein (TBP) in association with 12–13 additional TBP-associated factors (Tora, 2002). In addition to histone-fold motifs, which are important for TAF-TAF interactions within TFIID (Leurent et al., 2004 and references therein), several other conserved motifs (HMG, NHR1, DUF1546, LisH, WD-40, PHD) have been identified. The relevance of these domains, if any, for promoter activation is not clear yet. In contrast, the double bromodomain of metazoan TAF1 has been shown to preferentially bind diacetylated histone H4 peptides and with a lower affinity also binds monoacetylated peptides (Jacobson et al., 2000). To summarize, while it is clear that histone acetylation and methylation cooperate in the activation of promoters, the crosstalk between these modifications at the level of the basal transcription machinery has not been explored fully.

Mass spectrometry (MS)-based proteomics has become a powerful tool in biology and for identification of histone modifications, in particular ([Aebersold and Mann, 2003] and [Taverna et al., 2007]). We have previously used quantitative proteomics to dissect regulatory pathways ([Mann, 2006], [Olsen et al., 2006] and [Ong and Mann, 2005]). Here we employ a screen to identify peptide-protein interactions by stable isotope labeling by amino acids in cell culture (SILAC) (Schulze and Mann, 2004). In addition to several known interaction partners for H3K4me3, we unexpectedly
Addendum

identified subunits of the basal transcription factor TFIID as specific binders. We show that H3K4me3 binding is mediated by the PHD finger of the TAF3 subunit and that the methylation mark is essential for this PHD to bind native nucleosomes in vitro and in vivo. Overexpression of TAF3 can enhance transcription stimulation by the Ash2L subunit of Set1/MLL (mixed-lineage leukemia) histone methyltransferases in a PHD-dependent manner. Furthermore, we demonstrate that asymmetric dimethylation of H3R2 can selectively inhibit, and that acetylation of H3K9 and H3K14 potentiates, TFIID interaction with the K4-methylated H3 tail. These experiments reveal a direct connection between the basal transcription apparatus and transcriptionally active chromatin and expand our view of how histone modifications support transcriptional activation by TFIID.

Results

TFIID Binds to Trimethylated Histone H3 Lysine 4 In Vitro and In Vivo

In order to identify novel proteins interacting with H3K4me3, we set up a peptide pull-down approach making use of the SILAC technology. In SILAC, one or more amino acids are replaced with their stable isotope-labeled counterpart, thus generating “light”- and “heavy”-labeled proteins, which are distinguishable and quantifiable by MS. The experimental approach for identification of interaction partners of modified histones is depicted in Figure 1A. Nuclear extracts derived from human HeLaS3 cells grown in light or heavy medium are incubated with immobilized histone peptides in the nonmethylated and methylated form, respectively. After incubation and extensive washing, beads from both pull-downs are pooled, boiled in loading buffer, and run on an SDS-PAGE gel. Following in-gel trypsin digestion and peptide extraction, the peptide mixtures were analyzed by high resolution MS on a linear ion trap-orbitrap instrument. The vast majority of proteins (>99%) bound equally well to the methylated and nonmethylated peptide and appeared in the mass spectra with equal intensity in the light and heavy form. Proteins interacting in a methylation-dependent manner are present with a higher intensity in the heavy form. This quantitative filter therefore distinguishes specific binders.

When performing this screen for H3K4me3, we identified several known interactors with a significant SILAC ratio. They included BPTF (Figure 1B) and CHD1 (Figure 1C), which have been reported to bind methylated H3K4 via a PHD finger and a double chromodomain, respectively ([Wysocka et al., 2006] and [Li et al., 2006]). A number of other novel interactors were identified, which included several PHD finger-containing proteins (see Table S2). Surprisingly, we also identified all TAF subunits of TFIID with highly significant ratios (30:1), indicating that the TFIID complex specifically binds to trimethylated H3K4 (Figures 1D, 1F, S1–S5, and Table S1). The TATA-binding protein (TBP) showed a SILAC ratio of 4, substantially lower than the ratios of the TFIID TAFs. While still significant, this lower ratio can be explained by the presence of TBP complexes lacking TFIID TAFs, which would bind nonspecifically to bait and control. This is also exemplified by the TAFI63 (or TAF1B) subunit of SL1 (ratio 0.98, Figure 1F). We verified the preferential interaction between TFIID and H3K4me3 by immunoblotting (Figure 1E).

To investigate the functional consequences of this interaction we performed RNAi experiments against WDR5, which is known to preferentially reduce global levels of the H3K4me3 marks but not of H3K4 me2 and H3K4me1 ([Dou et al., 2006] and [Steward et al., 2006]). Figure 2A shows that WDR5 siRNA treatment of U2OS cells indeed resulted in a pronounced reduction of H3K4me3. We examined the mRNA levels of a number of genes to identify promoters sensitive to H3K4me3 loss. As indicated by Figure 2B, WDR5 siRNA treatment reduced the mRNA levels of the HMG-CoA reductase, RPL34, RPS10, and RPL31 genes by about 2-fold. In contrast, fibronectin and the β-actin reference mRNAs were not affected. Analysis of chromatin immunoprecipitation (ChIP) samples by
quantitative PCR showed that the WDR5 siRNA reduced TBP association to the HMG-CoA reductase but not to the fibronectin core promoter (Figure 2C). As expected, H3K4me3 modification was reduced on both promoters. An extensive ChIP analysis of TBP and TAF1 binding was performed for the RPL34, RPL31, and RPS10 genes by scanning their 5′ regions using different PCR primer pairs (Figures 2D, 2E and S6). WDR5 siRNA treatment also reduced TBP and TAF1 association to these promoters. TBP and TAF1 ChIP signals were also detected more downstream overlapping with the peak of H3K4me3. This corresponds well with recently published ChIP-chip results for these factors ([Bernstein et al., 2005] and [Heintzman et al., 2007]).

Taken together, these results indicate that the entire TFIID complex but not other TBP-containing complexes specifically associates with H3K4me3 peptides in vitro and that the H3K4me3 mark is important for retention of TFIID and the transcriptional activity of a subset of promoters in vivo.

Figure 1. The TFIID Complex Binds to Trimethylated Histone H3 Lysine 4 In Vitro. 
(A–D) Schematic representation of the SILAC-based histone peptide pull-down approach (Schulze and Mann, 2004). Representative spectra of BPTF (B), CHD1 (C), and TAF1 (D) peptides. The heavy peptide from the SILAC pair is clearly more intense, demonstrating specific binding of these proteins. The gray and red ovals indicate the relative amount of protein binding to the unmodified or H3K4me3 peptide, respectively.

(E) The preferential interaction between TFIID and H3K4me3 was confirmed by pull-downs of nuclear extracts using unmodified (−), H3K4me3 or H3K9me3 peptides. Bound proteins were analyzed by immunoblotting using TAF1 and HP1α antibodies.

(F) Quantification of the interaction of all TFIID-TAFs with H3K4me3 as revealed by SILAC screening. SILAC ratios represent the relative abundance of the heavy to the light peptide indicating specific binding to the H3K4me3 peptide.
Addendum

Figure 2. TFII D Occupancy and mRNA Expression Dependence on H3K4me3 In Vivo.
(A) Immunoblot analysis on U2OS cell lysates transfected with WDR5 or GAPDH control siRNAs.
(B) Quantitative RT-PCR analysis of siRNA-treated cells. Levels of the different mRNA were normalized on β-actin mRNA levels, which were not affected by the WDR5 knockdown. Error bars indicate standard deviation of three to four biological replicates.
(C) ChIP experiments were performed with siRNA-treated cells to assess H3K4Me3 levels and TBP occupancy on the indicated pol II promoters. An H3 core domain antibody was used as a control. The results are represented as percentage of input for histone antibodies and as folds over background for TBP. Dark gray and light gray bars indicate results with WDR5 or GAPDH control siRNA-transfected cells, respectively. Error bars indicate standard deviation of three biological replicates.
(D and E) A representative ChIP analysis of TAF1, TBP, and H3K4me3 association to the RPL34 and RPL31 loci in siRNA-treated cells. H3K4me3 levels are relative to total H3 levels. TBP and TAF1 levels are represented as percentage of input. Location + 1 indicates the transcription start site.

The TAF3-PHD Finger Directly Binds to Trimethylated Histone H3 Lysine 4
Several papers described interactions between PHD fingers containing proteins (including BPTF) and methylated H3K4 or H3K36 peptides ([Li et al., 2006], [Pena et al., 2006], [Shi et al., 2006] and [Wysocka et al., 2006]). Interestingly, the metazoan TAF3 subunit of TFII D contains a PHD finger at its extreme C terminus ([Gangloff et al., 2001] and [Pointud et al., 2001]). Alignment of the PHD fingers of TAF3 and ING2 (Figure 3A) revealed conservation of residues involved in H3K4me3 interaction (Ruthenburg et al., 2007). This suggests that the PHD finger of TAF3 could be critical for the interaction between TFII D and methylated H3K4. To test this hypothesis we generated HeLa cell lines expressing tagged versions of full-length mouse TAF3 or TAF3ΔC80 lacking the C-terminal PHD finger (see Figure S7). Nuclear extracts derived from these cell lines were tested for binding to H3 peptides. Whereas the tagged full-length mTAF3 protein bound specifically to the H3K4me3 peptide (Figure 3B, compare lane 1 and 2), no binding of the tagged mTAF3ΔC80 protein could be observed (lane 3 and 4). The TAF7 subunit of TFII D was also retained by the H3K4me3 peptide, but to a lesser extent in the TAF3ΔC80 extracts (compare lanes 2 and 4). This may be explained by a dominant-negative effect of the TAF3ΔC80 protein, which replaces part of the endogenous TAF3 in
Selective binding of TFIID to histone H3K4me3

the pool of cellular TFIID. Further analysis indicated that the tagged TAF3 proteins are incorporated into TFIID complexes (Figure S7). Thus, TFIID complexes harboring the truncated TAF3ΔC80 subunit are defective for H3K4me3 binding.

Figure 3. The Isolated PHD Finger of TAF3 Binds with High Specificity to H3K4me3.
(A) Domain representation of the TAF3 subunit of TFIID indicating the N-terminal histone-fold domain (residues 9–82) and C-terminal PHD finger (residues 870–914). Alignment of the PHDs of mouse TAF3 (862–918), human TAF3 (860–914), mouse ING2 (205–266), and human ING2 (352–406). Arrowheads indicate positions of mutants tested in (D).
(B) Nuclear extracts derived from HeLa stable cell lines expressing either full-length HA-tagged TAF3 or a mutant lacking the PHD finger were used for pull-downs using the indicated peptides. Immunoblotting was performed for the presence of endogenous TAF7 and exogenous TAF3 proteins in the pull-down eluates. Please note that the TAF7 signal in the H3K4me3 pull-down of the TAF3ΔPHD lysate most likely corresponds to TFIID complexes containing endogenous TAF3.
(C) Bacterial lysates containing GST-mTAF3 (857–924) were incubated with streptavidin-beads coated with the indicated H3 peptides. Bound proteins were analyzed by Coomassie staining. Arrowheads indicate position of the GST-fusion protein.
(D) Wild-type and mutant TAF3-PHD proteins (as indicated to the right) were analyzed as in (C) for binding to K4-methylated peptides as indicated above.
(E) Competitive binding of GST-fusion proteins of the PHD of mouse ING2 and TAF3 to H3 peptides indicated to the right. The top panel shows 2.5% of the input. The percentage of TAF3-PHD lysate is 0%, 5%, 10%, 20%, 50%, 80%, and 100% (from left to right) of the total used for binding.
(F) Dissociation constants of TAF3-PHD binding to H3 peptides (residues 1–17) carrying the indicated modifications as determined by tryptophan fluorescence.

To further investigate this we assessed the ability of a GST fusion of the isolated TAF3-PHD finger to bind methylated histone peptides. As shown in Figure 3C, the PHD finger bound very efficiently to
methylated H3K4, with a clear preference for H3K4me3, whereas no binding to methylated H3K9 or H3K36 was observed. Based on the sequence alignment we mutated selected residues of the TAF3-PHD finger expected to be involved in H3K4me3 binding (Figure 3A). GST-TAF3 PHD fusions carrying the M882A, D886A, and D890A/W891A mutations were unable to bind H3K4me3 (Figure 3D). The control A901V mutation did not affect binding efficiency. Previous analysis showed that mutation of equivalent residues in the PHDs of mouse ING2 or human BPTF reduced H3K4me3 binding 20- to 500-fold ([Li et al., 2006] and [Pena et al., 2006]). Our mutational analysis underscores the specificity of the interaction and verifies the importance of several residues conserved between PHD fingers (Ruthenburg et al., 2007). To directly compare the affinities between the PHD fingers of TAF3 and ING2 for H3K4me3, we performed a competition experiment under conditions of PHD finger excess over peptide. As shown in Figure 3E, the ING2 and TAF3-PHD finger display similar binding affinities for H3K4me2, whereas the TAF3 PHD clearly has a much higher affinity for H3K4me3 compared to the ING2 PHD. This indicates that the TAF3-PHD finger discriminates strongly between the H3K4me2 and H3K4me3 marks.

To accurately determine the affinities for modified H3 peptides we performed equilibrium binding assays with the TAF3-PHD finger using tryptophan fluorescence. Using the same methodology, the dissociation constant of the ING2 PHD for H3K4me3 was determined to be 1.5 µM (Pena et al., 2006). In our experiments and as expected from Figure 3E, binding of the TAF3 PHD to the H3K4me3 peptide was much stronger with a dissociation constant of 0.16 µM (Figure 3F). The higher affinity of the TAF3 PHD compared to ING2 PHD most likely relates to subtle differences in the geometry of the aromatic cage involved in binding of the trimethyl lysine (I869 and W891 in mTAF3 versus Y215 and W238 in mING2). Structural analysis of the PHD finger of TAF3 will provide further insight into this. Taken together, H3 peptide-binding assays identified the PHD finger of TAF3 as necessary and sufficient for recognition of the H3K4me3 mark.

The TAF3 PHD Finger Binds to Nucleosomes Containing Trimethylated Histone H3K4

The experiments described so far indicate that the TFIID complex binds via the PHD of TAF3 with a high selectivity to histone H3K4me3 peptides. To extend these observations we isolated nucleosomes from HeLa cells and various yeast strains (Figure 4A and data not shown) to investigate whether the TAF3 PHD finger is able to recognize the H3K4me3 mark in native chromatin. As shown in Figures 4B and 4C, immunoblotting with modification-specific antibodies revealed that the bound nucleosomes were enriched for H3K4me3 and contained only low levels of H3K4me2 or H3K4me1. DNA analysis of bound yeast nucleosomes indicated retention of mono-, di-, and trinucleosomes (Figure 4C). In agreement with the peptide pull-down, the two PHD mutants D886A and D890A/W891A were unable to bind yeast nucleosomes, indicating the specificity of binding. To address the importance of the H3K4 trimethylation mark for the binding, we isolated nucleosomes from yeast strains carrying deletions in genes responsible for different histone methylation marks (Figure 4D). The SET1, SET2, and DOT1 genes encode histone methyltransferase (HMT) enzymes for H3K4, H3K36, and H3K79, respectively (Martin and Zhang, 2005). SPP1 encodes a subunit of the Set1p/COMPASS HMT complex and has been shown to be important for efficient H3K4 trimethylation, but not di- or monomethylation of H3K4 ([Dehe et al., 2006] and [Schneider et al., 2005]). As predicted the TAF3 PHD finger binds nucleosomes from set2Δ and dot1Δ strains indicating that neither H3K36 nor H3K79 methylation is required for binding to nucleosomes (Figure 4E). However, in the set1Δ and spp1Δ strains, in which H3K4 (tri-)methylation is abolished,
Selective binding of TFIID to histone H3K4me3 nucleosome binding was lost. These results indicate that the TAF3 PHD finger binds specifically to native H3K4me3-marked nucleosomes.

Figure 4. The PHD Finger of TAF3 Specifically Interacts with H3K4 Trimethylated Native Nucleosomes. (A) Soluble native mononucleosomes from HeLa cells were analyzed by agarose-gel electrophoresis for DNA content and SDS-PAGE followed by Coomassie staining for histones. (B) Beads were coated with GST or GST-TAF3PHD and incubated with native HeLa nucleosomes. Bound material was analyzed by immunoblot analysis using modification-specific antibodies. (C) Beads were coated with GST or GST-TAF3PHD (WT or mutants) and incubated with native yeast nucleosomes. Precipitated material was analyzed as in (A) and (B). (D) Nucleosomal preparations isolated from various mutant yeast strains were subjected to immunoblot analysis to determine the H3 methylation status. (E) The indicated nucleosomal preparations were subjected to GST pull-down and analysis as in (B).

Asymmetric Dimethylation of H3R2 Selectively Inhibits Recognition of the H3K4me3 Mark by TFIID and TAF3 PHD

The structure of the WDR5 protein bound to an N-terminal peptide of histone H3 indicated that methylation of H3R2 negatively affected WDR5 binding (for review, see Sims and Reinberg, 2006). Furthermore, in yeast, asymmetric dimethylation of H3R2 inhibits the binding of the PHD finger protein SPP1 to methylated H3K4 (A. Kirmizis and T. Kouzarides, personal communication). To investigate a potential inhibitory effect of H3R2 methylation on the H3K4me3 binding of PHD finger
Addendum

proteins we compared several differentially methylated histone H3 peptides. This revealed that asymmetric dimethylation of H3R2 drastically reduces the binding of the TAF3 PHD to H3K4me3, whereas this effect was not observed for the ING2 or BPTF PHD (Figure 5A). This was confirmed in a competition experiment (Figure 5B), which revealed that the ING2 and TAF3-PHD finger have a similar affinity for the doubly methylated peptide (R2me2a/K4me3). In agreement with the GST-PHD pull-downs, H3R2me2a results in a 25-fold lower binding constant of the H3K4me3 modified peptide, whereas H3R2me2s and H3R2me1 display little effect (Figure 3F). Together, these analyses indicate that asymmetric dimethylation of H3R2 counters the high affinity of the PHD finger of TAF3 but not of ING2 and BPTF for the H3K4me3 mark.

Figure 5. H3R2me2a Selectively Inhibits Binding of TFIID to H3K4me3.
(A) Pull-downs using the indicated peptides were performed as described in Figure 3.
(B) Competition experiment between GST-TAF3-PHD and GST-ING2-PHD as described in Figure 3E using the H3K4me3 and H3R2me2a/K4me3 peptides.
(C) A triple SILAC peptide pull-down was performed using an unmodified (H3), an H3K4me3, and an H3K4me3/H3R2me2a doubly methylated peptide. The spectra show the relative binding of TAF1 (left), TAF7 (middle), and BPTF (right) to these three peptides as indicated. The gray ovals represent the relative protein binding to the unmodified H3 peptide, the red ovals represent the binding to the H3K4me3 peptide, and the black ovals represent the binding to the doubly modified H3K4me3/H3R2me2a peptide. Note that the ratio for the TAF subunits is somewhat lower compared to those shown in Figure 1, which is caused by a variation in washing efficiency.

To further investigate the effect of asymmetric dimethylation of H3R2 on the binding of the TFIID complex to trimethylated H3K4, we extended our quantitative proteomics approach to simultaneously assay three different histone modification marks (H3 unmodified, H3K4me3, and H3R2me2a/K4me3)
Selective binding of TFIID to histone H3K4me3

with three differentially SILAC-labeled extracts. This assay allows direct visualization of potential agonistic and antagonistic effects of modifications. The spectra in Figure 5C show that H3K4me3 binding by the TFIID complex (exemplified by its subunits TAF1 and TAF7) is compromised by the H3R2me2a modification. Importantly, binding of the BPTF protein was not affected (Figure 5C), indicating that the effect is specific to the TAF PHD finger. Together, these experiments reveal crosstalk between different histone modifications at the level of TAF3-PHD and TFIID-complex binding.

Acetylation of H3K9 and H3K14 Is Agonistic to H3K4me3

Similar to trimethylation of H3K4, acetylation of the adjacent lysines (K9 and K14) is generally associated with active promoters (Millar and Grunstein, 2006). The TAF1 subunit of metazoan TFIID contains an acetyl-lysine binding activity residing in its double bromodomain (Jacobson et al., 2000). We compared the combined effects of acetylation and methylation of the histone H3 tail on TFIID and BPTF binding using the SILAC triple pull-down approach. As shown in Figure 6A, K9 and K14 acetylation had a minor effect on retention of TAF1 (and other TFIID TAFs, data not shown). BPTF, which contains both a PHD and a bromodomain, showed a similar binding behavior. The presence of the H3K4me3 mark in addition to H3K9,14Ac strongly augmented retention of TAF1 (and the other TFIID TAFs) and of BPTF (additional increase in SILAC ratio 10-fold). In contrast, the BAF180/polybromo subunit of the human BAF chromatin-remodeling complex, which contains multiple bromodomains but no PHD finger (Nicolas and Goodwin, 1996), bound equally strong to the H3K9,14Ac and H3K4me3/K9,14Ac peptides (SILAC ratio > 8-fold over the unmodified peptide; right panel in Figure 6A).

We next tested whether acetylation enhances the binding of TFIID to H3K4me3 in an experiment comparing unmodified H3 with H3K4me3 and H3K4me3/K9,14Ac peptides. As shown in Figure 6B, both TAF1 and BPTF binding to H3K4me3 was markedly increased by acetylation at H3K9 and H3K14. As expected we failed to observe preferential binding of BAF180/polybromo to the H3K4me3 peptide, whereas the protein was retained efficiently by the H3 peptide containing both acetylation and methylation marks. Affinity measurements and H3 peptide pull-downs also indicated that acetylation on K9 and K14 had little effect on the affinity of the isolated PHD of TAF3 for the H3K4me3 peptide (Figures 3F and S8).

Collectively, these experiments indicate that both for TFIID and BPTF the PHD-mediated interaction to H3 tails is augmented by K9 and K14 acetylation. We propose that the combinatorial effects of H3K4me3 binding via TAF3 and H3K9,14Ac binding via TAF1 results in a strong interaction between TFIID and transcriptionally active promoters.

TAF3 Can Act as a Transcriptional Coactivator in a PHD Finger-Dependent Manner

To obtain further support for the hypothesis that TAF3 acts as a transcriptional cofactor by recognizing H3K4me3 we examined the effect of siRNA-mediated knockdown of TAF3 on the expression of the WDR5-dependent genes. Quantitative RT-PCR analysis indicated that similar to the WDR5 siRNA, treatment of U2OS cells with TAF3 siRNA reduced the mRNA levels for HMG-CoA reductase, RPL34, RPL31, and RPS10 but not for fibronectin (Figure 7A). Next, we employed a luciferase reporter assay in which the Ash2L subunit of Set1/MLL histone methyltransferase complexes was targeted to a test promoter as a Gal4 DNA-binding domain fusion. The evolutionary conserved Ash2L protein has been shown to be essential for efficient H3K4 di- and trimethylation and incorporates into the MLL1 complex via direct interaction with the Rbpb5 subunit ([Dou et al.,
Addendum

We reasoned that this could provide a functional assay for TAF3-dependent activation of transcription. Expression of the Gal4-Ash2L resulted in activation of promoter activity in human U2OS osteosarcoma cells. Coexpression of wild-type TAF3 greatly enhanced Gal4-Ash2L-mediated activation in a Gal4 binding site-dependent manner (Figure 7B). In contrast, the PHD finger point mutants M882A and D886A were not efficient in coactivation. Immunoblotting of transfected cell lysates indicated a similar expression of the transfected proteins (Figure 7C). Together, these experiments indicate that TAF3 can act as a transcriptional cofactor and support the model of Figure 7D in which methylation of histone H3K4 provides a binding site for the TFIID complex, resulting in enhanced recruitment or stability of the RNA polymerase II preinitiation complex.

Figure 6. Acetylation of H3K9 and K14 Acts Agonistically with H3K4me3 to Anchor TFIID on the Histone H3 Tail.
(A and B) Triple SILAC pull-downs were performed using peptides as indicated to investigate the relative contribution of H3K4me3 and H3K9,14Ac on TFIID binding.

Discussion

In this study we provide evidence that trimethylation of lysine 4 of histone H3 serves as a high-affinity binding site for mammalian TFIID and that the PHD finger of the TAF3 subunit mediates this interaction. We investigated crosstalk of H3K4me3 with other modifications and found an inhibitory effect of H3R2me2a on TFIID binding, whereas H3K9,14Ac enhanced TFIID binding. This
Selective binding of TFIID to histone H3K4me3 correlates well with active transcription and the genome-wide distributions of these histone marks and TAF1 ([Bernstein et al., 2005], [Heintzman et al., 2007] and [Millar and Grunstein, 2006]).

![Graphs and images](image)

**Figure 7. Transcriptional Activation by TAF3.**

(A) Quantitative RT-PCR analysis of siRNA-treated cells. Levels of the different mRNA were normalized on β-actin mRNA levels, which were not affected by the TAF3 knockdown. Error bars indicate standard deviation of two biological replicates. Knockdown of TAF3 expression was assessed at the mRNA level as appropriate TAF3 antibodies are not available.

(B) U2OS cells were transiently transfected in triplicates (errors bars indicate standard deviations) with the 5XGal4MLP-Luc reporter plasmid and pMT2-HA empty vector or pMT2-HA expression plasmids for the wild-type TAF3 or its M882A and D886A mutants in the presence or absence of Gal4-Ash2L as indicated below the panels. The experiment shown in the left panel is representative of three biological replicates. In the right panel U2OS cells were transfected in duplicates with the luciferase reporter construct TK-Luc lacking Gal4-binding sites.
and in the presence or absence of Gal4-Ash2L and TAF3 plasmids as indicated. The graphs represent the fold activation relative to the transfection with Gal4-Ash2L alone.

(C) U2OS cells were transfected with the indicated constructs and analyzed for protein expression by SDS-PAGE and immunoblotting. The tubulin-loading control indicates that less protein was loaded in the TAF3-D886A lane.

(D) Model depicting the different chromatin- and DNA-mediated interactions involved in anchoring the TFIID complex to the core promoter. In red stars the H3K4me3 methylation mark recognized by the PHD domain of TAF3 is indicated and in yellow the (di-)acetylation mark interacting with the double-bromo domain of TAF1. In addition, DNA-sequence specific interactions of TBP with the TATA box, of TAF1/2 with the initiator element, and of TAF6/9 with the downstream promoter element (DPE) are shown.

Function of the H3K4me3 Mark as a Docking Site
Methylation of H3K4 is a universal histone mark in the eukaryotic kingdom, and it has been studied in most detail in yeast cells. Trimethylation of H3K4 has been proposed to link transcription initiation to elongation and mRNA processing ([Krogan et al., 2003] and [Morillon et al., 2005]). Strikingly, yeast TAF3 lacks a PHD finger, indicating that this domain has been acquired relatively late in evolution (Gangloff et al., 2001). Several observations suggest that H3K4me3 may serve new or additional roles in higher eukaryotes. First, the H3K4me3 mark remains limited to the first 1 kb of a gene ([Bernstein et al., 2005] and [Heintzman et al., 2007]). Accordingly, the H3K4me3 modification is much less abundant in human nucleosomes compared to yeast (Figure 4) (Taverna et al., 2007). Second, whereas the yeast genome encodes 18 PHDs in 14 different proteins, the PHD is much more prevalent in higher eukaryotes genomes (38 PHDs in flies and 150 in humans) (Bienz, 2006). The observation that eight of the 18 yeast PHDs are able to bind methylated H3K4 (Shi et al., 2007) raises the question how many of the human PHD fingers can bind methylated H3K4. Our proteomic screen of HeLa cell extracts identifies seven PHD finger-containing proteins specifically binding H3K4me3 (Table S2). Importantly, multiple complexes with opposing activities are now known to compete for H3K4me3 binding: repression via ING2/Sin3A (Shi et al., 2006), chromatin remodeling via BPTF/NURF (Wysocka et al., 2006), and gene activation by TFIID (this paper). The finding that the TAF3 PHD has a 10- to 20-fold higher affinity for H3K4me3 than the PHDs of ING2 and BPTF indicates that via TAF3 TFIID can effectively compete with these (opposing) activities.

Relevance of TFIID Binding to H3K4me3 for Transcription
Current models for the function of H3K4 methylation are largely derived from yeast studies. They are supported by an elegant in vitro study using human proteins, which suggests that the H3K4me3 appears after the basal transcription machinery has assembled on the promoter (Pavri et al., 2006). In this scenario TAF3-dependent anchoring of TFIID could stimulate subsequent rounds of transcription. Other observations suggest the possibility that in human cells H3K4 methylation can act at the promoter activation step since gene-specific transcription factors like activated nuclear receptors and CREB can interact directly with subunits of MLL complexes ([Dreijerink et al., 2006], [Ernst et al., 2001], [Garcia-Bassets et al., 2007] and [Lee et al., 2006]). We found that loss of the H3K4me3 mark resulted in a reduced TFIID association and transcriptional activity of certain promoters, which also depend on TAF3 expression. In addition, transient transfections (Figure 7B) show that expression of TAF3 enhances the transcriptional stimulation by the Ash2L subunit of H3K4 methylase complexes. Taken together, this suggests that TAF3-mediated binding of TFIID to H3K4me3-marked nucleosomes can serve either to anchor TFIID to activated promoters or to recruit TFIID during promoter activation.

Recruitment of TFIID to gene promoters is governed both by its interaction with gene-specific transcription factors and coactivators and by the “quality” of the core promoter (Albright and Tjian,
Selective binding of TFIID to histone H3K4me3

2000). Whereas TBP binds with high affinity (Kd 1 nM) to a canonical TATA box, several other TFIID subunits can recognize core promoter elements like the INR (TAF1/2), DCE (TAF1), (DPE TAF6/9), and MTE (TAF1) with lower affinities ([Juven-Gershon et al., 2006] and [Lee et al., 2005]). Only a subset of mammalian promoters contains a classical TATA box, which stresses the importance of the other DNA elements. Our study indicates that histone modifications of the core promoter should also be taken into consideration (Figure 7D). A previous study of the yeast PHO5 promoter showed that nucleosome sliding can induce dependence on acetylation of the H4 tail and a bromodomain-containing protein ([Martinez-Campa et al., 2004]). We propose that H3K4 trimethylation plays an important role in defining a core promoter as it creates a binding site for the TAF3 subunit of TFIID. Strikingly, the fibronectin and β-actin promoters contain a TATA box and are insensitive to loss of H3K4me3 or TAF3. In contrast to the H3K4me3/TAF3-dependent HMG-CoA reductase promoter, the RPL34, RPL31, and RPS10 promoters all lack a canonical TATA box. It is tempting to speculate that the TAF3-H3K4me3 interaction may be most important for promoters lacking canonical core promoter DNA elements. Large-scale identification of TAF3-dependent promoters would allow testing of these models and determine the relevance of the H3K4me3-TFIID interactions for their transcription.

Crosstalk between Different Histone Modifications

An important aspect of the “histone code” hypothesis is that the modifications act in combination ([Strahl and Allis, 2000]). Our analysis provides two new examples of combinatorial control for binding to methylated H3K4. For the isolated TAF3 PHD we observed that asymmetric methylation of H3R2 inhibits H3K4me3 binding of TAF3 but not of ING2 or BPTF. The H3R2me2a modification also inhibits binding of the TFIID complex. Genome-wide analysis indicated the co-occurrence of the H3K4me3 and H3K9,14Ac marks in human cells ([Bernstein et al., 2005] and [Heintzman et al., 2007]). Our proteomic analysis clearly shows that TFIID binding to H3K4me3 is enhanced by coincident H3K9,14Ac modification. Most likely, this effect is mediated by the double bromodomain of TAF1. Whereas this domain preferentially binds to diacetylated H4 tails, it also binds to singly acetylated tails with low affinity (Kd 40 µM) ([Jacobson et al., 2000]). Consistently, the TFIID complex shows little enrichment with H3K9,14Ac compared to the nonacetylated peptide in the SILAC experiments. The dominance of H3K4me3 in TFIID binding can be readily explained by the higher affinity of the TAF3 PHD for this mark.

In conclusion, we have employed high-accuracy quantitative proteomics using the SILAC technology to discover a direct connection between histone trimethylation and the basal transcription machinery. One striking observation described in this work is the remarkable affinity of the TAF3 PHD finger for H3K4me3 and the cooperative effect with acetylation of H3K9,K14. These facts and given the general presence of H3K4me3 on active promoters as well as the broad importance of the TFIID complex for eukaryotic transcription suggest that the interaction between TFIID and H3K4me3 is highly relevant for the activation of genes in mammalian cells. An interesting question that remains is whether the interaction between H3K4me3 and TFIID is secondary following promoter recruitment or whether the interaction can serve as a primary recruitment signal for TFIID to promoters. Our findings direct future experiments to resolve these two possibilities.
Addendum

**Experimental Procedures**

**Protein Extraction and Peptide Pull-Downs**

HeLaS3 cells were grown in suspension in the presence of normal lysine and arginine or heavy versions of these (Isotec). Nuclear extracts were prepared as described (Dignam et al., 1983). Histone peptides containing the N-terminal 17 amino acids of the histone H3 followed by two glycines and a biotinylated lysine were synthesized using the Fmoc strategy as described (Schulze and Mann, 2004). Peptides were immobilized on Dynabeads MyOne Streptavidin C1 (Dynal) and subsequently incubated with 350 µg of nuclear extract (diluted to 0.6 mg/ml) in binding buffer (150 mM NaCl, 50 mM Tris-HCl at pH8.0, 1% NP40, 1 mM DTT, and complete protease inhibitors -EDTA [Roche]) for 2 hr at 4°C in a rotation wheel. Beads were washed five times with 1 ml of binding buffer containing 400 mM NaCl and finally twice with 1 ml of binding buffer. Beads from both pull-downs (unmodified and modified peptide) were subsequently pooled, and bound proteins were eluted and analyzed on 4%–12% gradient gels (Invitrogen) by colloidal blue staining (Invitrogen). For immunoblotting, antibodies against TAF1 (Upstate, #05-500), and HP1α (Upstate, #07-346) were used.

**Mass Spectrometry of Proteins**

After trypsin digestion of gel slices peptides were extracted, desalted using StageTips (Rappsilber et al., 2003), and analyzed using a nano-HPLC Agilent 1100 nanoflow system connected online to an linear trap quadropole (LTQ)-Orbitrap mass spectrometer (Thermo Fisher, Bremen). The mass spectrometer was operated in the data-dependent mode to automatically switch between MS and MS2. The instrument was operated with “lock-mass option” as recently described (Olsen et al., 2005). Survey spectra were acquired with 60,000 resolution in the orbitrap, while acquiring up to five tandem mass spectra in the LTQ part of the instrument. Fragment spectra were searched using the MASCOT software program (Matrix Science, UK). MSQuant (http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_plusSign=%2B&_targetURL=http%253A%252F%252Fmsquant.sourceforge.net%252F) was used to verify and quantify the resulting peptide pairs. The median ratio of all ratios between 2 and 0.5 was used to determine the standard deviation from the 1:1 ratio. The protein ratios for all TAFs were more than ten standard deviations away from this median value. Each pull-down experiment was performed at least three times with independent batches of extracts, with similar results.

**Mammalian Cell Culture, Chromatin Immunoprecipitation (ChiP), mRNA Analyses, and Transient Reporter Assays**

Human U2OS osteosarcoma cells were transfected with siRNAs (Dharmacon) according to the manufacturer's instructions. Formaldehyde-crosslinked chromatin samples were prepared as described ([Bernstein et al., 2005], [Denissov et al., 2007] and [Heintzman et al., 2007]). The ChiP samples were analyzed by qPCR, and mRNA levels were determined by quantitative RT-PCR as described (Dreijerink et al., 2006). PCR primers and siRNA sequences are available upon request. For luciferase assays U2OS cells were transfected in duplicates or triplicates (as indicated) by using FuGENE 6 (Roche Applied Science). Firefly reporter luciferase constructs 5XGal4MLP-Luc or TK-Luc (100 ng) were cotransfected with 50 ng TK promoter-driven Renilla luciferase plasmid for differences in transfection efficiency and 250 ng pMT2-HA-mTAF3 constructs. Cell lysates were
prepared 38 hr after transfection, and luciferase activity was determined using the Dual-Luciferase Reporter Assay System (Promega).

To analyze levels of transfected proteins, U2OS cells were plated on 6-well tissue culture plates and transfected with 500 ng Gal4-Ash2L expression plasmid and 500 ng mTAF3 wild-type (WT) or mutant plasmids. Total cell lysates were analyzed by immunoblotting using anti-HA (3F10) or Gal4-DBD (RK5C1, Santa Cruz, CA) antibodies and enhanced chemiluminescence (ECL, Amersham Biosciences).

**H3 Peptide- and Nucleosome-Binding Experiments**

H3 peptides immobilized on streptavidin-conjugated dynabeads were incubated with crude bacterial GST-TAF-PHD induced lysates in binding buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 0.1% NP40, 10 µM ZnCl2, 1 mM DTT, and protease inhibitors) for 2–3 hr at 4°C. Following extensive washing bound proteins were eluted and analyzed by SDS-PAGE and Coomassie brilliant blue R-250 (Biorad).

For nucleosome-binding experiments glutathione-agarose beads were coated with bacterial lysates containing GST or GST-TAF3PHD (WT or mutant). Mononucleosomes were mixed with GST-coated agarose beads and incubated overnight at 4°C in H3-binding buffer. Beads were washed five times with H3-binding buffer. Twenty-five percent was used for extraction of DNA using the QiaQuick PCR purification kit (QIAGEN) and resolved on a 1.5% agarose gel. The remainder was prepared for immunoblot analysis. Antibodies specific for H3, H3K4me1, H3K4me2, H3K4me3, and H4 (ab1791, ab8895, ab7766, ab8580, and ab31827) were obtained from Abcam. Antibodies against H3K36me3 (GTX80647), H3K79me2, and H2A (#07–366 and #07–146) were obtained from GeneTex, Inc. and Upstate Biotechnology, respectively. GST-fusion proteins were detected with a GST antibody conjugated to HRP (Amersham, 27–4588).

**H3 Peptide Affinity Measurements**

The TAF3-PHD(857–924) was expressed as a GST-fusion protein and purified essentially as described (Domínguez et al., 2004). The resulting protein was further purified by anion exchange chromatography and estimated to be > 95% pure by Coomassie staining. For Kd measurements the TAF3-PHD protein was diluted to 100 nM for H3K4me3, H3R2me1/K4me3, H3R2me2s/K4me3, to 200 nM for H3K4me3/K9,14Ac, and to 1 µM for H3R2me2a/K4me3, H3K9,14Ac, H3K4me2, H3K4me1, and unmodified peptides in H3-binding buffer plus 0.001% Tween-20. Tryptophane fluorescence was determined using a Varian Cary Eclipse fluorescence spectrophotometer using 1.2 ml cuvettes with stirring. The excitation wavelength was 280 nm and slit width 2.5 nm. Emission was determined in the range of 335–355 nm with a 10 nm (for 100 nM PHD), 5 nm (for 200 nM PHD), or a 2.5 nm slit width (for 1 µM PHD) with an averaging time of 5 s and 2 nm data interval. Each measurement was repeated at least three times with two different batches of H3 peptide.

**Acknowledgments**

We gratefully acknowledge Drs. Davidson, Tora, Gozani, and Patel for providing plasmids, Dr. Stewart for various yeast strains, Dr. W. Herr for WDR5 antibody, and Dr. Rehmann for help with the tryptophan fluorescence measurements. We are indebted to Dr. Lüscher for communication of unpublished results and providing the Gal4-Ash2L plasmid. We also thank Drs. van Ingen, Wienk, and Boelens for designing the TAF3-PHD mutants. We also thank Dr. Kraus, Dr. Rehmann, and members of the Stunnenberg, Mann, and Timmers labs for stimulating discussions.
This work was supported by grants to the Timmers lab from the European Union (STREP LSHG-CT-2004-502950) and the Netherlands Proteomics Centre. Furthermore, the work in the Mann and Stunnenberg labs is supported by HEROIC, an Integrated Project funded by the European Union under the 6th Framework Programme (LSHG-CT-2005-018883). The research of M.V. is supported by a TALENT grant of the Netherlands Organization for Scientific Research (NWO) and by a fellowship of the Dutch Cancer Society (KWF/NKB).

References


Selective binding of TFIID to histone H3K4me3


Addendum

fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36, J. Biol. Chem. 282 (2007), pp. 2450–2455.
Summary
Coordinated regulation of gene expression in a living cell is dependent on a highly complex and broad network of various functional DNA elements, transcription factors, chromatin modifying complexes and non-coding RNA molecules. Although many recent studies aim to find certain aspects of the gene regulation, the majority of the principles and details in genome functioning remain to be uncovered. In this thesis I describe our novel findings that provide further insight into the molecular regulatory mechanisms of expression of human genes.

In Chapter 1 I present a broad introduction into the transcription field. In parts 1, 2 and 3 I describe concepts and mechanisms in gene regulation as well as recent investigations towards understanding transcription mechanisms. In part 4 of the introduction I describe a state-of-the-art methodology for genome research.

In Chapter 2 we identified a large number of functional genomic loci comprising promoters of a representative set of active human genes. About half of the loci corresponded to the promoters of novel hitherto uncharacterized genes suggesting that a substantial portion of human genes still is not annotated. We performed a comprehensive binding profiling of general transcription factors and analyzed their specific requirement for different classes of genes. We found that certain factors that were previously considered to be specific for one class of genes, are more broadly required for transcription of other classes suggesting novel regulatory links between them. Furthermore, based on the distinct binding patterns of the factors we suggested that there are distinct mechanisms of transcription initiation that are characteristic for certain classes of human promoters.

In Chapter 3 the function of the general factor TFIIA in transcription of human genes was investigated via the analysis of the binding profiles at a large number of different genomic loci. The comparison of the profiles together with 25 different transcription factors revealed a significant correlation between the binding of TFIIA and TAF subunits of TFIID complex suggesting a close functional relationship between these factors. Using a high resolution mapping of genomic binding sites for TFIIA and TBP we found that many active promoters do not display significant TFIIA binding. Additional data analysis suggested that there are large groups of TFIIA-dependent and independent genes that have specific biological functions. These data suggest specific and unique functions for TFIIA in transcription of human genes and support existence (Chapter 2) of distinct transcription initiation mechanisms.

In Chapter 4 I specifically investigate transcription by RNA polymerase I. I measured the binding of RNAP I-specific factors throughout the rDNA gene and found a number of unexpected features implying a unique spatial structure. I concluded that the promoter, enhancer and terminator of the rDNA gene are linked together around the transcription initiation complex SL1. In addition, I found efficient contacts between the SL1 complex and the entire transcription region of the rDNA gene in different phases of the cell cycle. To resolve the topology of an active rDNA gene I propose a unique spatio-dynamic structural organization of the RNAP I transcription in the nucleolus. This model integrates accumulated knowledge on rRNA transcription and provides simple and efficient solutions for many unresolved questions about nucleolar functioning.

Chapter 5 describes our investigation in functioning of E2F transcription repressor complexes during cellular senescence. The function of endogenous E2Fs was blocked by overexpression of mutated form E2F-DB lacking both the C-terminal transactivation and the pRB binding domains. This resulted in activation of different E2F target genes, including cell cycle suppressor p19ARF, and prevented cells to undergo both spontaneous and RASV12-induced senescence and to overcome negative effects of overexpressed p19ARF and p53. The results suggest that RB-mediated repression by endogenous E2F is required for cells to respond to senescence-associated induction of p19ARF and p53.
Samenvatting
Samenvatting

In een levende cel is gecoördineerde regulatie van genexpressie afhankelijk van een zeer complex en uitgebreid netwerk van verschillende functionele DNA elementen, transcriptie factoren, complexen die chromatine modificeren en niet-coderende RNA moleculen. Ofschoon er tegenwoordig veel onderzoek wordt verricht aan gen regulatie is er nog zeer veel onbekend. Met de nieuwe bevindingen beschreven in dit proefschrift willen we bijdragen aan een dieper inzicht in de moleculaire mechanismen van menselijke genexpressie.

Hoofdstuk 1 geeft een breed overzicht over transcriptie. In deel 1, 2 en 3 van dit hoofdstuk worden mechanismen en concepten van genregulatie beschreven, alsmede recentelijke ontdekkingen die bijdragen aan het begrijpen van transcriptie mechanismen. In deel 4 worden hyper moderne technologieën en methodieken voor genoom analyse beschreven.

Hoofdstuk 2 beschrijft de ontdekking van een groot aantal functionele genomische elementen binnen de promotoren van een representatieve groep actieve, menselijke genen. Van ongeveer de helft van deze elementen werd vastgesteld dat het gaat om promotoren van nieuwe, tot op heden onbekende, genen. Dit duidt erop dat een aanzienlijk deel van de menselijke genen tot nu toe niet geannoteerd is. Ook bepaalden we de plaatsen op het genoom waar algemene transcriptie-factoren binden, en analyseerden we verschillende groepen van genen het belang van deze binding. Voor een aantal transcriptie-factoren konden we laten zien dat ze niet speciek zijn voor één bepaalde groep genen, zoals eerder werd aangenomen, maar ook voor andere groepen, hetgeen suggereert dat deze groepen wat betreft regulatie met elkaar verbonden zijn. Gebaseerd op de verschillende bindingsprofielen van de factoren denken we ook dat er verschillende mechanismen van transcriptie initiatie zijn, elk specifieк voor bepaalde groepen van menselijke promotoren.

Hoofdstuk 3 wordt de functie van de algemene factor TFIIA bij transcriptie van menselijke genen bestudeerd, door de binding op een groot aantal verschillende delen van het genoom in kaart te brengen. Door dit bindingsprofiel te vergelijken met 25 andere transcriptiefactoren werd vastgesteld dat er een correlatie is tussen de binding van TFIIA en de TAF subunits van het TFIID complex. Dit wijst erop dat TFIIA en TFIID functioneel erg nauw met elkaar verbonden zijn. Door met hoge resolutie genomische bindingsplaatsen van TFIIA en TBP te bepalen konnen we laten zien dat op veel actieve promotoren geen significante TFIIA binding plaatsvindt. Aanvullende data analyse wijst erop dat er grote groepen TFIIA afhankelijke genen zijn, alsmede grote groepen TFIIA onafhankelijke genen, met specifieke biologische functies. Dit suggereert specifieke en unieke functies voor TFIIA bij de transcriptie van menselijke genen en ondersteunt het bestaan van verschillende transcriptie initiatie mechanismen zoals beschreven in hoofdstuk 2.

De focus van hoofdstuk 4 ligt op de transcriptie door RNA polymerase I. Door binding van RNAP I specifieke factoren in het rDNA gen te bepalen werden een aantal onverwachte bevindingen gedaan, wat duide op een unieke ruimtelijke structuur. Hieruit concludeerden we dat de promoter, enhancer en terminator van het rDNA gen met elkaar in verbinding staan bij het transcriptie initiatie complex SL1. Ook vonden we gedurende diverse fases van de celcyclus efficiënte binding van het SL1 complex met de hele transcriptionele regio van het rDNA gen. Om de topologie van actieve rDNA genen op te lossen wordt een unieke plaats-afhankelijke structurele organisatie van de nucleolus voorgesteld. Dit model bevat de toenemende kennis over rDNA transcriptie en geeft simpele en efficiënte oplossingen voor veel onopgeloste vragen over het functioneren van de nucleolus.

In Hoofdstuk 5 wordt het functioneren van de transcriptionele repressie complexen E2F gedurende verouderen van cellen onderzocht. Het functioneren van het cel-eigen E2F werd geblokkeerd door de gemuteerde vorm E2F-DB tot overexpressie te brengen, welke zowel het C-terminale transactivatie domein als ook pRB bindings domein mist. Dit resulteerde in activatie van een aantal genen waaraan
Samenvatting

E2F gebonden zat, inclusief de cel-cyclus remmer p19\textsuperscript{ARF}. Ook verhinderde het zowel de spontane als RAS\textsuperscript{V12} geïnduceerde veroudering van cellen. Daarnaast hielp het de cellen de negatieve effecten van tot overexpressie gebrachte p19\textsuperscript{ARF} en p53 te compenseren. Deze resultaten wijzen erop dat cellen RB afhankelijke blokkade door cel-eigen E2F nodig hebben om het door ouderdom geïnduceerde p19\textsuperscript{ARF} en p53 aan te kunnen.
Acknowledgements

I would like to express my gratitude to all the people who helped me to prepare this thesis although not all of them are mentioned here. First of all, I would like to say lots of thanks to my supervisor Henk Stunnenberg who invited me to perform research in Nijmegen and supported me for many years. During the long time of my work in his laboratory I learned a lot. It is a specialized knowledge about transcription which I included in the rather long general introduction, and also scientific thinking during resolving of complex tasks and organizational communication. His great enthusiasm and ideas supported an intensive working environment and enabled me to make a number of novel observations and obtain a lot of results. Furthermore, via highly extensive communication with him during preparation of the manuscript (Chapter 2) I had an opportunity to learn really a lot about scientific writing. I cannot give a value to these years which played an important role in my formation as a specialist and improving as a person. I am very grateful to Dimitra Mitsiou from who I could learn methodology, experiments design and planning – all the important things needed for experimenter. Her help was crucial in my work, especially during very difficult first years. It was very good to meet her husband Giannis and another friend from Greece – Greta Braliou. I would like to say a lot of thanks to Catelijne Stortelers and Greta for hospitality and help.

I thank all the other colleagues from the laboratory for their openness and help, and particularly Colin Logie for careful reading and correction of my writings. I want to thank Ron Wehrens whose expertise and willing help allowed us to perform an analysis of our data. Of course, I want to say great thanks to all our collaborators because their help made possible an existence of this thesis.

I appreciate very much Hennie Claassen and the secretary Maria van den Goor for their great help regarding my stay in The Netherlands.

I should say great thanks to my friend Sergey Gololobov who was always very supportive and responsive during all the 4 years until he came back to Novosibirsk.

I must say about my inexpressible gratitude to my family: my wife Oksana and children Ilja, Maria and Anna for so many things that I could write another thesis about it.
Curriculum Vitae

I was born on 11th January 1975 in Korsakov, Russia. In 1991 - 1992 I studied at the Physics-Mathematics School in Novosibirsk, Russia. In 1992 - 1997 I continued my study at the Novosibirsk State University at the Faculty of Natural Sciences. I specialized in molecular genetics during my research work in 1995 - 2000 at the Institute of Cytology and Genetics, Novosibirsk. In October 2000 I started a PhD research program at the Department of Molecular Biology at the University of Nijmegen under supervision of prof. Hendrik Stunnenberg. From 2005 I continued working at this department as a researcher.

List of publications

