UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Mechanisms of Core Promoter Sequence-dependent RNA Polymerase II Transcription

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Muyu Xu

December 2012

Dissertation Committee:

Dr. Ernest Martinez, Chairperson
Dr. Frances M. Sladek
Dr. Frank Sauer

The Dissertation of Muyu Xu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

**Acknowledgements**

I have grown up during my five year PhD study in UCRiverside. I am very thankful to my committee, my friends and my family.

First of all, I would like to give my sincerest thanks to my mentor: Dr. Ernest Martinez. I have learnt many things from Ernest and I will tell two short stories here. I remember I was so impressed by the fact that I can repeat ten-year-old experiments using ten-year-old transcription factors from Ernest when I first joined my laboratory at the beginning of 2008. From then on, I began to record my experiments much more clearly and preserve my factors much more carefully. The other story happened during the revision of my paper. At that time, Ernest was so strict to my data and sometimes I felt frustrated. However, after my paper got accepted by Genes & Development only one day after resubmission, I was so convinced that there is never over attention to details. I will never forget the delicious Spanish seafood from Emma at Christmas party every year.

I also thank Dr. Frances M. Sladek and Dr. Frank Sauer for their constructive suggestions and precious energy during countless rounds of committee meetings. I thank Dr. Bob Roeder, Dr. Sohail Malik and other researchers for their generous reagents and my laboratory members for discussion and technical support.

Last but not least, I would like to give my deepest thanks to my parents. As farmers in China, it was extremely difficult for you to support my study. However, you tried your best to support me from kindergarten to college and provide me the education opportunities. It is you who make me understand that great parents don't need to be knowledgeable and rich. You are my biggest support to fight for my PhD degree in USA.

Part of Chapter 2, Subchapter 2.1 and 2.2 are reprinted from *Genes & Development* 25: 2513-2524. PMID: 22156211. Xu M, Sharma P, Pan S, Malik S, Roeder RG, Martinez E. Core promoter-selective function of HMGA1 and Mediator in Initiator-dependent transcription, copyright (2011), with permission from CSH Press. I want to thank Dr. Ernest Martinez for Figures 2.3-2.5, Dr. Songqin Pan for mass spec analysis in Figure 2.5 and Dr. Priyanka Sharma for Figures 2.2.5 and 2.2.6. This work is supported by grants NIH (CA129325) to R.G.R. and NSF (MCB-0448488; MCB-1021696) to E.M.

Part of Chapter 3 is reprinted from the *Journal of Biological Chemistry* 283: 33808-33815. PMID: 18838386. Wang YL, Faiola F, Xu M, Pan S, Martinez E. Human ATAC is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein, copyright (2008), with permission from ASBMB. I thank Dr. Yuan-liang Wang for Figure 3.8 and Dr. Francesco Faiola for Figure 3.9. This work is supported by grants NIH (CA100464) and NSF (MCB-0448488) to E.M.

ABSTRACT OF THE DISSERTATION

Mechanisms of Core Promoter Sequence-dependent RNA Polymerase II Transcription

by

Muyu Xu

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology
University of California, Riverside, December 2012
Dr. Ernest Martinez, Chairperson

The TATA-box, Initiator (INR), Downstream Promoter Element (DPE), Motif Ten Element (MTE), TFIIB Recognition Element (BRE) and the other core promoter elements contribute to the diverse architecture of core promoters and are paramount for transcriptional activation. Diverse core promoters can communicate with enhancer-bound activators to contribute in the second level gene expression regulation. However, the mechanisms of transcription initiation catalyzed by different core promoters remain unknown and/or controversial. Because TFIID and TFIIB bind most of the core promoter elements, many scientists believe that different core promoters are regulated by the same set of general transcription factors. In contrast, other scientists including us insist that additional core promoter sequence-specific transcription factors besides the general transcription machinery are required to regulate transcription from different core promoters. Several lines of evidence support this hypothesis: a TAFs and

Initiator dependent Cofactor 1 (TIC1) fraction requires for the TATA/INR synergy; a TIC2 fraction supports TATA-less core promoter directed transcription; CK2, PC4 and Mediator facilitate the Sp1-activated INR/DPE transcription in mammalian system and NC2 mediates the INR/DPE transcription in *Drosophila* system. Here, we further purify the TIC1 fraction and identified HMGA1 and Mediator as the effective components that support TATA/INR synergy *in vitro*. In addition, we also verify the TATA/INR specific role of HMGA1 in mammalian cells. Furthermore, we demonstrate HMGA1 interacts with TFIID and Mediator, and the acidic COOH-tail of HMGA1 is required but not sufficient for HMGA1 to interact with TFIID and Mediator. Accordingly, HMGA1 COOH-tail is also required to support the maximal transcriptional synergy between the TATA-box and the INR. Finally, analysis of activated transcription by Gal4-fusion activators and the β-Actin gene (*ACTB*) promoter upstream activating sequences further demonstrates that preferential communication between activators and core promoters contributes to the gene expression regulation. Amazingly, a strict TATA-specificificity by the *ACTB* upstream activating sequences is revealed for the first time.

**Table of Contents**

## List of Figures

## List of Tables

# Chapter 1

## Introduction

**RNA polymerase II transcription and General Transcription Factors**

Following the discovery of the double helix DNA and single strand mRNA, the Central Dogma states that RNA is transcribed from DNA. The process of RNA synthesis is transcription. Transcription involves multiple events in eukaryotic cells, including activator binding, chromatin de-condensation and remodeling, co-activator bridging, general machinery recruitment to core promoter, transcription initiation, elongation and termination. RNA polymerase, the first identified factor of the general transcription machinery, carries out transcription. RNA polymerase activity was first observed in mouse liver extracts in 1959 (Weiss and Gladstone 1959). Soon, RNA polymerase activity was also discovered in *E. coli* (Stevens 1960). It was believed at that time that only one unique polymerase existed in all organisms. To date, four different forms of RNA polymerase have been discovered (Thomas and Chiang 2006). Dr. Roeder and Dr. Rutter were the first to identify different forms of RNA polymerase activities using sea urchin embryo nuclear extract fractionated by DEAE-sephadex chromatography. RNA Polymerase I eluted at low salt concentration fraction, followed by RNA polymerase II (RNA Pol II) and III (Roeder and Rutter 1969). After the discovery and purifications of RNA Pol II, fractions from soluble nuclear extract were found to be required for accurate RNA Pol II initiation (Weil et al., 1979). Chromatography with Phosphocellulose-11 (P11) of cell extract demonstrated that Fraction A (0.1 M [KCL] flow-through), C (0.6 M [KCL]) and D (1 M [KCL]) were necessary for accurate transcription (Matsui et al., 1980). The effective protein factors in A and D were named

TFIIA and TFIID, while fraction C contained TFIIB, TFIIE, TFIIF, and TFIIH (Sawadogo and Roeder 1985; Reinberg and Roeder 1987; Flores et al., 1989, 1992). RNA Pol II together with General Transcription Factors (GTFs) including TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH support basal level transcription *in vitro*.

**Core promoter elements**

Core promoter DNA sequences can extend 40 base pairs (bps) upstream and downstream of the transcription start site (the first nucleotide of mRNA transcript). When first discovered, core promoters were not considered to be diverse. However, subsequent studies revealed a structural and functional diversity of core promoters (Smale 2001; Smale and Kadonaga 2003). Recently, the contribution of core promoters in combinatorial gene regulation are being appreciated more and more.

The TATA-box (also named the Goldberg-Hogness box) was the first core promoter element identified among protein coding genes. It was discovered through the comparisons of the 5'-flanking sequences upstream of start codon of a number of *Drosophila*, mammalian, and viral genes (Goldberg, PhD thesis 1979; Breathnach and Chambon 1981). Originally, TATA-box was found between 25-30 bp upstream of transcription start site in virtually every class II genes, leading to the hypothesis that all core promoters may contain a TATA-box (Breathnach and Chambon 1981; Smale and Kadonaga 2003). In *Saccharomyces cerevisiae*,

the TATA-box is located at 40-120 bp upstream of the transcription start site and is also important for transcription initiation (Li et al., 1994). Mutagenesis and DNA binding experiments revealed that the consensus sequences of the TATA-box is TATA(A/T)A(G/A) (Smale and Kadonaga 2003; Thomas and Chiang 2006). As transcription start sites from more and more genes were characterized, the prevalence of the TATA-box disappeared. A recent study from our laboratory and Dr. Sladek's laboratory showed that only 10-24% of 10,271 analyzed human genes contain a TATA-box (Yang et al., 2007). A similar ratio was published in a previous statistics study (Jin et al., 2006). Early studies indicated that a fraction partially purified from nuclear extracts (TFIID fraction) was required to support the specific transcription initiation from TATA-box containing promoters (Parker and Topol 1984; Sawadogo and Roeder 1985). The TFIID fraction appeared heterogeneous, and contained components that were later identified as the TATA Binding Protein (TBP) and many TBP-Associated Factors (TAFs) that stablely associate with TBP to form the TFIID complex. TBP was originally purified as a single polypeptide from *Saccharomyces cerevisiae* and proved to bind specifically to the TATA-box by DNase I footprinting (Buratowski et al., 1988; Burley and Roeder 1996). Crystal structure of the TBP-TATA-box complex revealed that TBP binds to the minor groove of the TATA-box and bends the TATA-box (Kim et al., 1993). The TAF components of the TFIID complex were later purified by affinity binding to TBP and sequenced, cloned and analyzed. TAFs have multiple functions, such as core promoter binding/selectivity, coactivator function, protein kinase, histone acetyltransferase, and H1 ubiquitin-conjugating activities (Burley and Roeder 1996; Martinez 2002).

The early comparisons of the 5'-sequences upstream of start codon not only discovered the TATA-box, but also identified a Pyrimidine-rich element surrounding the +1(A) start site, which was speculated to be a functional core promoter element (Breathnach and Chambon 1981). The clear definition of Initiator (INR) was established by Dr. Smale through the study of the mouse lymphocyte-specific Terminal deoxynucleotidyl Transferase (TdT) promoter. Interestingly, INR can direct accurate transcription without a TATA-box and the strength of transcription directed by INR was similar to the TATA-box directed transcription. In addition, INR strongly synergized with the TATA-box when located 25-30 bp downstream of the TATA-box (Smale and Baltimore 1989; O'Shea-Greenfield and Smale 1992). The activator Sp1 can stimulate both TATA-box and INR core promoter elements (Smale et al., 1990). By mutagenesis and statistics analyses, the INR consensus DNA sequence was defined as YYANWYY(Y=A/T, N=A/T/C/G) in mammalian and TCA(G/T)TY in *Drosophila* (Breathnach and Chambon 1981; Smale et al., 1990; Yang et al., 2007). INR is the most prevalent core promoter element among different organisms and about half of human genes contain the INR (Jin et al., 2006; Yang et al., 2007). TAFs within TFIID were required to support transcription from INR-containing TATA-less promoters (Martinez et al., 1994). Interestingly, human TAF1 and *Drosophila* TAF2 together preferentially bound to INR-like sequences (Chalkley and Verrijzer 1999). Although the synergistic binding of TFIID to TATA-box and INR contributes to their transcriptional synergy (Emami et al., 1997), a TIC1 fraction purified from nuclear extract was also required to support the synergy (Martinez et al., 1998). The effective components in TIC1 fraction have not been purified to homogeneity.

Following the discovery of TATA-box and INR, more and more core promoter elements were identified. DPE was identified by alignment of INR containing TATA-less *Drosophila* core promoters. In contrast to the TATA-box, DPE locates at 30 bp downstream of the transcription start site (+1) and can not function without an INR (Burke and Kadonaga 1996). The core promoters of about half of *Drosophila* genes contain a DPE. The consensus DPE DNA sequences are (A/G)G(A/T)(C/T)(G/A/C) (Smale and Kadonaga 2003). TAF6 and TAF9 bind the DPE by photo-crosslinking studies (Burke and Kadonaga 1997). Surprisingly, NC2, which represses TATA-containing promoters, activates DPE-containing promoters in crude *Drosophila* embryonic extracts and *in vivo* (Willy et al., 2000; Hsu et al., 2008).

Upstream and downstream TFIIB Recognition Element ($BRE^u$ and $BRE^d$) are bound by TFIIB, thus the name. TBP-TFIIB-DNA co-crystal structure revealed that TFIIB binds to the major groove upstream and the minor groove downstream of the TATA-box (Nicolov et al., 1995; Smale and Kadonaga 2003).

The MTE was first identified by computer analyses of core promoter sequences and further experimentally characterized (Ohler et al., 2002; Lim et al., 2004). Interestingly, the binding factor for MTE also appeared to be TAF6 and TAF9 (Theisen et al., 2010). The DCE is bound by TAF1 and was identified in human β-globin and Adenovirus major late promoters and substitutes DPE function in certain promoters (Lewis et al., 2000).

Hepatitics B Virus X gene core promoter elment 1 and 2 (XCPE1/2) were identified by the study of human Hepatitics B Virus X gene. XCPE1/2 are present in about 1% of human genes, preferentially in TATA-less promoters. Surprisingly, TBP together with TFIIB and Meidator are sufficient to support XCPE1/2 transcription *in vitro* without TAFs (Tokusumi et al., 2007; Anish et al., 2009).

Lately, the TCT motif was identified as a pyrimidine-rich core promoter motif to present in virtually all ribosomal protein genes. TCT motif can't bind to TFIID, which makes it different from INR. A single nucleotide mutation from "TCT" to "TCA" changes TCT into a functional INR (Parry et al., 2010). Besides all core promoter elements described above, most of which are bound by TFIID and TFIIB (summarized in Fig. 1.1, Fig 1.2 and Table 1.1), it is highly possible that other distinct unknown core promoter elements exist. Identification of all core promoter elements is an important step to uncover the mechanisms of transcription.

**Mediator complex**

Yeast Mediator was first purified as a novel fraction to mediate activated transcription, thus the name (Kelleher et al., 1990). Human Mediator was first purified from Hela cells as the Thyroid Receptor-associated protein complex (TRAP). TRAP was able to mediate TRα-activated transcription *in vitro* (Fondell et al., 1996). At the same period, Mediator was identified by different laboratories, such as SRB/MED-containing cofactor complex (SMCC;

Ito et al., 1999) and vitamin D receptor interacting protein complex (DRIP; Rachez et al., 1998).

Mediator is recruited to regulatory DNA sequences by direct protein-protein interactions with a variety of activators. Activators further induce conformational changes in Mediator that can affect its functions (Malik and Roeder 2010; Meyer et al., 2010; Taatjes 2010). Mediator also interacts physically with RNA Pol II and several GTFs to facilitate their assembly at the core promoter (Kornberg 2005; Malik and Roeder 2010). Accordingly, Mediator associates with both enhancers and core promoters in mammalian cells and has been shown to interact with cohesin to form a complex that bridges enhancers to core promoters via DNA looping (Heintzman et al. 2009; Kagey et al. 2010). Besides facilitating activator-dependent recruitment of the general transcription machinery, Mediator also activates post-recruitment steps in transcription and stimulates phosphorylation of the COOH-terminal repeat domain (CTD) of RNA Pol II (Kornberg 2005; Malik and Roeder 2010). These previous observations suggest that Mediator contributes to differential gene regulation by integrating signals from enhancers and gene-specific activators and may control the activity of the general transcription machinery at the core promoter of most genes. Consistent with this, Mediator is required for optimal activator-independent (i.e., basal) transcription from most core promoters in either yeast or metazoan cell-free transcription extracts *in vitro* (Baek et al., 2002; Reeves and Hahn 2003; Takagi and Kornberg 2006). Intriguingly, however, the stimulatory effect of Mediator on basal transcription is much less apparent in purified systems reconstituted with

non-limiting concentrations of the general transcription machinery (Takagi and Kornberg 2006). This suggests that additional factors may be required for efficient Mediator-dependent stimulation of the general transcription machinery (Malik and Roeder 2010). In addition, Mediator subunit Med26 $NH_2$-terminus domain was shown to interact with TFIID and p-TEFb elongation complex to serve as a molecular switch to control the transcription from initiation to elongation (Takahashi et al., 2011).

**HMGA1**

Histones are the major players in chromatin packing and High Mobility Group (HMG) protein families are the second largest group of low molecular weight architectural proteins in chromatin. HMG proteins play important roles in DNA packing, DNA damage repair, chromatin remodeling, transcriptional regulation and cancer (Reeves and Beckerbauer 2001; Zhang and Wang 2010). HMG protein families can be classified into three groups: HMGA, HMGB and HMGN. HMGA proteins contain three AT-hook AT-rich DNA minor groove binding domains and a highly acidic COOH-tail (Fig. 1.3); HMGB proteins contain two “HMG” box that binds to structural DNA with no sequences specificity and a highly acidic COOH-tail as HMGA; HMGN proteins bind to nucleosomes (Zhang and Wang 2008). This study focuses on the function of HMGA1 (isoforms a and b) in transcription.

HMGA1a and 1b, formerly known as HMG-I and HMG-Y, are differential splicing products from the same gene. Compared with HMGA1a, HMGA1b has a 11 amino acids deletion between the first and the second AT-hooks (Reeves and Beckerbauer 2001; Fig. 1.3). HMGA1 proteins are flexible by themselves, however, once bound to DNA and/or protein, HMGA1 can adapt different structures to function as molecular glue which can stabilize DNA-protein and protein-protein interactions (Fusco and Fedele 2007). The structure of HMGA1 with INF-β promoter DNA has been solved by NMR studies, which showed HMGA1 bound to the minor groove of AT-rich DNA through the core "RGR" motif together with the proximal flanking amino acids (Huth et al., 1997). The most well-known transcriptional function of HMGA1 is derived from the studies of human IFN-beta promoter. HMGA1 is required for NFκB (p65/p50), c-Jun and the other IFN-related proteins to form a stable complex to stimulate virus-induced IFN-beta transcription (Yie et al., 1999). In addition, studies from Fusco's group demonstrated that HMGA1 can promote tumor formation *in vivo* (Fusco and Fedele 2007). DNA microarray analysis was performed with *Hmga1* gene knockout mouse embryonic stem cells to identify HMGA1-dependent genes (Martinez Hoyos et al., 2004).

Interestingly, HMGA1 can interact with GTFs, such as TFIIF and TAF3 (Sgarra et al., 2008; Malini et al., 2011). The interaction of HMGA1 with GTFs indicated that HMGA1 may regulate transcription at core promoters. Evolutionary, the AT-hooks and the acidic COOH-tail of HMGA1 are highly conserved from *Danio rerio* to *Homo sapiens*. Although HMGA1 proteins have not been identified for Fungis, *C.elegans* and sea urchins, HMGA1-like proteins can be predicted

from their genomes (Fig 1.4 and Table 1.2). The evolutionary conservation of HMGA1 AT-hooks and COOH-tail indicates that HMGA1 is important for all organisms. Interestingly, AT-hooks are also present as the affiliated DNA binding domains in various transcription factors, such as SNF2 and *Drosophila* TAF1 (Aravind and Landsman 1998), suggesting that AT-hooks might have important function.

**Negative Cofactor 2 and DNA Topoisomerase I**

Negative Cofactor 2 (NC2) is considered to be a general repressor of TATA-box directed transcription. The crystal structure of NC2 demonstrates that NC2 can compete with TFIIB for binding TBP (Kamada et al., 2001). Surprisingly, NC2 activated INR/DPE promoter transcription in a *Drosophila* cell free system and *in vivo* (Willy et al., 2000; Hsu et al., 2008). In the presence of other unidentified protein factors (e.g. TIC1 fraction), NC2 was shown to preferentially repress TATA-only core promoter transcription (Malecova et al., 2007).

DNA Topoisomerase I (Topo I) consists of 765 amino acids with molecular size around 91 kD and is well known for its role in turning ward DNA by transiently introducing nicks, allowing strand passage and re-ligation (Wang 2002; Leppard and Champoux 2005). Topo I can also function as a transcriptional coactivator by interacting directly with activators and components of the general transcription machinery, such as TBP, to stimulate transcription (Merino et al., 1993). The coactivator function of Topo I was in part due to its ability to

enhance TFIID-TFIIA-promoter complex formation. At high concentration, Topo I represses transcription from TATA-containing but not TATA-less promoters (Shykind et al., 1997). Topo I was originally purified from the phosphocellulose column at 0.85 M KCl fraction with the "upstream factor stimulatory activity" (USA fraction) (Thomas and Chiang 2006).

**TAFs and Initiator- dependent Co-factors (TICs)**

TATA-box and INR core promoter elements strongly synergize *in vitro* (Smale and Baltimore 1989; Martinez et al., 1998). TFIID was shown to bind core promoter containing a TATA-box and an INR at a distance of 25-30 bp (TATA/INR) and a TATA-box only (TATA) core promoter in a similar fashion. TFIIA supports the differential binding of TFIID to TATA/INR and TATA core promoters (Martinez et al., 1994; Emami et al., 1997). However, the TATA-box and INR did not synergize (i.e. TATA/INR and TATA core promoters had same transcription strength) in purified transcription systems reconstituted with purified GTFs and RNA Pol II. A TIC-1 fraction partially purified through several chromatography steps (Phosphocellulose 0.85M KCl, DEAE-cellulose 0.12 M KCl and Heparin-sepharose 0.3-0.55 M KCl) was required to support TATA/INR synergy (Martinez et al., 1998). Western blot analysis demonstrated the effective components of TIC1 fraction were different from RNA Pol II, TFIID, TFII I or YY1, which were reported to be responsible for INR function (Seto et al., 1991; Carcamo et al., 1991; Martinez et al., 1998). The effective factors in the TIC1 fraction that support TATA/INR synergy have not been identified.

**Beta-actin Promoter**

Human β-Actin (*ACTB*) promoter sequences contain two proximal activating elements: the CCAAT box and the CCArGG box. Mutagenesis studies demonstrated that both elements are required to support *ACTB* transcription (Danalition et al., 1991). NF-Y binds the CCAAT-box and $p67^{SRF}$ (SRF) binds the CCArGG-box. NF-Y was previously fused to the Gal4 DNA binding domain to analyze its core promoter preferences. Interestingly, the results showed that Gal4-NF-Y (Gal4-NF-A, Gal4-NF-B and Gal4-NF-C trimeric factor) stimulated promoters possessing either TATA-box or INR in a similar manner (Silvio et al., 1999). The core promoter preference of SRF has not been investigated yet. *ACTB* expression was not affected in HMGA1-knockout mouse embryonic stem cells and HMGA1-knockdown pancreatic cancer cells (Martinez Hoyos et al., 2004; Kolb S et al., 2007). These results suggest that *ACTB* transcription may be regulated through HMGA1-independent pathway.

**Contribution of core promoters to combinatorial gene expression**

After the original sequencing of the human genome, 26383 protein coding genes were identified and ontology analysis showed that 6% of human genes are coding for transcription factors (Venter et al., 2001). Since transcription factors are so limited, elaborate strategies might be required to regulate gene expression. Combinatorial binding of activators to the upstream enhancer/promoter sequences is one of the strategies that can largely increase the

patterns of gene expression (Smale 2001). The well known example of virus-induced interferon-beta gene regulation provides evidence to support this strategy: The interferon-beta promoter contains four positive regulatory DNA elements and interferon-beta transcription can be induced by viruses only when there are more than two positive regulatory elements present in the promoter (Thanos and Maniatis 1992). This suggested that the combinational binding of at least two activators is required to regulate interferon-beta transcription. The other strategy of transcription regulation is selective communication of enhancers/proximal promoters with the core promoter elements. The early example of the existence of two functional distinct TATA-boxes was first discovered in the yeast *his3* gene promoter. The *his3* promoter contains both a nonconsensus TATA-box and a consensus TATA-box, but only the consensus TATA-box is inducible to the upstream activating sequences (Struhl 1986). Furthermore, myoglobin gene enhancer selectively activated its own consensus TATA, but it could not activate the TATA from the *Simian virus 40* (SV40) early promoter. However, after the SV40 TATA sequences was mutated from "TATTTAT" to "TATAAAA", it responded to the myoglobin enhancer (Wafald et al., 1990).

The study of the murine lymphocyte-specific TdT promoter demonstrated that the TdT upstream promoter sequences function preferentially through the INR element but not the TATA-box (Smale and Baltimore 1989; Garraway et al., 1996). A subsequent study proved that Elf-1, which binds to the upstream sequences of the TdT promoter, might be responsible for the activation of the INR (Ernst et al., 1996). Another important study from Dr.

Kadonaga's laboratory strongly supported the existence of preferential communications between enhancers/proximal promoters and core promoter elements, which increase the diversity of the transcriptional outputs. In this study, a DNA fragment containing loxP-INR-DPE-GFP-FRT-mini white-loxP-TATA-INR-GFP-FRT (loxPs are target sites for CRE recombinase and FRTs are target sites for FLP recombinase) was inserted into *Drosophila* genome by P-element mediated transposon. Eighteen stably inserted *Drosophila* lines were analyzed further by crossing them with *Drosophila* lines expressing either FLP or Cre recombinases, leaving either INR-DPE-GFP or TATA-INR-GFP in the same positions in the genome. Among these 18 lines, 14 lines showed similar GFP expression levels regardless of different core promoters, but three lines expressed more GFP when using INR-DPE as a core promoter and one line expressed more GFP when using TATA-INR as a core promoter (Bluter and Kadonaga, 2001). Because the basal transcription levels from INR-DPE or TATA-INR core promoters are similar in *Drosophila* embryonic extract, the reasonable explanation for the different transcription outputs is that certain enhancers/proximal promoters (4 out of 18) have preferences to function through specific core promoter elements, which increased the diversity of transcriptional output. To date, there are only a limited number of examples for how enhancers/proximal promoters select core promoter elements are achieved. Selective communication between enhancers and specific core promoters could also facilitate the recruitment of enhancer to promoter when it is far away from the transcription start site.

**Significance of my project**

Core promoter elements are diversely present in genes. Selective communication between enhancers and core promoters contributes to the gene regulation. To date, mechanistic studies of different core promoter-directed transcription are very limited. NC2 represses TATA-box containing promoters but activates INR/DPE synergy in *Drosophila in vitro* and in cultured cells (Willy et al., 2000; Hus et al., 2008). In contrast, PC4 and CK2 were shown to support INR/DPE synergy in a purified mammalian *in vitro* transcription system and in cells instead of NC2 (Lewis et al., 2005). Here we demonstrate for the first time the concerted requirement of HMGA1 and Mediator to support TATA and INR synergy in a purified transcription system. Our studies demonstrate a novel role of HMGA1 functions as a core promoter-specific transcription factor both *in vitro* and *in vivo*. Our results also extend the well-known function of HMGA1 in the assembly of enhanceosomes to the activation of the basal transcription. In addition, we also demonstrate that Mediator can preferentially support TATA/INR synergy in the presence of HMGA1 and TFIID/TAFs, which is a novel core promoter-specific function of Mediator besides its previously described general coactivator function. Mechanistically, we show that HMGA1 interacts with both TFIID and Mediator, and the acidic COOH-tail of HMGA1 is required but not sufficient for the interaction. Accordingly, we also show that the HMGA1 COOH-tail is required to support the maximal TATA/INR synergy *in vitro*. Finally, by analyses of core promoter preference of activators, we identify a novel TATA-specificity of the *ACTB* upstream activating sequences,

providing the second example of a strict core promoter element preference of a proximal promoter.

**Figure legends**

**Figure 1.1** Core promoter elements and their relative positions to the transcription start site. Relative positions of core promoter elements to transcription start site (TSS). BRE and TATA-box locate upstream of TSS; INR, TCT and XCPE1/2 overlap with TSS; MTE, DPE and DCE locate downstream of TSS (This figure is adapted from Dr. James Kadonaga's laboratory research website, UCSD. http://biology.ucsd.edu/faculty/kadonaga.html).

**Figure 1.2** TFIID and TFIIB bind to most of the core promoter elements. TBP binds to the TATA-box. TAF 1 and TAF2 together bind to the INR. TAF6 and TAF9 bind to DPE as well as MTE. TAF1 crosslinked with DCE. TFIIB binds to $BRE^{u}$ and $BRE^{d}$. This figure is adapted from Fig 2 of Thomas and Chiang 2006. TAF6 and TAF9 bind to MTE was demonstrated in recent study from Theisen et al., 2010.

**Figure 1.3** AT-hooks and acidic COOH-tail of HMGA1. AT-hooks are colored light green and COOH-tail is colored yellow. The amino acid sequences for the AT-hooks of HMGA1, AT-hooks of *Drosophila* TAF1 and acidic COOH-tail of HMGA1 are described below. The AT-hook core amino acids "RGR" are highlighted in blue color. The phosphorylation sites of COOH-tail by CK2 are labeled with "*".

**Figure 1.4** The amino acid sequence alignments and the evolution tree of HMGA1 from *Xenopus* to *Homo sapiens.* (**A**) The sequences of HMGA1 from different species are aligned with ClustalW2 program. The highly conserved amino acid sequences are labeled with "*". The AT-hooks and COOH-tail are described under the amino acid sequences. (**B**) The evolution tree on the bottom part demonstrates the relative differences of the amino acid sequences of HMGA1 in different species. The numbers represent the differences of HMGA1 amino acids between the hypothetical HMGA1 ancestor and the species from *Homo sapiens* to *Xenopus* (*Homo sapiens* and *Canis Familiaris* are closest to the common HMGA1 ancestor with difference at 0.00521, following by *Mus musculus* with 0.02089. *Xenopus* HMGA1 has the biggest difference from HMGA1 ancestor sequences with a number of 0.36453). The length of lines represents the relative lengths of evolution time for different species.

**Table 1.1** Core promoter elements identified in eukaryotes to date. The table lists the relative positions, consensus sequences, prevalence, bound factors and the references. N/A (Not Avaiable). TATA-less (prefer to present in TATA-less gene promoters). RP genes (ribosomal protein genes).

**Table 1.2** Predicted HMGA1-like proteins from NCBI protein blast with Reference Protein Database. Human HMGA1b amino acid seqences was used to blast in Reference Protein Database. Only the predicted HMGA1-like proteins from representative species were listed in the table according to their identities to *Homo sapiens* HMGA1b.

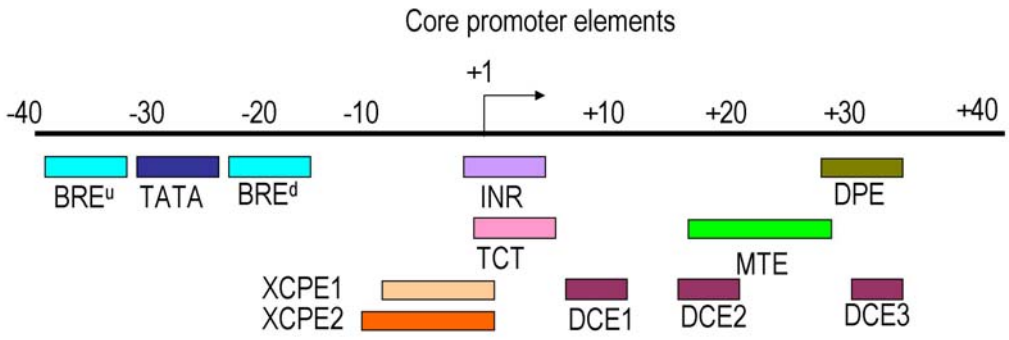



**Figure 1.1** Core promoter elements and their relative positions to the transcription start site (http://biology.ucsd.edu/faculty/kadonaga.html).

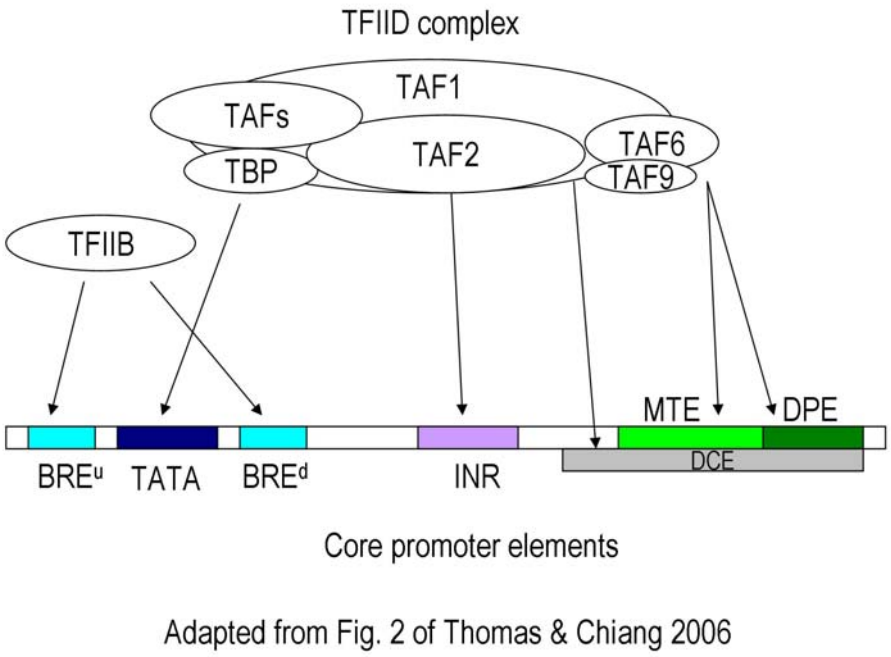



**Fig 1.2** TFIID and TFIIB bind to most of the core promoter elements.

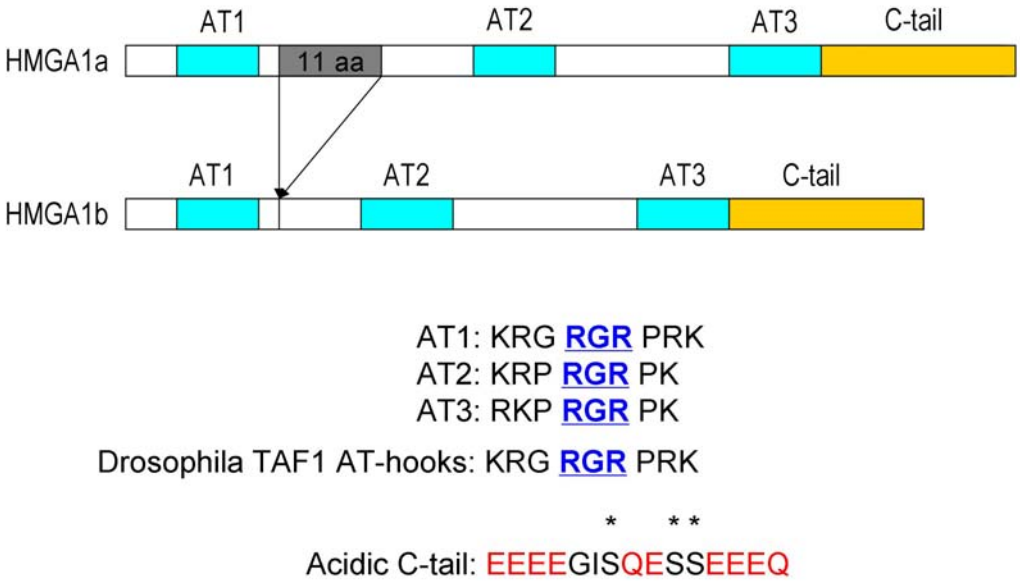



**Figure 1.3** AT-hooks and COOH-tail of HMGA1.

**A.**

```
Homo sapiens          MSE-SSSKSSQPLASKQEKDGTEKRGRGRPRKQPPKEPSEVPTPKRPRGRPKGSKNKGAA 59
Canis familiaris      MSE-SSSKSSQPLASKQEKDGTEKRGRGRPRKQPPKEPSEVPTPKRPRGRPKGSKNKGAA 59
Mus musculus          MSE-SGSKSSQPLASKQEKDGTEKRGRGRPRKQPQKEPSEVPTPKRPRGRPKGSKNKGAA 59
Gallus gallus         MSD-AGAKPSPPLASKGEKDAAEKRGRGRPRKKP-EDPSEAPTPKRPRGRPKGSKNKASS 58
Danio rerio           MSD-SEKQT----VSLKDKDGVEKRGRGRPRKHP-KESSGSPSAKKPRGRPKGSKNKGPS 54
Xenopus               MSSREGARQSSSAEQPASPSQSPKRGRGRPRKPQKEPTAGEPSPKRPRGRPKGSKNKSPS 60
                      **.    :      .  . .   *********   : .:  *:.**********..:
                                                AT1                 AT2
Homo sapiens          KT--RKTTTTPGRKPRGRPKK-LEKE--EEEG-ISQESSEEEQ-- 96
Canis familiaris      KT--RKATTTPGRKPRGRPKK-LEKE--EEEG-ISQESSEEEQ-- 96
Mus musculus          KT--RKVTTAPGRKPRGRPKK-LEKE--EEEG-ISQESSEEEQ-- 96
Gallus gallus         KG--RKSSVTPGMKPRGRPKK-PQQD--EEEVNISQESSEEEQ-- 96
Danio rerio           K---RKSSTS-GSKAKGKPKK-EEKEKPQDSSEDAEEDEDEEQ-- 92
Xenopus               KSAQKKEEASGEKRPRGRPRKWPQQEKKSGREQTAETSSQESEDD 105
                      *   :*  .:   :.:*:*:*  :::  .     :: ..:*.:
                                AT3                COOH-tail
```

**B.**

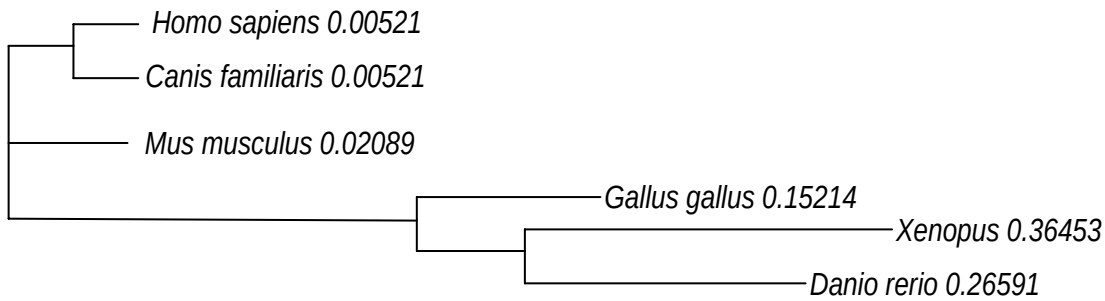



**Figure 1.4** The amino acid sequence alignments and the evolution tree of HMGA1 from *Xenopus* to *Homo sapiens.*

| Elements | Positions | Consensus sequence (5'-3') | Prevalence ( human) | Bound factors | References |
|---|---|---|---|---|---|
| TATA | -31 to -24 | TATA(A/T)A(A/T)(A/G) | 24% | TBP | Goldberg, phD thesis, 1979 |
| INR | -2 to +5 | YYANWYY (Y=C/T, N=A/T/C/G, W=A/T) | 50% | TAF1 and TAF2 | Smale and Baltimore 1989 |
| DPE | +28 to +34 | (A/G)G(A/T)(C/T)CGTG | 12% | TAF6 and TAF9 | Burke and Kadonaga 1996 |
| $BRE^{u}$ | -38 to -32 | (G/C)(G/C)(G/A)CGCC | 16.4% | TFIIB | Largrange et al., 1998 |
| DCE | sub elements<br>+6 to +11<br>+16 to +21<br>+30 to +34 | Core sequences<br>S1: CTTC<br>S2: CTGT<br>S3: AGC | N/A | TAF1 | Lewis et al., 2000 |
| MTE | +18 to +29 | C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C) | 50.6% | TAF6 and TAF9 | Ohler et al., 2002 |
| $BRE^{d}$ | -23 to -17 | (G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G) | 16% | TFIIB | Deng and Roberts 2005 |
| XCPE1 | -8 to +2 | (G/A/T)(G/C)G(T/C)GG(G/A)A(G/C)(A/C) | TATA-less | TAFs-independent | Tokusumi et al., 2007 |
| XCPE2 | -10 to +2 | (A/C/G)C(C/T)C(G/A)TT(G/A)C(C/A)(C/T) | TATA-less | TAFs-independent | Anish et al., 2009 |
| TCT | -1 to +6 | YCTYTYY (Y=C/T) | RP genes | N/A | Parry et al., 2010 |

**Table 1.1** Core promoter elements identified in eukaryotes to date.

| Species | Accession number | Identity | Notes |
|---|---|---|---|
| *Strongylocentrotus purpuratus* | XP_800800.1 | 58% | hypothetical protein |
| *Nectria haematococca* | XP_961052.1 | 45% | hypothetical protein |
| *Caenorhabditis elegans* | NP_503022.1 | 41% | hypothetical protein |
| *Arabidopsis* | XP_002875083.1 | 40% | hypothetical protein |
| *Drosophila ananassae* | XP_001954751.1 | 38% | hypothetical protein |

**Table 1.2** Predicted HMGA1-like proteins from NCBI protein blast with Reference Protein Database.

## Chapter 2

# Biochemical identification of HMGA1 and Mediator as components that support the synergy of TATA-box and INR core promoter elements *in vitro*

Publication:

**Xu M**, Sharma P, Pan S, Malik S, Roeder RG, and Martinez E, 2011. Core promoter-selective function of HMGA1 and Mediator in Initiator-dependent transcription. *Genes & Development* **25**: 2513-2524. [PMID: 22156211]. With the copyright permission from CSH Press, 2011.

**Abstract**

TIC1 fraction partially purified through several chromatography steps from Hela nuclear extract was shown to be required to support TATA/INR synergy by adding back to a purified system containing GTFs and RNA Pol II. Following further chromatographic purifications, HMGA1 and Mediator were identified as the effective factors that support TATA/INR synergy. Consistent with previous results, TAFs were also required together with HMGA1 and Mediator to stimulate TATA/INR synergy. However, neither HMGA1 nor Mediator could independently support maximal INR function in TATA-containing core promoter. In addition, HMGA1 and Mediator counteracted negative regulators of TBP/TATA-directed transcription (e.g. NC2 and Topoisomerase I) in an INR-dependent manner. In sumarry, HMGA1, Mediator and TFIID together stimulated TATA/INR synergy, while HMGA1, NC2 and Topo I together repressed TATA-only core promoter transcription. The concerted functions from all the positive and negative factors contribute to the extraordinary transcriptional differences between TATA/INR and TATA-only core promoters.

## Introduction

Regulation of gene-specific transcription in eukaryotes is controlled by the combinatorial interplay of a variety of regulatory DNA elements located in promoter proximal and distal (e.g., enhancer) regions and core promoter elements located within the transcription initiation region (i.e., the core promoter). Regulatory elements are recognized by cognate sequence-specific DNA-binding regulators (activators or repressors), which in turn recruit a diversity of coregulators (Roeder 2005). Activators often assemble cooperatively at enhancers to form stereo-specific activating complexes (e.g., enhanceosomes). Architectural DNA-binding proteins, such as HMGA1 (formerly HMGI/Y), have been shown to further assist in the formation of specific enhanceosomes (Thanos and Maniatis 1995; Reeves 2003). HMGA family proteins do not have an intrinsic transcription regulatory domain or a strict DNA sequence specificity but bind to the minor groove of AT-rich or structured DNA through "AT-hook" motifs and to numerous sequence-specific regulators. HMGA1 is thought to act as a chaperone to induce or stabilize DNA and/or protein conformations that facilitate cooperative binding of activators to specific enhancers (Reeves and Beckerbauer 2001; Reeves 2003; Panne 2008).

Once recruited by activators to regulatory DNA sequences, different classes of coactivators interplay to modify chromatin structure and/or directly interact with the general transcription machinery to enhance transcription by RNA Pol II (Roeder 2005). The

multi-protein Mediator complex belongs to the latter class of coactivators and has emerged as the prevalent "general coregulator" required for transcription of most protein-coding genes in eukaryotes (Kornberg 2005; Malik and Roeder 2010). Mediator is recruited to regulatory DNA sequences by direct protein-protein interactions with a variety of activators, which further induce structural shifts in Mediator (Malik and Roeder 2010; Meyer et al., 2010; Taatjes 2010). Mediator also interacts physically with RNA Pol II and with several GTFs and facilitates their assembly at the core promoter (Kornberg, 2005; Malik and Roeder 2010). Accordingly, Mediator associates with both enhancers and core promoters in mammalian cells and has been shown to interact with cohesin in a complex that bridges enhancers to core promoters via DNA looping (Heintzman et al., 2009; Kagey et al., 2010). Besides facilitating activator-dependent recruitment of the general transcription machinery, Mediator also activates post-recruitment steps in transcription and stimulates phosphorylation of the C-terminal repeat domain (CTD) of RNA Pol II (Kornberg 2005; Malik and Roeder 2010). These previous observations suggest that Mediator contributes to differential gene regulation by integrating signals emanating mostly from enhancers and gene-specific activators and may control the activity of the general transcription machinery at the core promoter of most genes. Consistent with this, Mediator is required for optimal activator-independent (i.e., basal) transcription from most target core promoters analyzed thus far (Baek et al., 2002; Kim et al., 1994; Mittler et al., 2001; Park et al., 2001; Takagi and Kornberg 2006). Intriguingly, however, the stimulatory effect of Mediator on basal transcription is much less apparent in purified systems reconstituted with non-limiting concentrations of the general transcription

machinery (Mittler et al., 2001; Nair et al., 2005; Takagi and Kornberg 2006). This suggests that additional factors may be required for efficient Mediator-dependent stimulation of the general transcription machinery a (Malik and Roeder 2010).

The core promoter is the ultimate target of activators and Mediator, and is defined as the DNA region where the GTFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) and RNA Pol II assemble to form a functional preinitiation complex (PIC) (Juven-Gershon et al., 2008). Accordingly, this region is generally "nucleosome-free" or marked with unstable nucleosome variants at active or poised genes *in vivo* (Jin et al., 2009). It has long been known that core promoter DNA sequences play an important regulatory role by influencing the transcriptional response of genes to distal activators and enhancers. However, how this is accomplished has remained obscure (Smale 2001; Juven-Gershon et al., 2008). The TATA box is bound by TBP, the INR is recognized by TAF1 in conjunction with TAF2, the DPE photocrosslinks to TAF6 and TAF9, and the DCE photocrosslinks to TAF1 (Juven-Gershon et al., 2008). Accordingly, specific combinations of core promoter elements synergize in basal and activated transcription in crude nuclear extracts and in cultured cells and cooperatively recruit TFIID to the core promoter *in vitro* (O'Shea-Greenfield and Smale 1992; Burke and Kadonaga 1996; Emami et al., 1997; Juven-Gershon et al., 2008). Intriguingly, however, the intrinsic basal activities and synergistic functions of most core promoter elements can not be recapitulated in systems reconstituted with purified GTFs and RNA Pol II, suggesting that cooperative binding of TFIID to core elements can only partly explain their synergy, and that additional

factors may be required for optimal core promoter sequence-dependent RNA Pol II activity. Indeed, INR function at mammalian TATA-less promoters and the strong synergy of TATA-box and INR elements were shown to require distinct "TAF- and INR-dependent Cofactors" (TICs) whose identities have remained elusive (Martinez et al., 1998; Malecová et al., 2007). Similarly, the TFIID-dependent DPE function and its synergy with the INR require the negative cofactor NC2 in *Drosophila* extract (Willy et al., 2000). In contrast, protein kinase CK2 and PC4 were required for DPE-INR synergy in mammalian extract (Lewis et al., 2005). In yeast, the general transcription machinery may also require additional factors for efficient transcription from promoters with weak TATA boxes (Bjornsdottir and Myers 2008). Thus, the factors and mechanisms that regulate the general transcription machinery in a core promoter-specific manner may be diverse and remain poorly defined.

Here, we present the biochemical identification of HMGA1 and Mediator as core promoter-selective cofactors required for the TFIID/TAF-dependent transcription stimulatory function of the INR element and its synergy with TATA-box. HMGA1 functionally cooperates with Mediator and TFIID, and elicits an INR-specific basal transcription stimulatory activity of Mediator, which requires TAFs and counteracts the negative regulation of TATA-dependent transcription by NC2. Our results suggest a possible core promoter-dependent architectural or allosteric regulation of the general transcription machinery by HMGA1.

## Results

### Purification and mass spectrometry identification of HMGA1 as one of the components that support the synergy of TATA-box and INR core promoter elements

We previously partially purified a TFIID/TAF-dependent stimulatory activity (called TIC1) that restored INR function and the synergy of TATA and INR elements in a purified basal transcription system reconstituted with immunoaffinity-purified Flag-tagged TFIID, $Ni^{2+}$-affinity-purified native TFIIA, recombinant 6His-tagged TFIIB, TFIIE, and TFIIF, and purified native TFIIH and RNA Pol II (Martinez et al., 1998). To identify the active components of the crude TIC1 fractions, more extensive chromatographic fractionations were performed and the TIC1 activity in chromatographic fractions was analyzed by complementation of the purified basal transcription system (see Materials and Methods). We followed the ability of TIC1 to stimulate basal transcription selectively from a core promoter containing both TATA and INR consensus elements in a synergistic configuration (TATA/INR) but not from a derivative "TATA-only" core promoter (TATA) that differs only by point mutations that inactivate the INR (Fig. 2.2). The TIC1 activity was purified through seven chromatographic steps (Fig. 2.3; 2.4; 2.5), although fractionation on Q-Sepharose resulted in a significant loss of activity (see below, and Fig 2.3). A protein of ~19 KDa (p19) consistently co-fractionated with the TIC1 activity (Fig. 2.4; 2.5) and was enriched in the final TIC1 "Phenyl" fraction, which also contained two other protein bands, p110 and p9 (Fig. 2.5).

Tandem mass spectrometry analyses (LC-MS/MS) identified these proteins as DNA Topoisomerase I (p110), HMGA1 (p19), SRP14 (also in p19) and SRP9 (p9) (Fig. 2.5). SRP14/9 are abundant cytosolic (and nucleolar) proteins that heterodimerize and function within the signal recognition particle (SRP) in cotranslational targeting of proteins to the endoplasmic reticulum (Koch et al., 2003); hence they were considered contaminants and were not investigated further.

Given their roles as architectural factors and transcription coregulators, Topo I and HMGA1 were further tested for TIC1 activity as purified recombinant proteins (Fig. 2.1; 2.6; 2.9). Purified recombinant Topo I did not have INR-dependent activity by itself and at higher concentrations repressed both TATA/INR and TATA promoters to a similar extent (Fig. 2.9). In contrast, recombinant HMGA1b selectively stimulated the TATA/INR core promoter without affecting the TATA template (Fig. 2.6). Although modest, this INR-dependent stimulatory activity of recombinant HMGA1b was absolutely dependent on TAFs within TFIID (Fig. 2.7). Notably, at higher concentrations HMGA1b repressed transcription selectively from the TATA core promoter (Fig. 2.6); and TAFs were required to antagonize this repressive effect on the TATA/INR promoter (Fig. 2.7). The other major HMGA1 splicing isoform, HMGA1a, which only differs from HMGA1b by an extra 11 amino acids (Fig. 2.5), functioned similarly (Fig. 2.8). For all subsequent experiments we used recombinant HMGA1b at concentrations that activate TATA/INR but do not inhibit TATA-only transcription.

**Identification of Mediator complex as one of the components that support the synergy of TATA-box and INR core promoter elements**

Since significant TIC1 activity was lost during the Q-Sepharose fractionation step, which also separated Mediator from HMGA1 (Fig. 2.3), we considered the possibility that Mediator could be required for efficient HMGA1-mediated stimulation of INR-dependent transcription. As expected from numerous previous reports, a highly purified Mediator preparation that contains the various forms of Mediator, including CDK8-containing and CDK8-lacking complexes (Fig. 2.1-B), did not have core promoter selectivity in the reconstituted system in the absence of HMGA1, and only weakly stimulated basal transcription from both TATA/INR and TATA promoters (Fig. 2.10-A, lanes 2-4). In contrast, a significant (~5 fold) preferential stimulation of TATA/INR was observed in the presence of HMGA1 (Fig. 2.10). Notably, highly purified Mediator from P11-0.85 functioned similarly as Mediator purified directly from Hela nuclear extract (Fig. 2.11). The INR-dependent basal stimulatory activity of HMGA1 and Mediator was not observed in the purified system reconstituted with TBP, but required TFIID/TAFs (Fig. 2.12). Similarly, HMGA1 and Mediator stimulated basal transcription from the natural adenovirus major late core promoter (MLP), which is of the TATA/INR type, and had only a marginal effect on the natural human Hsp70 core promoter, which has an identical consensus TATA-box but no INR (Fig. 2.13). Thus, the INR-dependent activity of HMGA1 and Mediator is observed with different DNA sequences flanking the consensus TATA-box and INR elements (Fig. 2.2). The core promoter-selective activity of

HMGA1 and Mediator was similarly observed with the HMGA1a isoform (Fig. 2.8) and on linear templates (Fig. 2.13), indicating that a superhelical DNA structure is not required. The above results identify HMGA1 and Mediator as key positive components of the TIC1 activity and show that while Mediator (or TFIID) has no significant core promoter-selective transcription activity per se, in the presence of HMGA1, Mediator preferentially stimulates TATA/INR-containing core promoters by potentiating the TFIID/TAF-dependent synergy of TATA and INR elements.

**HMGA1 and Mediator counteracted negative regulators of TBP/TATA-directed transcription in an INR-dependent manner**

The above results suggested that the large (>40-fold) differential activity of TATA/INR versus TATA promoters observed in nuclear extracts is not solely the result of positive effects of HMGA1 and Mediator on TATA/INR but also involves selective repression of TATA-only transcription by a nuclear extract component(s) (see Fig. 2.12, lane 1 versus 5). Since NC2 (also known as DR1/DRAP1) inhibits TATA-dependent transcription and its inhibitory activity is counteracted by the INR in a TAF-dependent manner in nuclear extracts but not in a purified system (Malecová et al., 2007), we tested whether the differential core promoter-selective repressive effect of NC2 is dependent on HMGA1 and Mediator. As expected, purified recombinant NC2 (Fig. 2.14) repressed both TATA/INR and TATA core promoters similarly in the purified system (Fig. 2.15-A, lanes 1 vs 2). However, in the

presence of HMGA1 and Mediator, the TATA/INR promoter became more resistant to NC2 repression, while the TATA promoter was efficiently repressed (Fig. 2.15-A, lanes 3-5). Thus, besides potentiating TATA/INR synergy, HMGA1 and Mediator also antagonize NC2-mediated repression on TATA in an INR-dependent manner leading to an increased differential activity of TATA/INR versus TATA in the presence of NC2 (~9-fold).

Similar to NC2 (Malecová et al., 2007), Topo I (also known as PC3 or Dr2) was shown to repress basal TATA-dependent but not TATA-less INR-dependent transcription (Kretzschmar et al., 1993; Merino et al., 1993). Although at low concentrations Topo I did not have this repressive effect (Fig. 2.9), in the presence of HMGA1, a selective repression of the TATA core promoter was observed both in the absence and presence of Mediator (Fig. 2.15-B). In contrast, the TATA/INR promoter was stimulated in the presence of Topo I, HMGA1, and Mediator leading to a high (~18-fold) differential core promoter activity (Fig. 2.15-B, lane 3).

In summary, all the above results suggest that the differential activity of the general Pol II transcription machinery on TATA and TATA/INR core promoters is the result not only of positive cooperative effects of HMGA1 and Mediator on TFIID/TAF-dependent INR function, but also of antagonistic INR-dependent effects of HMGA1, Mediator, and TAFs on negative regulators of TBP/TATA-directed transcription, such as NC2 and Topo I (Summarized in Fig. 2.16). Interestingly, HMGA1 itself has both positive effects (in concert with TAFs and

Mediator) and negative effects (in concert with Topo I, or by itself at high concentrations) on TATA-dependent transcription, which depend on the presence or absence of a synergistic INR element.

## Discussion

The factors and mechanisms responsible for the strong synergistic stimulation of RNA Pol II-dependent transcription by TATA and INR core promoter elements have remained poorly understood. Previous results indicated that TFIID/TAF-dependent INR function in synergy with the TATA-box not only entails a TFIIA-dependent cooperative recruitment of TFIID to core promoters containing both elements in a synergistic configuration (Emami et al., 1997), but also involves TAF-dependent cofactors that are distinct from GTFs and have remained elusive (Martinez et al., 1998). Here we identified these cofactors as the architectural protein HMGA1 and the Mediator coregulator complex. Significantly, we found that the basal transcription stimulatory function of Mediator, which up to now has been considered ‘‘general’’ or invariant on all core promoters, can be stimulated by HMGA1 and TAFs in an INR-dependent manner (Fig. 2.3; Fig 2.10). Our results thus unveil a ‘‘facultative’’ core promoter-dependent activity of Mediator and HMGA1 and a functional core promoter-selective cooperativity of HMGA1, Mediator, and TFIID/TAFs, as none of these factors alone (or in pairs) can significantly stimulate INR-dependent transcription by the purified transcription machinery. We note, however, that the maximal level of INR-dependent activation observed with crude nuclear extracts has yet to be reached in the purified system, which may suggest the involvement of additional cofactors or post-translational modifications that may be missing in the purified reconstituted

system. For instance, HMGA1 is a substrate for multiple post-translational modifications *in vivo*, which influence its DNA binding and transcription functions (Reeves 2003).

In addition to their TAF- and INR-dependent stimulatory activities, HMGA1 and Mediator also antagonize repression of the basal transcription machinery by NC2 and Topo I in an INR-dependent manner (Fig. 2.15). While not addressed here, this concerted anti-repressive activity of HMGA1 and Mediator could also more broadly antagonize the inhibitory effects of general chromatin components at specific promoters *in vivo*. Indeed, HMGA1 was shown to dynamically compete with histone H1 binding to chromatin in live mammalian cells (Catez et al., 2004). Hence, the combined stimulatory and anti-repressive effects of HMGA1, Mediator, and TAFs may account for the large differential activity of TATA and TATA/INR core promoters observed in more physiological cell-free extracts and in live cells (O'Shea-Greenfield and Smale 1992; Colgan and Manley 1995; Malecova´ et al., 2007). We further analyzed the role of HMGA1 in the synergy of TATA-box and INR *in vivo*.

## Materials and Methods

### Plasmid Constructs

Plasmid templates: the TATA/INR plasmid pG5TdT (-41TATA/+33), the TATA plasmid pG5TdT (-41TATA/Inr-+33), pHsp70(-33/+99)CAT (HSP70), and pML(-45/+65)CAT (MLP) have been described previously (see Materials and Methods at Martinez et al., 1994).

### TIC1 factors biochemical chromatography purification

TIC1 activity was purified from 100 ml HeLa nuclear extract by successive chromatographic steps at 4ºC, as summarized in Fig. 2.3. Phosphocellulose (P11) and DEAE-cellulose (DE52) steps were as previously described (Martinez et al., 1998; Ge et al., 1996). The DE52 flow-through (20 ml, 2 mg protein) in BC100 (20 mM Tris-HCl, pH 7.9 at 4ºC; 20% glycerol; 100 mM KCl; 0.05% IGEPAL CA-630; 10 mM 2-mercaptoethanol; 0.2 mM EDTA; 0.2 mM PMSF) was loaded onto Heparin-Sepharose (0.25 ml). TIC1 activity eluted between 0.3 and 0.55 M KCl concentration (in BC buffer) as determined by *in vitro* transcription after complementation of the basal system reconstituted with purified GTFs and RNA Pol II. TIC1-containing fractions were pooled and dialyzed against BC100 and loaded onto Q-Sepharose resin. TIC1 activity eluted mostly in the flow-through (but significant activity was lost at this step) and was then loaded onto DNA-cellulose (0.1 ml). TIC1 eluted

between 0.15 and 0.25 mM KCl. After dialysis of pooled fractions against BC100, 0.27 ml was bound in batch onto S-Sepharose (50 µl) and eluted between 0.25 and0.38 M KCl. TIC1-containing fractions were polled and 0.17 ml was adjusted to 1M KCl by adding 1 vol. of BC2000 (without IGEPAL-CA630) and 0.34 ml was adsorbed to Pheny-Sepharose (40 µl). TIC1 activity was recovered in the unbound fraction. For transcription analyses BSA was added to the unbound and bound fractions (to 0.3 mg/ml final concentration), which were then dialyzed against BC100. For SDS-PAGE and mass spectrometry analyses, 0.1 ml of TIC1-Phenyl unbound fraction was TCA precipitated and 10 µl were resolved by SDS-PAGE and stained with Coomassie. The three visible protein bands (p110, p19 and p9) were excised, de-stained, in-gel digested with trypsin, analyzed by liquid chromatography electrospray ionization tandem mass spectrometry (LC/ESI/MS/MS) and spectral data were used to search the NCBI non-redundant database, essentially as previously described (Wang et al., 2008). The identified peptides sequences with significant MASCOT ion scores >40 and $P<0.05$ are listed in Fig. 2.5.

**TFIIE/H (S0.3) purification**

The TFIIE/H (S0.3) fraction was purified as described previously for the TFIIE/F/H (Mono S) fraction (Ge et al., 1996) with modification of the last chromatographic step, which was performed on S-Sepharose, and only fractions containing TFIIE and TFIIH (but not TFIIF), as determined by Western blot, were pooled.

**Recombinant proteins purification**

Recombinant mouse 6His-HMGA1b expression vector was general gift from Dr. Fusco. 6His-HMGA1b was expressed in E. coli BL21-CodonPlus(DE3)-RIL (Stratagene), and purified with Talon Metal Affinity Resin (Clontech) from the soluble fraction for 6His-HMGA1b (Fig.2.1), essentially as previously described for 6His-TBP (Ge et al., 1996). Briefly, after about 4 hr of culture in 500 ml LB medium (containing ampicillin) at 30 C° (shaker), the bacteria, at a density of about 0.5 OD600, are induced with 0.5 mM isopropyl β-d-l-thiogalactopyronoside (IPTG, final concentration) and protein expression is allowed for 3 hr at 30 C°. The cells from a 500 ml culture are lysed in 10 ml of lysis buffer [20 mM HEPES, pH 7.9, at 4°; 10% glycerol; 500 mM KC1; 0.1% Nonidet P-40 (NP-40)] by sonication in an ice bath (three times for 5 min each with a Branson 450 sonifier, microtip at output 5 and 20% duty cycle). Cell debris is removed by centrifugation in a Ti45 rotor at 15,000 rpm for 10 min at 4 °. Imidazole is then added to the supernatant to a 5 mM final concentration and the resulting cleared lysate solution (containing soluble HMGA1b) is mixed with 0.4 ml $Ni^{2+}$ Talon resin (pre equilibrated in lysis buffer containing 5 mM imidazole) and incubated for 2-3 hr at 4 °, under constant rotation in a tightly closed small column. After letting the unbound proteins flow through the column, the resin is washed extensively with 40 ml BC500 containing 5 mM imidazole and then with 20 ml BC100 containing 30 mM imidazole. Bound HMGA1b is eluted from the resin with 150 mM imidazole in BC100, HMGA1-containing fractions are pooled (about 0.6 ml), analyzed by

SDS-PAGE, snap frozen in liquid nitrogen, and stored at -70 C°. NC2 complex expression vectors were generous gifts from Dr. Thomas Oelgeschlager and the purification of NC2 complex was described detailedly previously (Wang et al., 2008). In summary, Flag-NC2β/His-NC2α complex were incubated with TALON® Metal Affinity Resin, washed with LB-500, and eluted in LB-500 containing 250 mM imidazole. The resulting eluates were further subjected to incubation with anti-Flag M2 agarose for 15 h at 4°C. After extensive washes with LB-500, complexes were eluted with 0.3 mg/ml FLAG peptide in BC100. Recombinant HMGA1a was generous gift from Dr. Yinsheng Wang from Chemistry department of UCR. Recombinant Topo I was purchased from Genway.

**Flag-Nut2 Mediator Complex purification**

Mediator was immuno-affinity purified directly from nuclear extracts or P11 0.85 M KCl fraction of the Flag-NUT2/MED10 HeLa cell line (Malik and Roeder 2003), without any significant difference in core promoter selective activity in our assays (Fig. 2.11). Briefly, thaw 5 ml Flag-Nut2/MED10 nuclear extract on ice over night, spin at 15000 rpm/min for 15 mins with Ti45 Beckman rotor to remove any insoluble pellet. Transfer supernant to clean tubes and adjust NP-40 to final 0.05%, add 100 μl M2 resin to supernant and incubate over night. Following day, spin gentlely at 5000 rpm for 5 seconds and keep unbound; Wash with 5 ml BC-300 (0.05% NP-40) for 5 times, then transfer resin to 1.5 ml eppendorf tube and wash with BC-100 for 3 times. At last, elute 3 times with 100 μl BC-100 containing final 0.2 mg/ml

Flag peptide and analyze the elution by SDS-PAGE with Silver-staining. Note that these Mediator preparations contain diverse forms of Mediator, including CDK8-containing and CDK8-lacking (e.g., PC2) Mediator complexes (Malik and Roeder 2003).

### *In vitro* transcription assay

Transcription reactions and primer extension were described previously (Martinez et al., 1994; 1998). The 25mer primer (5'-AACAGCTATGACCATGATTACGCCA-3') was used for promoters TATA/INR and TATA primer extension. CAT-30mer (5'-GGTGGTATATCCAGTGATTTTTTTCTCCAT-3') was used for MLP and HSP70 transcripts primer extension. Purification of RNA Pol II and GTFs was described previously (Martinez et al. 1998). The purified basal transcription system consisted of 0.7 μL of RNA Pol II (DE), 1 μL of TFIIA ($Ni^{2+}$-NTA-agarose), 15 ng of recombinant 6His-TFIIB, 1 μL of Flag-tagged TFIID (about 5 ng of f:TBP per microliter), 20 ng of recombinant 6His-TFIIF, and 2 μL of TFIIE/H fraction (see above TFIIE/H(S0.3) purification). Alternatively, 20 ng of recombinant 6His-TFIIE and 0.15 μL of highly purified TFIIH (Q2) were used instead of the TFIIE/H fraction, with similar results. When indicated, 5-10 ng of recombinant 6His-TBP was used instead of Flag-tagged TFIID. The transcription reaction and experimental process were described as previously (Martinez et al., 1994). The X-ray films were scanned with HP precisionscan Pro 3.1 scanner and densitormetry analyses was performed with NIH ImageJ software.

**Figure legends**

**Figure 2.1** Gel analysis of TFIID, Mediator and recombinant HMGA1. **(A)** 4 μL Flag-affinity immunopurified TFIID from the Flag-TBP 3-10 HeLa nuclear extract. The major TAFs and TBP bands can be seen according to the protein marker bands on the left by silver-stained SDS-PAGE gel. **(B)** Flag-affinity immunopurified Mediator from the Flag-NUT2/MED10 HeLa nuclear extract and a mock purification from regular HeLa S3 nuclear extract; silver-stained SDS-PAGE gel. Positions of protein molecular weight markers are shown in KDa. This Mediator preparation is a mixture of several forms of the complex that may exist *in vivo*, including CDK8-containing and CDK8-lacking Mediator complexes (Malik and Roeder 2003). **(C)** Recombinant 6His-HMGA1b purified from the soluble fraction with Coomassie-stained SDS-PAGE gel.

**Figure 2.2** Schematic description of core promoters analyzed in this study. TATA/INR and TATA core promoters derived from the mouse TdT core promoter (-41/+33) with, respectively, an added TATA-box or a TATA-box and a mutated INR (base substituted in the INR are underlined). There are 5 Gal4 binding sites upstream of TATA/INR and TATA. The structures of the Ad2MLP and human HSP70 core promoters are indicated as well. All promoters have an identical 7-mer consensus TATAAAA box sequence (boxed). Consensus INR sequences are boxed.

**Figure 2.3** Purification process of TIC1 factors and separation of Mediator at Q-Sepharose-0.1. **(A)** Purification Scheme for TIC1. TIC1 activity was purified from the TIC1/3 activity fraction (identical to the USA fraction). See Materials and Methods for more detail. **(B)** Loss of TIC1 activity at the Q-Sepharose fractionation step. Reduced TIC1 activity is recovered in the flow-through fraction (Q-Seph. 0.1) as shown by the *in vitro* transcription experiment in the purified TFIID-dependent system (top). No significant activity was present in the bound fractions (not shown). The western blot (bottom) shows that the reduction in TIC1 activity correlates with the removal of Mediator (MED21/SRB7), which bound to Q-Sepharose.

**Figure 2.4** p19 protein band correlates with TIC1 activity. **(A, B)** Silver-stained SDS-PAGE gels (top) and *in vitro* transcription analyses (bottom panels) of fractions derived from the last two purification steps. Fractions with TIC1 activity are marked with a horizontal line. Asterisks mark proteins that do not correlate with TIC1 activity. The p110 band did not perfectly correlate with activity. The p19 band consistently correlated with TIC1 activity. U and W are unbound and wash fractions, respectively.

**Figure 2.5** Mass Spectrometry identification of the proteins in Phenyl-Sepharose fraction to be Topo I, HMGA1, SRP14 and SRP9. The TIC1 Phenyl-fraction was resolved by SDS-PAGE and stained with Coomassie blue. The proteins in the three bands p9, p19 and p110 were identified by tandem mass spectrometry (LC/ESI/MS/MS). The peptides identified

are indicated in boxes, including a peptide for HMGA1a. The 11 amino acid sequence indicated in bold is characteristic of the HMGA1a isoform, and is spliced out in HMGA1b.

**Figure 2.6** Characterization of the INR-specific function of recombinant HMGA1b. **(A)** *In vitro* transcription/primer extension assays with recombinant HMGA1b were performed with supercoiled templates in the purified system containing TFIID. Autoradiograms shown for TATA and TATA/INR are from the same gel exposure time. **(B)** Panel B shows a quantitation (mean ± S.D.) of more that 3 independent transcription experiments normalized to the promoter activities in the absence of HMGA1b.

**Figure 2.7** TAFs are required for HMGA1 to preferentially stimulate TATA/INR promoter. **(A, B)** HMGA1b does not stimulate the TATA/INR template in the purified system reconstituted with TBP instead of TFIID, and represses TATA/INR transcription at high concentrations. Panel B shows a quantitation (mean ± S.D.) of more that 3 independent transcription experiments normalized to the promoter activities in the absence of HMGA1b.

**Figure 2.8** Characterization of the INR-specific function of recombinant HMGA1a. HMGA1a preferentially stimulates TATA/INR transcription in the presence of Mediator. Purified recombinant HMGA1a was titrated from 25 ng to 75 ng in the purified TFIID-dependent system containing Mediator, as indicated.

**Figure 2.9** Topo I does not have core promoter selective function. **(A)** Purified recombinant 6His-Topo I purchased from Genway. **(B)** Recombinant 6His-Topo I was titrated in the purified TFIID-dependent basal system from 15 ng to 90 ng and transcripts from supercoiled TATA/INR and TATA templates were analyzed by primer extension. Transcription signals were quantitated using NIH ImageJ software.

**Figure 2.10** Mediator preferentially stimulates TATA/INR synergy in the presence of HMGA1. **(A)** *In vitro* transcription experiment with supercoiled TATA and TATA/INR templates. Mediator was titrated alone (lanes 2-4) or together with 40 ng recombinant HMGA1b (lanes 5-7) in the purified TFIID system. The histogram shows the relative transcription activities for each template (normalized to lane 1). **(B)** The individual and combined effects of HMGA1b and Mediator on basal transcription in the purified system were quantitated from more than 3 independent experiments and plotted for each supercoiled TATA and TATA/INR promoter template as relative activities (mean ± S.D.) normalized to promoter activities in the absence of HMGA1 and Mediator.

**Figure 2.11** Mediator purified from P11-0.85 fraction stimulates TATA/INR similarly to Mediator purified from nuclear extract. **(A)** Schematic purifications of Mediators. Mediator A was purified one step from nuclear extract by Flag-Nut2/Med10 affinity purification. Mediator B was purified through P11-0.85, followed by Falg-Nut2/Med10 affinity purification. Two steps purified Mediator B was supposed to be a little purer than one step

purified Mediator A. **(B)** *In vitro* transcription comparison of Mediator A and B with supercoiled TATA and TATA/INR templates in the absence of HMGA1 (Lane 1) or in the presence of 50 ng recombinant HMGA1 (Lane 2-5). 2 μl one step purified Mediator A was tested in Lane 2 and 1 μl, 2 μl and 4 μl two steps purified Mediator B was titrated in Lanes 3 to 5.

**Figure 2.12** TAFs are required for HMGA1 and Mediator to preferentially stimulate the TATA/INR promoter. The relative transcription activities and "selectivity ratio" of TATA/INR to TATA are shown. HeLa nuclear extract (NE) was used as a reference.

**Figure 2.13** HMGA1 and Mediator preferentially stimulate TATA/INR-type MLP with different flanking sequences. **(A, B)** Promoter activities were normalized to the activity of MLP in the TFIID system in the absence of HMGA1 and Mediator (A, lanes 1); and only the selectivity ratios (MLP/HSP70) are shown. In panel A, the corresponding relative transcription activities in lanes 1-5 were, respectively, for MLP: 1.00, 2.24, 3.27, 8.24, 1.89; and for HSP70: 0.67, 0.93, 1.02, 1.62, 1.26.

**Figure 2.14** SDS-PAGE analysis of recombinant NC2. The 6His- and Flag- tagged NC2 subunits were expressed from a polycistronic vector in E. coli, purified by metal-afinity and anti-FLAG immunoaffinity, and analyzed by SDS-PAGE. A Coomassie stained SDS-PAGE gel is shown. On the left side, the molecular mass of protein markers is shown in kDa.

**Figure 2.15** HMGA1 and Mediator counteract the repression of NC2 and Topo I in the presence of an INR. **(A, B)** Recombinant NC2 or Topo I were added to the purified TFIID-based system in the presence or absence of HMGA1 and Mediator, as indicated. Basal transcription was analyzed form supercoiled TATA and TATA/INR promoters. Autoradiograms in each of the two panels are from the same gel and film exposure. The relative transcription signals (normalized to lane 1) and the ratio of TATA/INR to TATA signals are shown

**Figure 2.16** Graphical summary. HMGA1, TopoI and Mediator preferentailly stimulate TATA/INR synergy. HMGA1 and Mediator counteract NC2 repression on TATA/INR core promoter. While HMGA1 (high concentration), Topo I and NC2 selectively repress TATA-only core promoter.

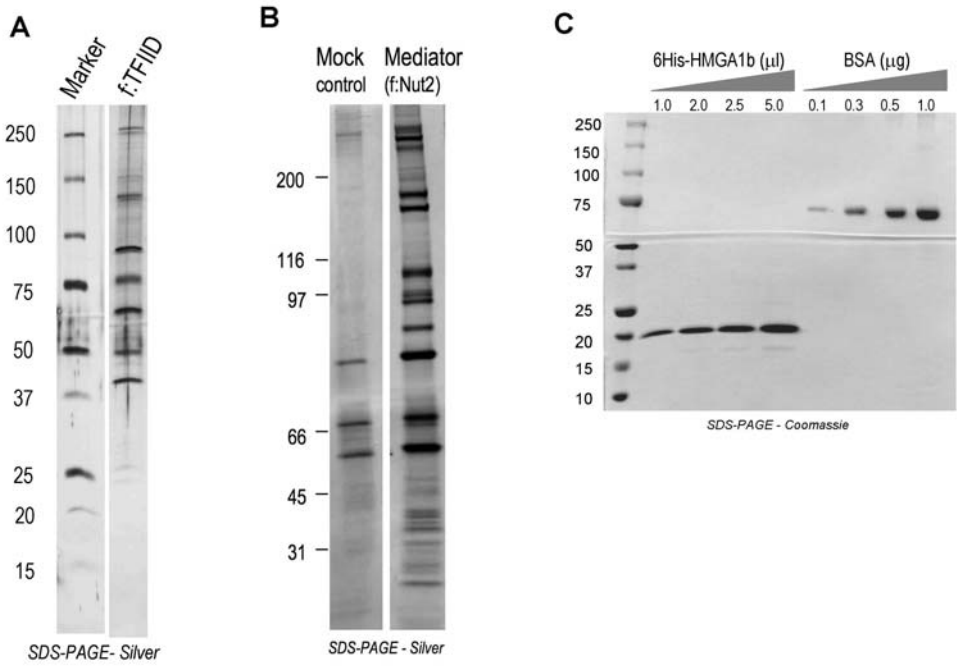



**Figure 2.1** Gel analysis of TFIID, Mediator and recombinant HMGA1.

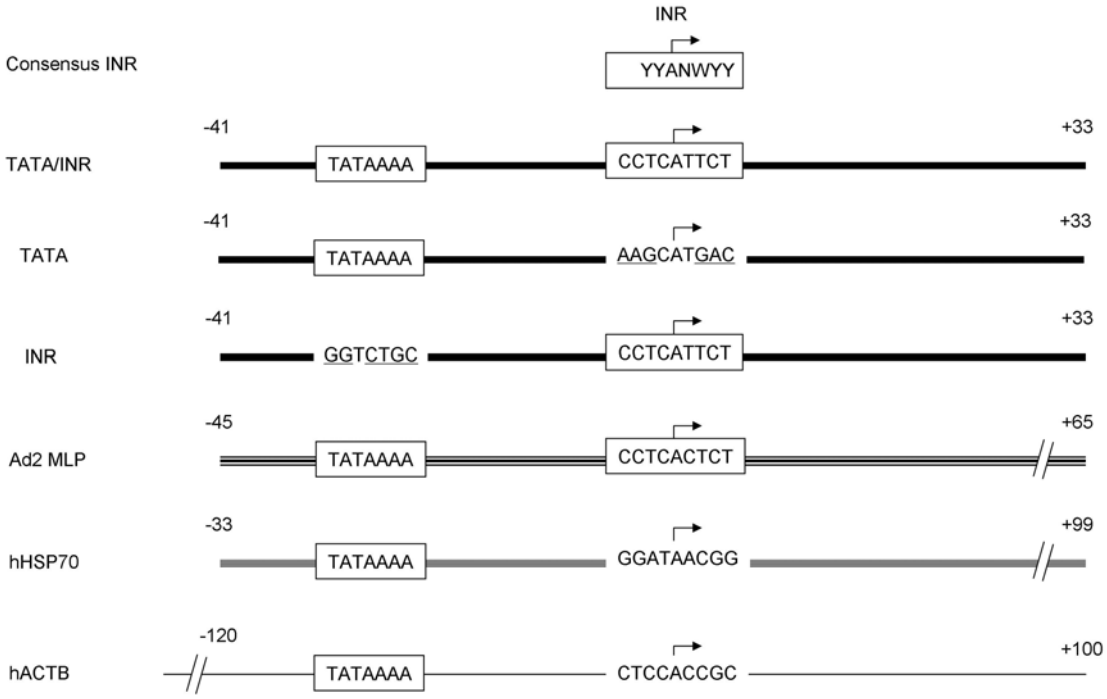



**Figure 2.2** Schematic description of core promoters analyzed in this study.

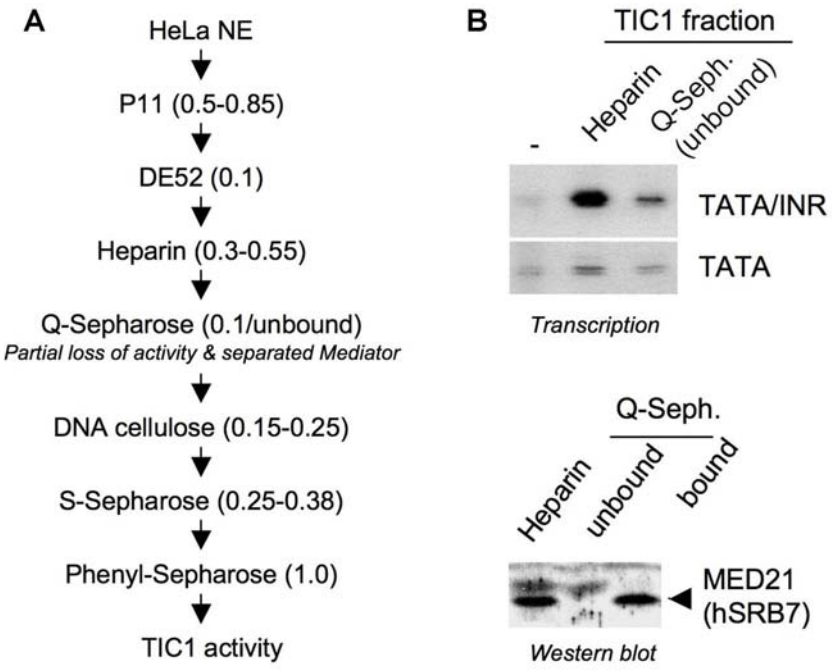



**Figure 2.3** Purification process of TIC1 factors and separation of Mediator at Q-Sepharose-0.1.

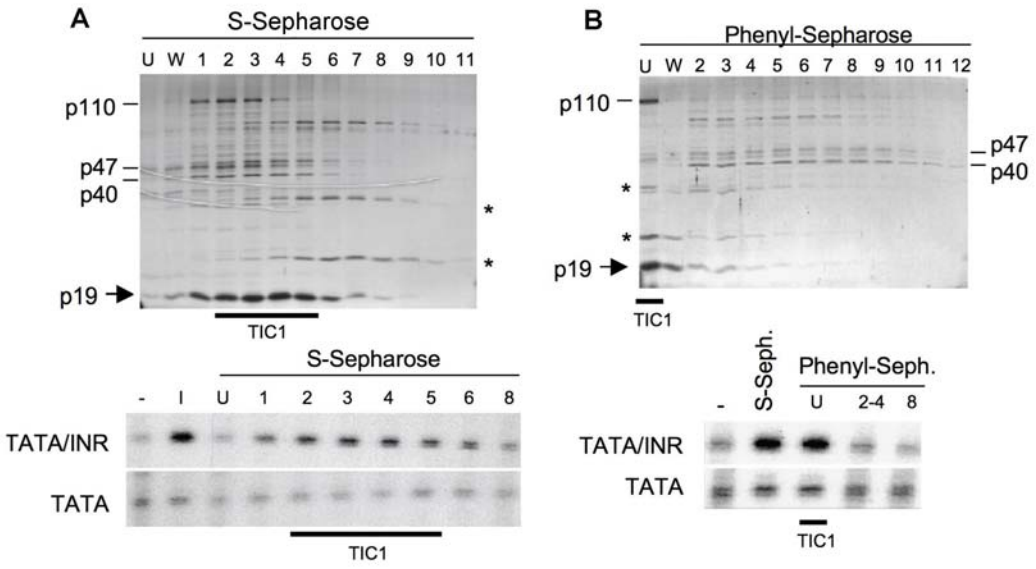



**Figure 2.4** p19 protein band correlates with TIC1 activity to stimulate TATA/INR synergy.

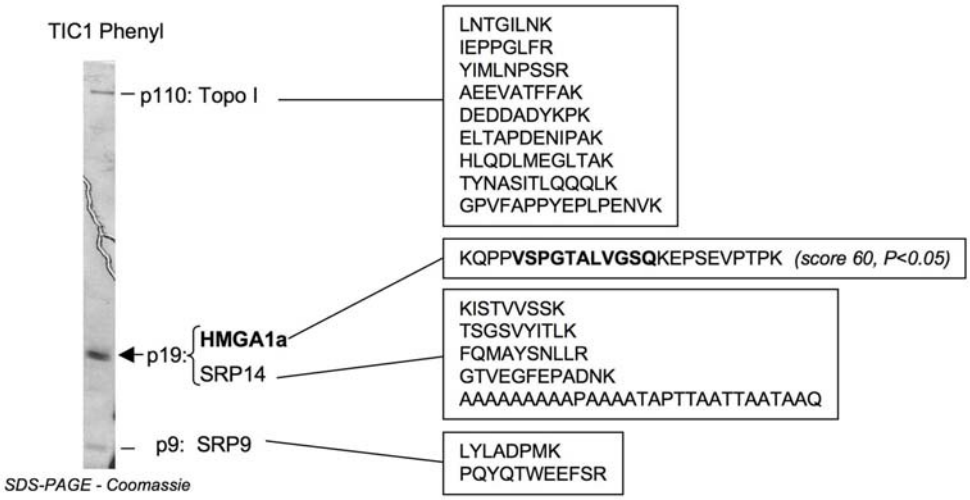



**Figure 2.5** Mass Spectrometry identification of the proteins in Phenyl-Sepharose fraction to be Topo I, HMGA1, SRP14 and SRP9.

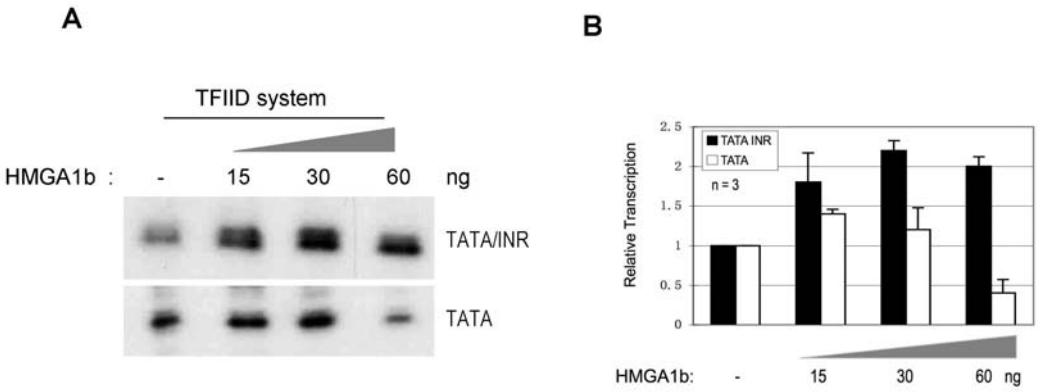



**Figure 2.6** Characterization of the INR-specific function of recombinant HMGA1b.

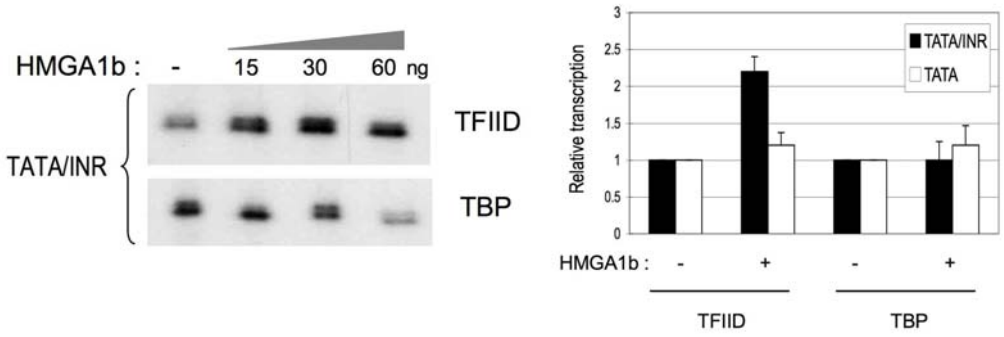



**Figure 2.7** TAFs are required for HMGA1 to preferentially stimulate TATA/INR promoter.

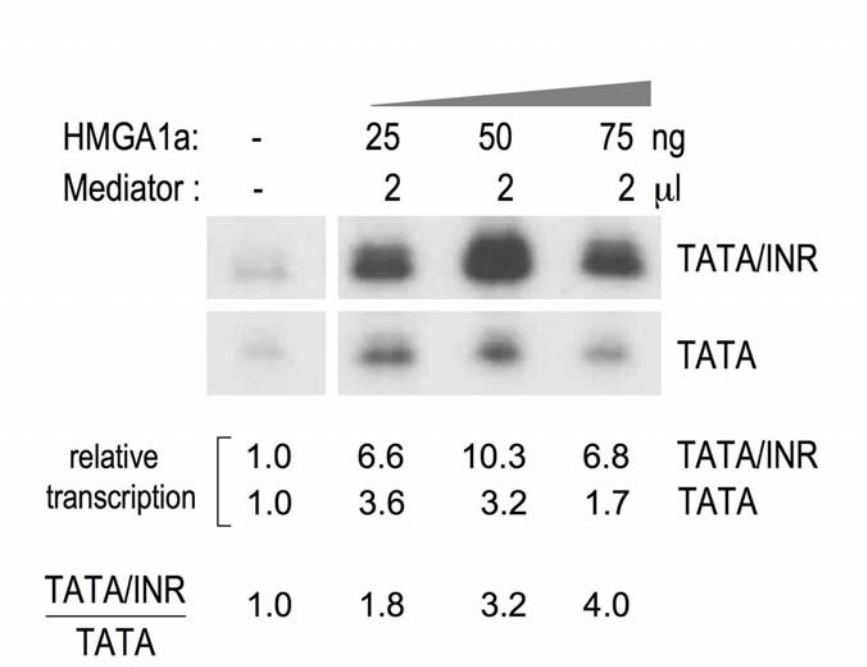



**Figure 2.8** Characterization of the INR-specific function of recombinant HMGA1a.

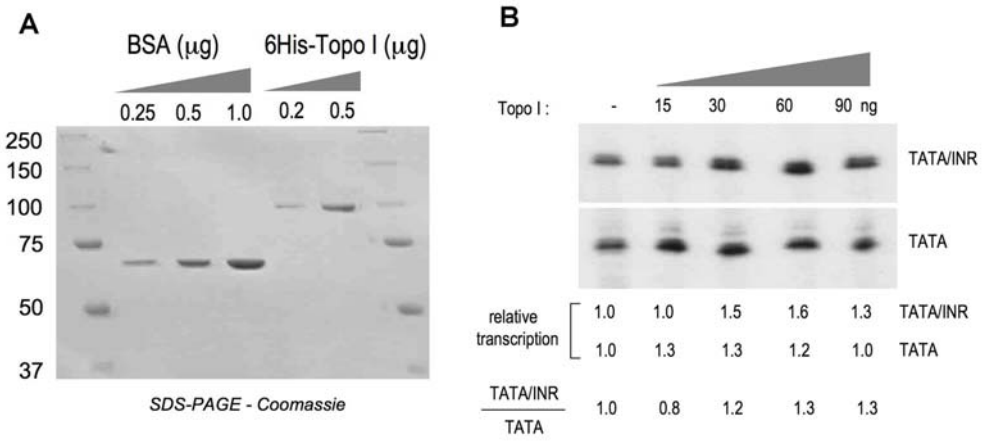



**Figure 2.9** Topo I does not have core promoter selective function.

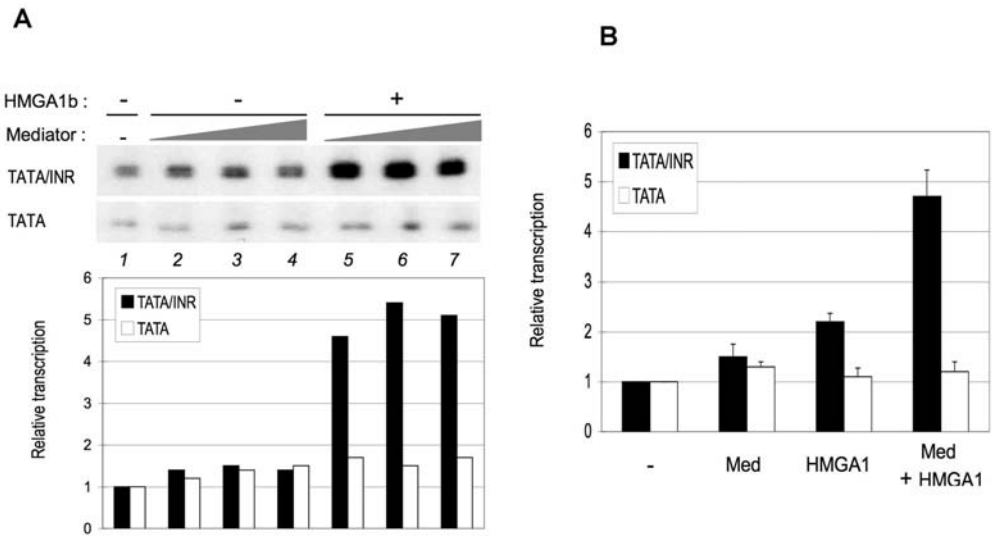



**Figure 2.10** Mediator preferentially stimulates TATA/INR synergy in the presence of HMGA1.

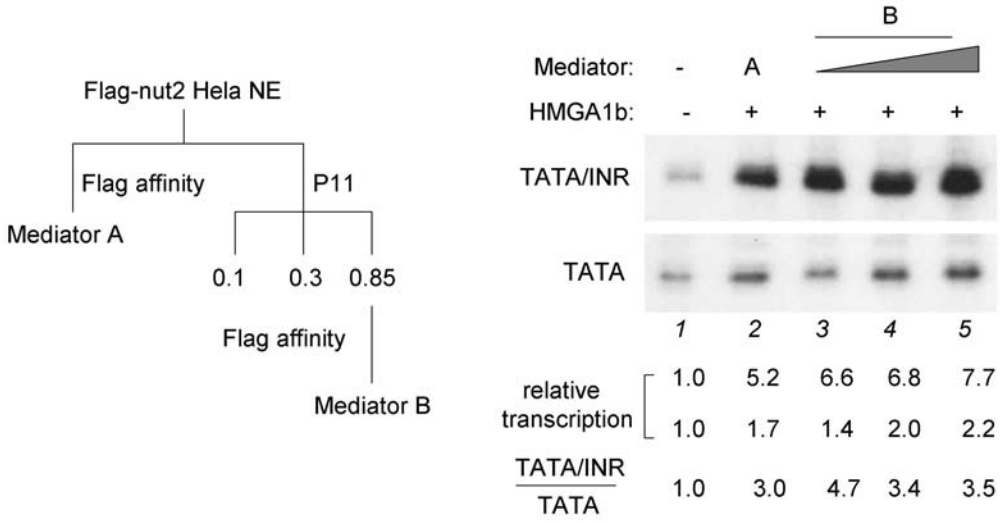



**Figure 2.11** Mediator purified from P11-0.85 fraction stimulates TATA/INR similarly to Mediator purified from nuclear extract.

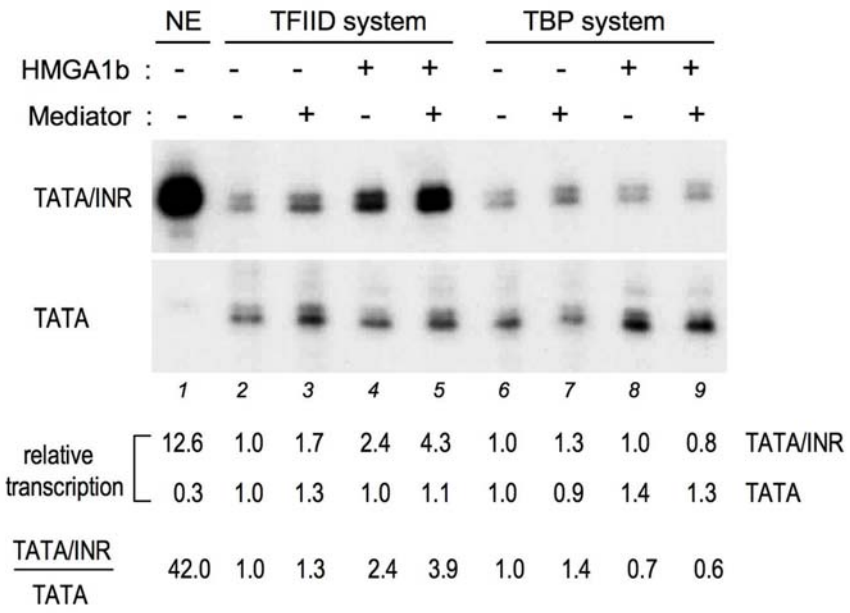



**Figure 2.12** TAFs are required for HMGA1 and Mediator to preferentially stimulate the TATA/INR promoter.

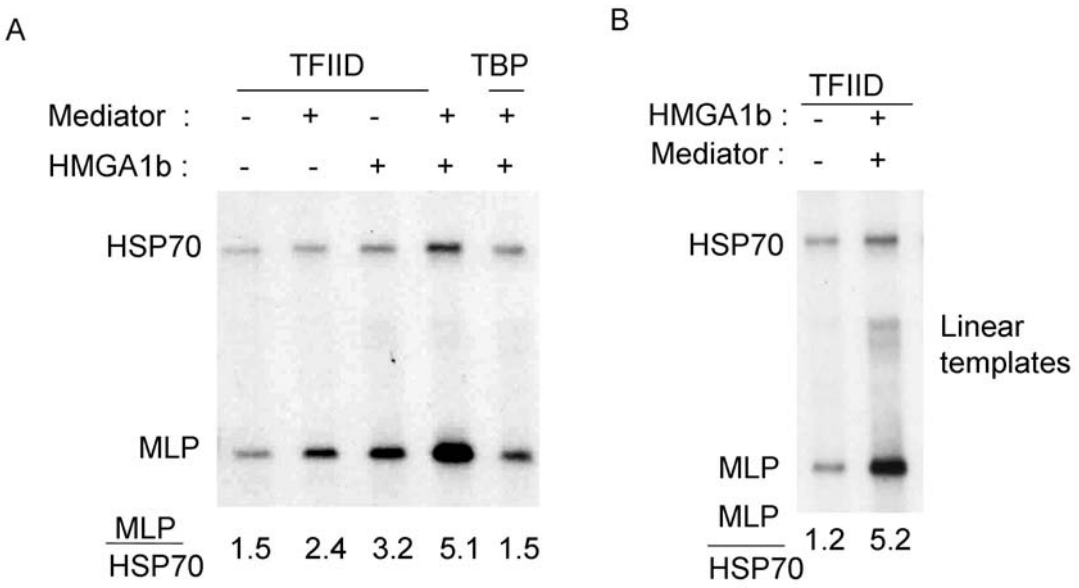



**Figure 2.13** HMGA1 and Mediator preferentially stimulate TATA/INR-type MLP with different flanking sequences.

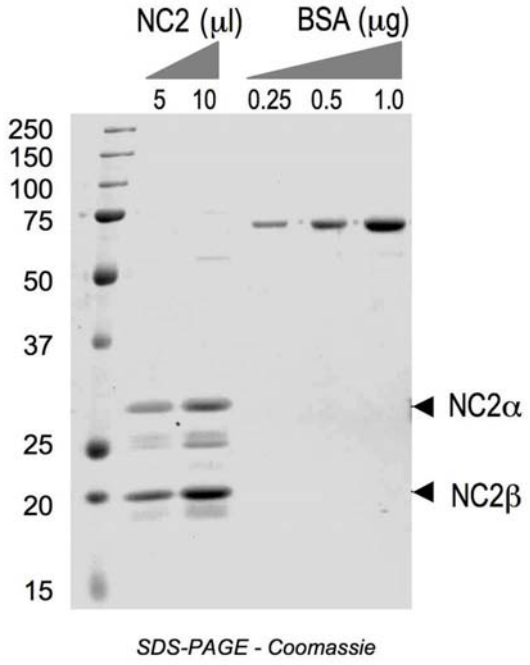



**Figure 2.14** SDS-PAGE analysis of recombinant NC2.

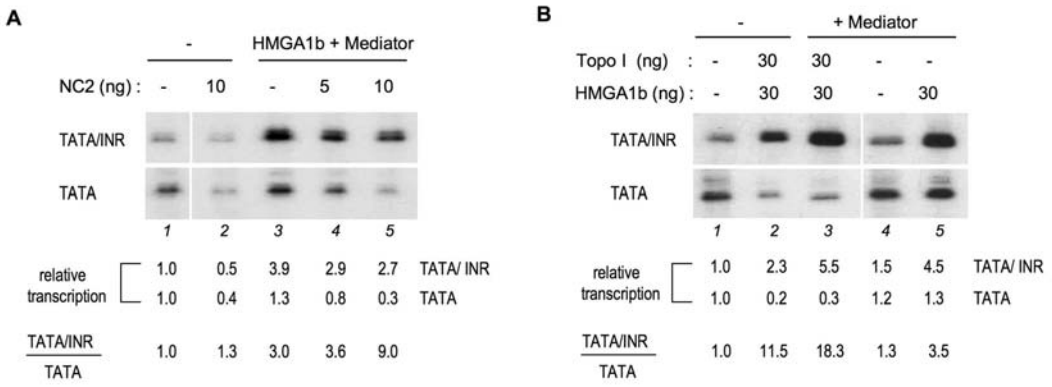



**Figure 2.15** HMGA1 and Mediator counteract the repression of NC2 and Topo I in the presence of an INR.

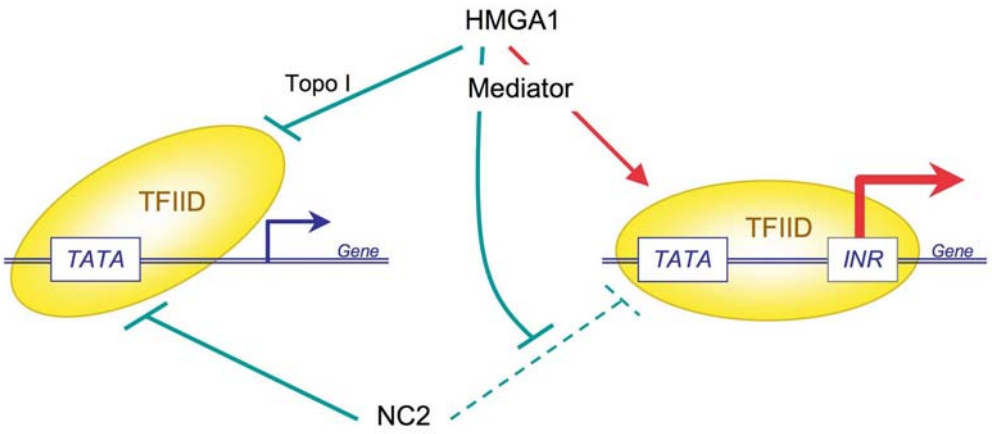



**Figure 2.16** Graphical summary.

**Subchapter 2.1**

## Role of HMGA1 in the synergy of TATA-box and INR elements in mammalian cells

**Abstract**

*In vitro* data (Chapter 2) demonstrates the requirement of HMGA1 and Mediator for TATA/INR synergy. Here, we further investigate the role of HMGA1 in the synergy of TATA-box and INR elements in mammalian cells. Interestingly, HMGA1 knockdown specifically reduces TATA/INR core promoter driven transcription without affecting TATA core promoter driven transcription. In addition, HMGA1-stimulated genes of the mouse embryonic stem cells tend to have both TATA and INR core promoter elements in a synergistic configuration, while HMGA1-repressed genes do not have any core promoter preference. In summary, both HMGA1 knockdown and knockout analysis confirm HMGA1 as TATA/INR core promoter-specific factor *in vivo*.

## Introduction

After we identified HMGA1 and Mediator as TATA/INR core promoter specific factors *in vitro* (Chapter 2), we further investigated their core promoter preferences *in vivo*. TATA-box and INR elements were previously reported to synergize with each other in cells when INR locates 25-30 bp downstream of TATA-box (Emami et al., 1995). In another study, TATA, INR, DPE and MTE core promoter elements were cloned together to construct a "Super Core Promoter", which strongly activated transcription *in vivo* by luciferase assay (Juven-Gershon et al., 2006). Mediator contains 30 subunits but HMGA1 is a single peptide protein which makes it easier to knockdown (Zhang and Wang 2008; Malik and Roeder 2010). HMGA1 have been knocked down in pancreatic cancer cells to investigate its functions in promoting cancer growth (Kolb et al., 2007).

Genome-wide core promoter element analyses of human genes showed that there are only 24% genes contain a TATA-box, while 46% genes contain an INR (Yang et al., 2007). In the other statistics analyses, 7995 mouse genes have been classified into different groups based on their core promoter elements and similar distributions of core promoter elements were obtained from the analyses (Jin et al., 2006). Genome-wide microarray was performed with HMGA1 knockout mouse embryonic stem cell and genes altered (stimulated/repressed) more than four-fold were listed (Martinez Hoyos et al., 2004). A genome-wide core promoter structure study of HMGA1-dependent genes would therefore be valuable.

## Results

### HMGA1 knockdown selectively decreased TATA/INR-driven transcription in HEK293 cells

To test the possible INR-dependent function of endogenous HMGA1 in mammalian cells, we analyzed the activity of the TATA/INR and TATA core promoters fused to a luciferase reporter gene in transfected HEK293 cells. As expected, the basal TATA/INR dependent luciferase activity was significantly higher than that of the TATA promoter (Fig. 2.1.2). Depletion of endogenous cellular HMGA1 by RNA interference (RNAi) using a specific short interfering RNA (siRNA) (Fig. 2.1.3) selectively inhibited TATA/INR but not TATA reporter activity, or the activity of the TATA-only β-actin promoter-luciferase (ACTB-Luc) reporter (Fig. 2.1.1; 2.1.4). Importantly, the TATA/INR promoter-selective effect of HMGA1 knockdown was confirmed by primer extension analysis of correctly initiated luciferase mRNA transcripts (Fig. 2.1.5). Moreover, the HMGA1 requirement for TATA/INR-dependent transcription was also observed with a different construct having different DNA sequences flanking the consensus TATA-box and INR elements (Fig 2.1.1 G5-TATA/INR-TK; Fig 2.1.6), suggesting that specific flanking sequences are not required for the TATA/INR-specific basal activity of HMGA1 either *in vitro* (Chapter 2) or *in vivo*.

**HMGA1 knockout preferentially affected TATA/INR containing genes in mouse embryonic stem cells**

To further investigate the possible core promoter-specific regulation of natural target genes by HMGA1, we performed a statistical analysis of published differential mRNA expression data obtained from mouse embryonic stem (ES) cells after the knockout of the *Hmga1* gene (Martinez-Hoyos et al., 2004). Of 13,059 Murine transcripts that were analyzed by Affymetrix oligo array, a total of 1,863 (14.3%) were differentially expressed by at least two-fold in HMGA1-knockout ES cells. To minimize indirect effects and false positives, we focused on the 250 differentially expressed gene transcripts (1.9%) that were most highly dependent on HMGA1, i.e., affected by at least four-fold and validated by RT-PCR (Martinez Hoyos et al., 2004). These included 103 transcripts from known mouse genes, of which 76 had experimentally validated transcription initiation sites and well-annotated core promoter elements from genome-wide bioinformatics studies (Jin et al., 2006). Of these 76 HMGA1-dependent transcripts, 44 were down-regulated (Table 2.1.2) and 32 were up-regulated (i.e., four-fold or more) in HMGA1-knockout ES cells. We separated the genes in these two groups according to the reported presence or absence of TATA-box and/or INR elements in their core promoters (Jin et al. 2006), and compared the frequencies of specific core promoter types in HMGA1-activated and HMGA1-repressed groups to the global frequencies of promoter types in the mouse genome (Table 2.1.1). Interestingly, the group of HMGA1-stimulated genes was significantly enriched in core promoters having both

TATA-box and INR elements (TATA/INR type), while other promoter types in this group did not significantly differ from their global frequencies in the genome. In contrast, the frequencies of most promoter types, including TATA/INR, in the HMGA1-repressed group of genes did not significantly differ from their global genomic frequencies. We note, however that the "none" category of promoters lacking both TATA-box and INR is underrepresented in the HMGA1-repressed group of genes, consistent with the fact that these promoters are generally GC-rich (AT-poor).

Altogether, these results are consistent with the *in vitro* transcription analysis and suggest a novel core promoter-dependent role of HMGA1 in gene-specific regulation in mammalian cells involving potentiation of the transcription synergy of TATA and INR elements.

**Discussion**

In support of a core promoter-selective stimulatory function of HMGA1 *in vivo*, we have further shown that endogenous HMGA1 in mammalian cells contributes to the transcription stimulatory activity of the INR at TATA- and INR- containing (TATA/INR) core promoters and that physiological HMGA1-activated (but not HMGA1-repressed) target genes in ES cells often have core promoters with TATA-box and INR elements in a synergistic configuration (Fig. 2.1.4, 2.1.5 and Table 2.1.1). Importantly, genes such as *Egr1*, *Nf2*, *Myc*, *Lipe* and *Rara* contains two accessions (Table 2.1.2), suggesting that these genes might have two transcription start sites. In order not to bring any bias, both of the accessions are analyzed according to the bioinformatic study at Jin V et al., 2006. Given the multiple post-translational modifications and diverse chromatin and gene regulatory roles of HMGA1 (Reeves 2003), including the novel core promoter-specific functions described here, the particular contributions of HMGA1 in regulation of specific genes are likely to be cell type- and context-dependent, consistent with the observed tissue specificity of HMGA1-dependent gene regulation (Martinez Hoyos et al., 2004).

Although there are 27.3% of TATA/INR containing genes affected by HMGA1 knockout (compared to 8% global percentage, Table 2.1.1), large portion of TATA/INR genes are not affected by HMGA1. The reasonable explanation would be the other upstream activators might function through the other coregulators other than HMGA1. Thus, it is important to test

the core promoter-specific cofactor functions of the other HMG proteins, such as HMGA2 and HMGB. Alternatively, other non-HMG architectural cofactors may also substitute for HMGA1 in cell-type-specific and gene-specific manner. For example, Topo I, HMGA1 and Mediator combination function best to support TATA/INR synergy *in vitro* (Chapter 2, Fig. 2.15). Topo I Chip-seq data showed that Topo I co-localized with RNA Pol II genome-widely (Cold spring harbor meeting of eukaryotic transcription mechanisms, 2011), which suggested Topo I may also be a good candidate as a core promoter specific cofactor for RNA Pol II. Furthermore, TAF1 of certain organisms (e.g. *Drosophila* and the other Fly families) containing HMGA1 AT-hook-1 like domains was able to bind to the transcription start sites of several *Drosophila* promoters in the absence of any architectural cofactor (Metcalf and Wassarman 2006), which suggests HMGA1 may not be required as cofactor for TFIID in some species. In summary, the requirement of HMGA1 to support TATA/INR synergy is likely to be gene-context-specific, cell-type-specific and species-specific.

## Materials and Methods

### Plasmid constructs

The plasmid pG5-TATA/INR-TK was a kind gift from Dr. Stephen Smale, UCLA (Emami et al., 1995) and the pG5-TK-Luc was a generous gift from Dr. Yang Shi. The TATA/INR-Luc and TATA-Luc reporters were generated by cloning the -41/+33 Kpn I-Hind III core promoter fragments of, respectively, TATA/INR and TATA plasmids between the Kpn I and Hind III sites of pGL3-Basic vector (Promega). ACTB-Luc reporter plasmid was obtained by PCR amplification of the human beta-ACTIN promoter (-120 to +100) with forward primer: 5'-TCTAGTgagctcGCGAAGCCGGTGAGTGAGCG-3', and reverse primer: 5'-TCACCGaagcttCCGGTCGGCTGGCCGGGCTT-3, and cloned between the Sac I and Hind III sites of pGL3-Basic vector.

### Cell culture, Transient Transfection and Luciferase Assay

HeLa and HEK293 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at 37 degree with 5% CO2. Transient transfections of HEK293 cells were performed with Lipofectamine$^{TM}$ 2000 (Invitrogen), and luciferase assays were performed as described previously (Faiola et al. 2005).

**Antibodies**

HMGA1 (sc-1564) and β-actin (sc-1616R) were purchased from Santa Cruz Biotechnology.

**RNA interference (RNAi) analysis**

For RNAi and analysis of luciferase reporter transcripts by primer extension, HEK293 cells were plated in 6-well plates (at ~60% confluence). About 16 h later the cells (at ~80% confluence) were transfected with Lipofectamine 2000 (Invitrogen) with 100 nM siRNA, according to manufacturer's protocol. After 20 h the cells were re-transfected with 100 nM siRNA and, where indicated, the reporter constructs (3 μg TATA or TATA/INR-Luc and/or 0.5 μg ACTB-Luc). After 48 h total RNA was extracted with the RNeasy kit (Qiagen) and 10 μg total RNA was annealed to a 32P-radiolabelled 24mer luciferase primer or TK primer (Kosovsky and Johanes 1995; Emami et al., 1995), and primer extension was performed as previously described (Martinez et al., 1994). Alternatively, whole cell extracts were prepared and analyzed for HMGA1 expression by Western blotting or for luciferase activity. Total mRNA was extracted with Qiagen Rneasy kit. HMGA1 mRNA transcripts was detected by real time PCR. 1 μg total RNA was used to reversely transcribed mRNA into cDNA by iScript cDNA kit (BioRad). cDNA concentration was measured and diluted 10-fold for real time PCR. Beta-actin primers were PCR internal control. HMGA1 primers were described as previously

(Kolb et al., 2007): HMGA1 forward: 5'-CAGCGAAGTGCCAACACCTAAG-3', HMGA1-reverse: 5'-CCTTGGTTTCCTTCCTGGAGTT-3'; Beta-ACTIN forward: 5'-TGACGGGGTCACCCACACTGTGCCCA-3'; reverse: 5'-CTAGAAGCATTTGCGGTGGACGATGGAGGG-3'. PCR signals for HMGA1 mRNA were normalized to β-ACTIN mRNA. The negative control siRNA (Silencer R Negative Control siRNA#1) was purchased from Applied Biosystems. HMGA1-specific siRNA was described previously (Kolb et al., 2007) and synthesized by Applied Biosystems (5'-GACCCGGAAAACCACCACAtt-3' and 5'-UGUGGUGGUUUUCCGGGUCtt-3').

**Core promoter statistics analysis**

The core promoter structures of 7995 mouse genes that are also conserved in human are described in the supplemental data of previous Bioinformatics study (Jin et al., 2006). According to the paper, the frequencies of gene promoters containing TATA/INR, TATA-only, INR-only and None are 8%, 8%, 33% and 51%, respectively. These percentages are the global percentages of different core promoters in mouse.

Genes altered more than four-fold by HMGA1 knock out in mouse embryonic stem cell were listed in a table previously (Martinez Hoyos et al., 2004). Core promoter structures of these genes are assorted based on previous results from the Bioinformatics study (Jin et al., 2006). The transcription start sites of 76 genes out of 103 in the table were well

experimentally characterized. Among these 76 genes, 44 genes were activated by HMGA1 and 32 were repressed by HMGA1. The core promoter distributions of HMGA1 simulated genes are TATA/INR (27%), TATA (14%), INR (23%) and None (36%). The core promoter distributions of HMGA1 repressed genes are TATA/INR (9%), TATA (16%), INR (47%) and None (28%). These ratios are compared with the global ratios of TATA/INR (8%), TATA (8%), INR (33%) and None (51%) by Fisher's exact test and the results are described in Table 2.1.1.

**Figure legends**

**Figure 2.1.1** Schematic structure of the promoters analyzed *in vivo*. ACTB-Luc promoter from -120 to +100 bp contains a CCAAT-box (-90) and a CCArGG-box (-60) upstream of TATA-box (-30). G5-TATA/INR-TK contains 5 Gal4-binding sites immediately upstream of TATA and a consensus TdT INR at +1 transcription start site. The flanking sequences between TATA/INR belong to the DNA sequences from pUC19 plasmid. The coding gene for G5-TATA/INR-TK is mouse thymidine kinase.

**Figure 2.1.2** TATA-box and INR cooperate to stimulate transcription in HEK293 cells. TATA/INR-Luc and TATA-Luc were transfected into HEK293 cells and the relative luciferase activities (mean ± S.D.) from 6 independent experiments (each in duplicate) are shown. The luciferase activity of TATA-Luc was arbitrary set to 1.

**Figure 2.1.3** Knockdown of HMGA1 in HEK293 cells. **(A)** HEK293 cells were transfected with mock, control siRNA or specific siRNA against HMGA1. HMGA1 mRNA transcripts were analyzed by real time PCR and *ACTB* mRNA transcripts were used as internal control. **(B)** HEK293 cells were transfected with mock, control siRNA or specific siRNA against HMGA1. Endogenous HMGA1 in TCA precipitated samples was analyzed by Western blot and total proteins with ponceau S staining were used as loading control.

**Figure 2.1.4** HMGA1 knockdown reduces TATA/INR-driven transcription but not TATA-only driven transcription in HEK293 cells. Reporter plasmids TATA/INR-Luc, TATA-Luc, or ACTB-Luc reporters were co-transfected into HEK293 cells with Control siRNA (black bars) or HMGA1-specific siRNA (open bars). The relative luciferase activities normalized to pGL3 basic vector. In left panel, luciferase activities from TATA-Luc were set arbitrarily to 1; in right panel, luciferase activities from ACTB-Luc co-transfected with Control siRNA were set arbitrarily to 1.

**Figure 2.1.5** HMGA1 knockdown preferentially reduces TATA/INR-directed mRNA transcripts. Total mRNA samples from HEK293 cells transfected with ACTB-Luc (lanes 1-6) with either TATA/INR-Luc (lanes 1-3) or TATA-Luc (lanes 4-6), and the indicated siRNAs, were analyzed by primer extension with the Luc 24mer primer. Lane 7 is control reaction with mRNA from mock transfected cells. The positions of correctly initiated transcripts are indicated for each promoter construct (length of 172 nts for ACTB-Luc and 128 nts for TATA-Luc and TATA/INR-Luc). Top and bottom autoradiograms are from top and bottom parts of the same gel; bottom autoradiogram is from a longer X-ray film exposure time.

**Figure 2.1.6** HMGA1 knockdown reduces mRNA transcript from TATA/INR promoter with different flanking DNA sequences. **(Top-basal)** Cells were co-transfected with siRNAs (control or HMGA1-specific) and the reporter plasmid Gal-TATA/INR-TK. Total RNA was extracted from transfected cells and analyzed by primer extension to detect the expression of

correctly initiated TK mRNAs. The arrowhead indicates specific primer extension products (84 nts in length). **(Bottom-activated)** Cells were co-transfected with siRNAs (control or HMGA1-specific) and the reporter G5-TATA/INR-Tk and activator plasmid Gal4-VP16. Total RNA was extracted from transfected cells and analyzed by primer extension to detect the expression of correct initiated TK mRNAs. The arrowhead indicates specific primer extension products (84 nts in length). The asterisk points to a non-specific band also present in the control lane from untransfected cells (lane 3).

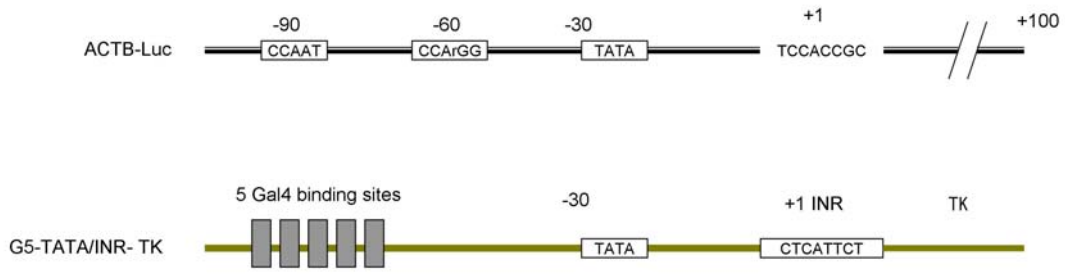



**Figure 2.1.1** Schematic structure of the promoters analyzed *in vivo*.

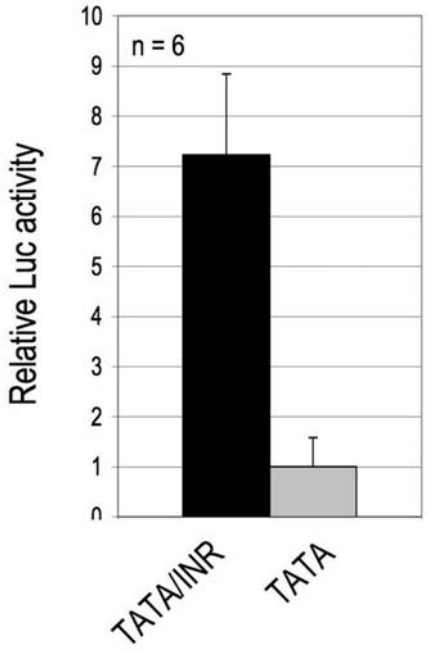
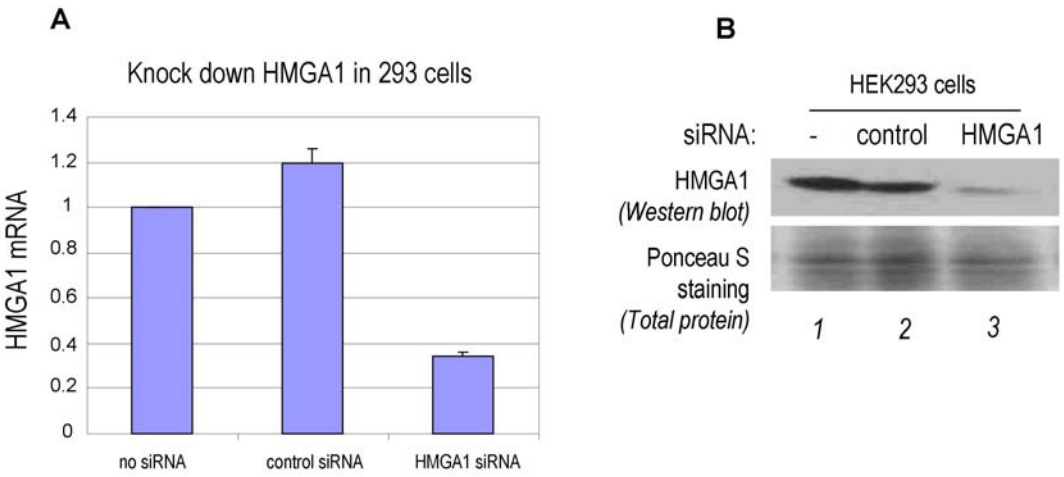



**Figure 2.1.2** TATA and INR cooperate to stimulate transcription in HEK293 cells.




Real time PCR internal control primer is primer of ACTB.

**Figure 2.1.3** Knockdown of HMGA1 in HEK293 cells.

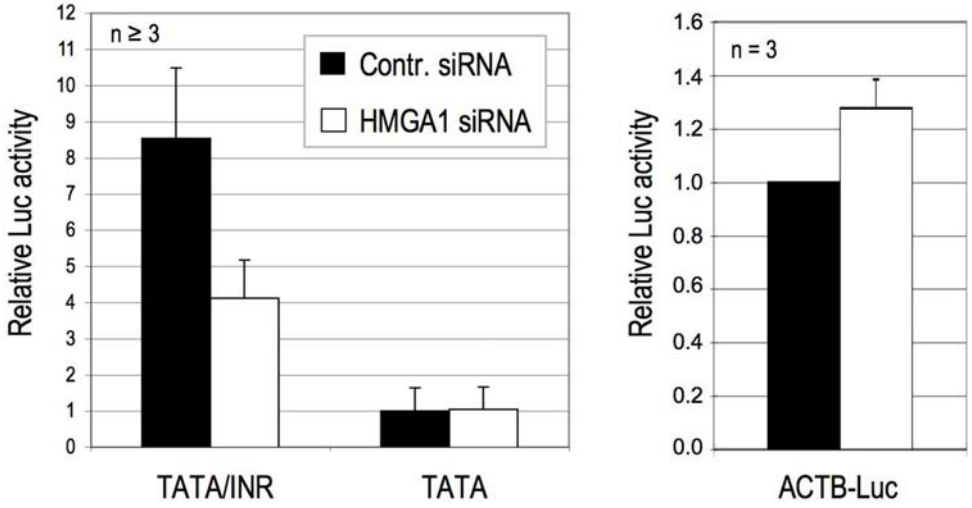



**Figure 2.1.4** HMGA1 knockdown reduces TATA/INR-driven transcription but not TATA-only driven transcription in 293 cells.

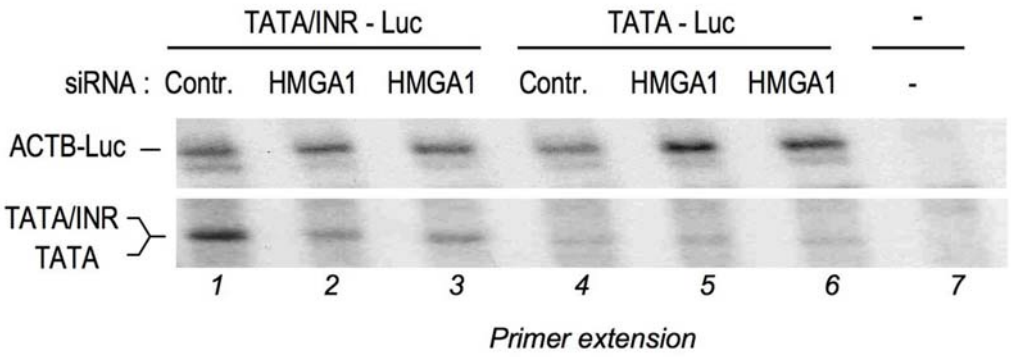



**Figure 2.1.5** HMGA1 knockdown preferentially reduces TATA/INR-directed mRNA transcripts.

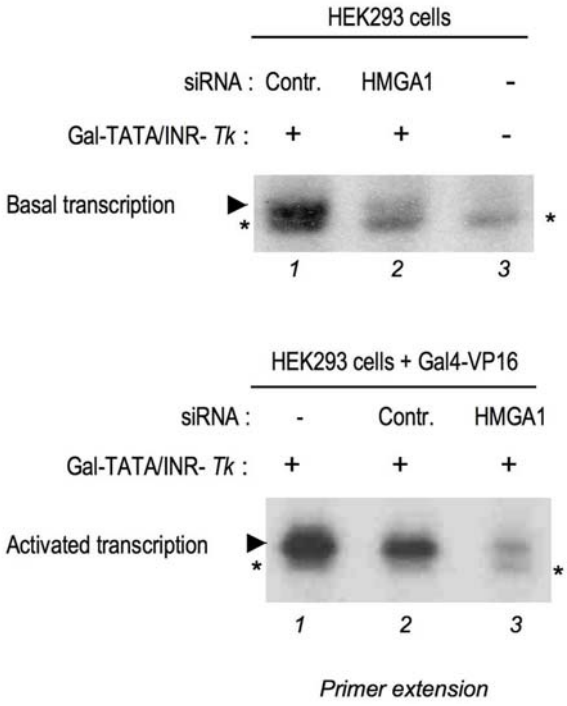



**Figure 2.1.6** HMGA1 knockdown reduces mRNA transcript from the TATA/INR promoter with different flanking DNA sequences.

| Core promoter type | All | TATA/INR | TATA-only | INR-only | None |
|---|---|---|---|---|---|
| Genes characterized[a] | 7995 (100%) | 640 (8%) | 664 (8%) | 2653 (33%) | 4038 (51%) |
| HMGA1-stimulated[b] | 44 (100%) | **12*** (27%)** | 6 (14%) | 10 (23%) | 16 (36%) |
| HMGA1-repressed[b] | 32 (100%) | 3 (9%) | 5 (16%) | 15 (47%) | **9* (28%)** |

Significant differences ($p<0.05$) from genomic frequencies analyzed by two-sided Fisher's test and indicated in bold; ***p = 0.000137; *p = 0.012694.
a. Frequencies of core promoters types in 7995 mouse genes with experimentally-validated TSSs (Jin et al., 2006).
b. Genes in (a) whose transcripts levels are altered by at least 4-fold in HMGA1 knockout mouse ES cells (Martinez Hoyos et al., 2004).

**Table 2.1.1** Frequencies of core promoter types for HMGA1-regulated genes in ES cells. (http://www.quantitativeskills.com/sisa/statistics/fisher.htm). Fisher exact test is performed using the "SISA" website as the url above. The numbers are filled as follows: Row 1 (640, 7355) and Row 2 (12, 32). The advised p-value is taken from the Fisher exact test results.

| Symbol | Gene ID, Accession | Structure | Symbol | Gene ID, Accession | Structure |
|---|---|---|---|---|---|
| *Oxt* | 18429, NM_012996 | TATA/INR | *Fstl1* | 14314, BC006185 | TATA |
| *Fos* | 14281, V00727 | | *Sox4* | 20677, NM_009238 | |
| *Tcfap2c* | 21420, NM_009335 | | *Myc* | 17869, NM_010849 | |
| *Egr1* | 13653, M22326 | | *Junb* | 16477, NM_008416 | |
| *Col1a2* | 12843, AK031577 | | *Id3* | 15903, M60523 | |
| *Col1a1* | 12842, X54876 | | *Mt3* | 17751, BC059725 | |
| *Acta2* | 11475, U63129 | | | | |
| *Cyp2a12* | 13085, M33313 | | | | |
| *Trh* | 22044, AK010666 | | | | |
| *Igfbp4* | 16010, AK031212 | | | | |
| *Cyr61* | 16007, M32490 | | | | |
| *Klf2* | 16598, NM_008452 | | | | |
| Symbol | Gene ID, Accession | Structure | Symbol | Gene ID, Accession | Structure |
| *Nf2* | 18016, L27105;<br>18016, NM_010898 | None | *Rara* | 19401, X56572;<br>19401, M60909 | INR |
| *Fstl1* | 14314, U06864 | | *Pim2* | 18715, AK043570 | |
| *Ptpru* | 19273, NM_011214 | | *Csk* | 12988, NM_007783 | |
| *Myc* | 17869, K03229 | | *Nfyc* | 18046, AK011327 | |
| *Nfix* | 18032, S81451 | | *Mgat1* | 17308, AK004760 | |
| *Egr1* | 13653, M28844 | | *Sod3* | 20657, NM_011435 | |
| *Hck* | 15162, NM_010407 | | *Trh* | 22044, X59387 | |
| *Tagln* | 21345, Z68618 | | *Igfbp3* | 16009, BC058261 | |
| *Mapt* | 17762, M18775 | | *Ccnd1* | 12443, M64403 | |
| *Sod3* | 20657, AF039602 | | | | |
| *Clcn6* | 26372, NM_011929 | | | | |
| *Cd19* | 12478, AK089835 | | | | |
| *Igfals* | 16005, BC020357 | | | | |
| *Lipe* | 16890, U08188;<br>16890, AF179427 | | | | |

**Table 2.1.2** Classification of the core promoters of genes stimulated more than four-fold by a HMGA1 knockout in mouse embryonic stem cells.

**Subchapter 2.2**

## The interactions of HMGA1 with TFIID and Mediator are required to support TATA/INR synergy

**Abstract**

Functionally, HMGA1, Mediator and TFIID cooperatively stimulate TATA/INR synergy *in vitro* and in cells (Chapter 2 and Subchapter 2.1). Here we show that HA-HMGA1 co-immunoprecipitats with both TFIID and CDK8-less Mediator. In addition, endogenous HMGA1 interacts with Med1/Med220. Co-immunoprecipitation (Co-IP) using various HMGA1 deletion mutants demonstrates that the acidic COOH-tail of HMGA1 is required but not sufficient to interact with TFIID and Mediator, and that the first $NH_2$-terminus AT-hook is not required for the protein interactions. Consistent with this, the COOH-tail of HMGA1 is required to support maximal TATA/INR synergy. Unexpectedly, Magnesium Agarose Gel Shift shows that HMGA1 does not further facilitate TFIID to bind either TATA/INR or TATA core promoters. Further investigation with other experimental assays is required to uncover the mechanisms responsible for these effects.

## Introduction

HMGA1 has been shown to assist in the formation of INF-beta enhanceosomes (Thanos and Maniatis 1995; Reeves 2003). HMGA protein families do not have an intrinsic transcription regulatory domain or a strict DNA sequence specificity but bind to the minor groove of AT-rich or structured DNA through "AT-hook" motifs and to numerous sequence-specific regulators. HMGA1 is thought to act as a chaperone to induce or stabilize DNA and/or protein conformations that facilitate cooperative binding of activators to specific enhancers (Reeves and Beckerbauer 2001; Reeves 2003; Panne 2008).

Co-IP conjugated with mass spectrometry has identified 11 proteins associated with HMGA1, most of which belong to four functional classes: transcriptional proteins, mRNA processing proteins, RNA helicases and protein chaperones (Sgarra et al., 2008). Among these proteins, TFII68 and TFIIF are proteins involving transcription initiation. In another separate study, a peptide fragment of TAF3 was shown to interact with HMGA1 (Malini et al., 2011). Altogether, these previous data suggested that HMGA1 may function in preinitiation complex formation through interactions with GTFs. Interestingly, *Drosophila* TAF1 isoforms with two HMGA1-like AT-hooks are able to bind to DNA probes containing *Drosophila* promoter transcription start site (Chalkley and Verrijzer 1999; Metcalf and Wassarman 2006). Thus, the physical interactions between HMGA1, TFIID, Mediator and the other GTFs should be investigated further.

TFIIA facilitated cooperative recruitment of TFIID to core promoters with both TATA-box and INR elements in synergistic configuration by Magnesium Agarose Gel Shift (Liberman and Berk 1991; Emami et al., 1997). HMGA1 was required for NFκB (p65/p50), c-Jun and the other IFN-related proteins to form a stable enhancerosome to stimulate virus-inducible IFN-beta transcription (Yie et al., 1999). Similarly, Magnesium Agarose Gel Shift assay could also be employed to test whether HMGA1 can facilitate the preferential binding of TFIID/TFIIA to TATA/INR promoter. Mechanistically, INR stimulated TATA-dependent transcription under single-round transcription condition in crude nuclear extracts (Yean and Gralla 1997), which suggested that INR is likely to function at the pre-initiation complex formation step. Because HMGA1 and Mediator stimulated INR function (Chapter 2 and Subchapter 2.1), further experiments need to be performed to analyze how HMGA1/Mediator function to facilitate pre-initiation complex assembly.

## Results

### HMGA1 interacts with TFIID and CDK8-less Mediator

To investigate the possible mechanisms for the functional cooperativity of HMGA1 with TFIID and Mediator, we tested their possible interactions in HEK293 cells. We first used a HEK293 cell line that expresses low levels of ectopic HA epitope-tagged HMGA1b (HA-HMGA1b) (Fig. 2.2.1). Immunoprecipitation with an anti-HA antibody and Western blot analyses demonstrated a specific interaction of HA-HMGA1b with all TFIID subunits tested (Fig. 2.2.2-A) and with several subunits of Mediator, but not with the CDK8 subunit (Fig. 2.2.2-B), which was proposed to be present only in negative isoforms of Mediator (Hans et al., 2006). We further confirmed the specific interaction of the Mediator complex with endogenous HMGA1 in HeLa and HEK293 cells by Co-IP with an anti-MED1 antibody (Fig. 2.2.3; and Fig. 2.2.4). Notably, endogenous HMGA1b was significantly enriched in the anti-MED1 immunoprecipitation (Fig. 2.2.3, lane 6; and Fig. 2.2.4). Moreover, the interactions observed were not affected by the presence of ethidium bromide (EB), suggesting that they were not indirect effects of DNA binding.

**Acidic COOH-tail of HMGA1 is required but not sufficient to interact with TFIID and Mediator and is required for maximal TATA/INR synergy**

To map the domain(s) of HMGA1b required for interaction with TFIID and Mediator, HA-HMGA1b wild type and several deletion mutants were transiently transfected into HEK293 cells and analyzed by co-immunoprecipitation and Western blotting. Interestingly, we found that the C-terminal acidic tail domain is important for HMGA1b interaction with both TFIID and Mediator (Fig. 2.2.5). To our knowledge this represents the first identification of HMGA1-interacting factors that depend on this conserved domain of HMGA proteins. Accordingly, the GST-HMGA1-ΔC mutant that lacks the COOH-tail could not interact with TFIID or Mediator. However, the COOH-tail domain consisting of 17 amino acids fused to Glutathione S-transferase (GST) was not sufficient to interact with either TFIID or Mediator. Altogether, the results from GST-pull down assays suggested that the $NH_2$-terminus and COOH-tail of HMGA1 are both required for HMGA1 protein interactions (Fig. 2.2.6). Interestingly, the first $NH_2$-terminus AT-hook of HMGA1 is not required for these protein interactions (Fig. 2.2.5).

We further tested the possible contribution of the C-terminal tail in the core promoter-specific basal stimulatory activity of HMGA1b by titrating purified recombinant HMGA1b wild type (wt) and a COOH-tail-deleted (ΔC) mutant (Fig. 2.2.7) in the purified transcription system in the presence of Mediator. While HMGA1b wt significantly stimulated

the MLP core promoter, but not the HSP70 promoter, at all concentrations tested (Fig. 2.2.8, lanes 2-4), the ΔC mutant had a drastically reduced activity and repressed transcription from both promoters at the highest concentration (Fig. 2.2.8, lanes 5-7). Similarly, HMGA1b ΔC mutant had a reduced INR-specific stimulatory activity on the TATA/INR core promoter and an increased repressive function at the highest concentration on both promoters (Fig. 2.2.8, right). However, the TATA/INR promoter was less sensitive to this repressive effect consistent with the INR-dependent activity of TAFs and Mediator in antagonizing the negative function of HMGA1 and other negative cofactors, described previously (Fig. 2.15). Thus, HMGA1 has both (i) a core promoter/INR-selective basal stimulatory function that cooperates with Mediator and TFIID/TAFs and requires the COOH-tail domain (and possibly the $NH_2$-terminus AT-hooks), and (ii) a negative function at the N-terminus (containing the three AT-hooks) that is suppressed by the COOH-tail domain and antagonized in an INR-dependent manner by Mediator and TFIID/TAFs. Hence, these results suggest that the interaction of TFIID and Mediator with HMGA1 is dependent on the acidic COOH-tail domain and important for their concerted core promoter-selective basal activities.

**HMGA1 can not facilitate TFIID binding to either TATA/INR or TATA core promoter probes by Magnesium Agarose Gel Shift assay**

TFIIA was shown to facilitate TFIID binding preferentially to TATA/INR core promoter probe by Magnesium Agarose Gel Shift (Emami et al., 1997). Since HMGA1 interacts with

TFIID and Mediator (see Results), we performed similar experiments to test the effect of HMGA1 on TATA/INR and TATA probes (Materials and Methods). Consistent with previous results: first, TBP could not bind to either TATA/INR or TATA in the absence of TFIIA; while in the presence of TFIIA, TBP-TFIIA complex bound to TATA/INR and TATA with similar pattern (Fig. 2.2.9-Top). Second, TFIID by itself bound to TATA/INR and TATA in a similar way. Third, when TFIIA was titrated in the presence of TFIID, TFIIA gradually stimulated the selective binding of TFIID to TATAT/INR (the slective effect on TATA/INR is weak), but TFIIA also had minor effect on TFIID binding to TATA probe (Fig. 2.2.10, lanes 2-5 and 12-15). The selective effect of TFIID-IIA on TATA/INR was not as dramatic as described in previous studies, suggesting there might be differences between either TFIID/TFIIA or the core promoter probes used in this study and those used in previous studies (Emami et al., 1997). When HMGA1 was titrated from 0.1 ng to 10 ng in the presence of TFIID, HMGA1 could not further facilitate TFIID binding to either TATA/INR or TATA probes (Fig. 2.2.10 lanes 6-10 and 16-20). Finally, HMGA1 was titrated in the presence of the TFIID-IIA complex and had no effect on TFIID-IIA binding to core promoters (Fig. 2.2.11). More assays (e.g. immobilized template recruitment assay) should be employed to study the mechanisms further (see discussion part).

**Discussion**

Although the detailed molecular mechanisms underlying the core promoter-selective cooperativity of HMGA1, Mediator, and TFIID/TAFs in INR-dependent transcription remain to be fully characterized, our results point to an important regulatory role of the conserved acidic COOH-tail domain of HMGA1. This acidic COOH-tail domain is conserved in HMGA proteins (HMGA1a/b and HMGA2) and appears to have important biological functions in regulation of cell proliferation and oncogenic transformation by HMGA proteins, whose overexpression is a hallmark of malignant tumours (Pierantoni et al., 2003; Fusco and Fedele 2007; Li et al., 2007). However, the molecular functions of this COOH-tail domain have remained unclear. The acidic COOH-tail domain of HMGA1 does not have an intrinsic transactivating function when fused to the DNA binding domain of GAL4 (Thanos and Maniatis 1992; Zhou et al., 1996), although it appears to be required for HMGA1 coactivation of some, but not all, HMGA1-dependent activators and target promoters in transfected cells (Yie et al., 1997; Chin et al., 1998). Here, we have shown that the COOH-tail domain of HMGA1 is required both for HMGA1 interaction with Mediator and TFIID in human cells (Fig. 2.2.5) and for maximal stimulation of INR-dependent basal transcription by Mediator and TFIID/TAFs *in vitro* (Fig. 2.2.8). In addition, the COOH-tail domain antagonizes the repressive function of the $NH_2$-terminus region containing the three AT-hooks (Fig. 2.2.8). This anti-repressive function of the COOH-tail correlates with the reported roles of the acidic COOH-tail domains of HMGA1 and HMGA2 in restricting the DNA-binding and

self-association activities of the $NH_2$-terminus region containing the AT-hooks (Nissen and Reeves 1995; Yie et al., 1997; Noro et al., 2003). Thus, the DNA-binding activity of HMGA1, which recognizes AT-rich and structured sequences, including TATA elements and nucleosomal DNA, could be altered in association with TFIID and/or Mediator. Alternatively (or in addition), HMGA1 could selectively interact with specific variants of Mediator or TFIID complexes or could affect their structure at enhancers and/or core promoters. Indeed, HMGA1 interacts selectively with a form of Mediator that lacks the CDK8 subunit (Fig. 2.2.2). In addition, structural effects of activators and TBP on Mediator conformation (Taatjes et al., 2002; Meyer et al., 2010; Taatjes 2010; Cai et al., 2010) and on isomerization of TFIID on core promoter DNA (Horikoshi et al., 1988; Lieberman and Berk 1994; Chi and Carey 1996) have been reported, which suggest a possible malleability of these complexes. In this context, it is interesting to note that the COOH-tail domain of HMGA1 is phosphorylated by CK2 (reviewed in Reeves, 2003), a protein kinase with reported core promoter-specific regulatory activities (Lewis et al., 2005). Thus, posttranslational modifications of HMGA1 could modulate its interactions with Mediator or TFIID and, hence, regulate its core promoter-selective functions.

TFIIA was shown to facilitate TFIID binding preferentially to TATA/INR core promoter probe by Magnesium Agarose Gel Shift (Emami et al., 1997), but HMGA1 had no effect on TFIID binding to either TATA/INR or TATA core promoters in the presence and absence of TFIIA (Fig. 2.2.10; Fig. 2.2.11). One explanation is that HMGA1 functions in recruitment

step(s) after TFIID binding, such as the recruitments of Mediator, TFIIB, TFIIF or RNA Pol II itself. Alternatively, HMGA1 may induce the conformation changes on TFIID, Mediator, RNA Pol II and core promoter DNA, which may largely increase the activity of RNA Pol II but may not have any effect on TFIID recruitment. Another possible explanation is that Magnesium Agarose Gel Shift assay may not be suitable to study TFIID recruitment facilitated by HMGA1. Immobilized template recruitment assay should be used to study TFIID recruitment because it works perfectly in testing the recruitment of preinitiation complex in the presence of Mediator (Lin and Carey 2012).

## Materials and Methods

### Plasmid constructs

The expression vector for 6His-HMGA1b (pET-His-HMGA1b), the mammalian expression vectors pHA-HMGA1b wild type and deletion mutants were generous gifts from Dr. Alfredo Fusco (Pierantoni et al., 2006). The bacterial expression vector for the COOH-tail deleted (ΔC) HMGA1b (pET-His-HMGA1bΔC) was obtained by replacing the Nde I-Xho I fragment of pET-His-HMGA1b encoding full-length HMGA1b with a Nde I-Xho I PCR fragment (forward primer: 5'-GATATACATATGAGCGAGTC-3' and reverse primer: 5'-TCTAGTCTCGAGCTTCTCCAGTTTCTTGGGTC-3') encoding HMGA1b residues 1-81. The GST-HMGA1-FL (full length 1-96), GST-ΔC (1-79) and GST-C-tail (80-96) are generated by Dr. Priyanka Sharma. Briefly, same forward primer with BamH I was used to amplify GST-FL and GST-ΔC (sequences: 5'-AATTggatccATGAGCGAGTCGGGCTC-3'). Reverse primer with for GST-FL: 5'-CTACgaattcTCACTGCTCCTCCTCAGAGGACT-3'. Reverse primer for ΔC: 5'-CTACgaattcTCACAGTTTCTTGGGTCTGCCCC-3'. The DNA for GST-C-tail was synthesized directly by Fisher Sci with BamH I and EcoR I (Sequences: 5'-CGTGGATCCGAGAAGGAGGAAGAGGAGGGCATCTCCCAGGAGTCCTCTGTGGT GGAGCAGTGAGAATTCATCG-3'). The DNA fragments of GST-HMGA1-FL, GST-ΔC and GST-C-tail are ligated into BamH I and EcoR I sites of pGEX-2T. All constructs were verified by DNA sequencing.

**Cell line, Antibodies, Western blot and Immunoprecipitation**

HEK293 cell line stably transfected with pHA-HMGA1b and expressing low levels of HA-tagged mouse HMGA1b were cultured as described previously (Materials and Methods, Subchapter 2.1). The antibodies obtained from commercial sources were HMGA1 (sc-1564), MED1/TRAP220 (sc-5334x), MED12/TRAP230 (sc-5374x), MED13/TRAP240 (sc-12013x), MED16/TRAP95 (sc-5363x), MED26/CRSP70(sc-48776x), CDK8(sc-1521), TAF1(sc-735x), TAF9 (sc-1247x), and b-actin (sc-1616R), all from Santa Cruz Biotechnology; and anti-HA antibody (12CA5) and anti-HA resin from Covance. TBP, TAF4, TAF5 and TAF6 were aliquots from Dr. Robert G. Roeder's laboratory. The NC2 antibodies was a kind gift from Dr. Thomas Oelgeschlager. Whole-cell extract preparation, immunoprecipitation, and Western blotting were essentially as previously described (Faiola et al. 2005). Where indicated, 50 μg/ml ethidium bromide was added to cell extracts before immunoprecipitation. For HA-HMGA1 immunoprecipitation, HA-HMGA1b stable cell line and 293 cell line were lysed with buffer (50 mM HEPES [pH 7.9] at 4°C, 250 mM NaCl, 0.1% IGEPALCA-630, 0.2 mM EDTA, 0.2 mM phenylmethylsulfonyl fluoride [PMSF], 2 mM 2-mercaptoethanol). Lysates were diluted with BC-0 (20 mM HEPES [pH 7.9] at 4°C, 20% glycerol, 0.2 mM EDTA, 0.05% IGEPAL CA-630, 0.2 mM PMSF, 5 mM 2-mercaptoethanol) to reach 179 mM NaCl final concentration (immunoprecipitation [IP] buffer), and cell extracts from one 10-cm plate were incubated with 10 μL of HA-resin for 12 h at 4°C under constant rotation. Immunoprecipitates were washed three times with IP buffer, resolved by sodium-dodecyl

sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and analyzed by Western blotting with the indicated antibodies. For immunoprecipitation of endogenous Med1, 293 cells were lysed as described above. Hela nuclear extract was used in Fig. 2.2.2-B and 293 whole cell lysate was used in Fig. 2.2.3, respectively. The indicated control Goat IgG (2-4 μg) or anti-Med1 IgG (SC-5334x, 2-4 μg) were incubated with the lysates for 14 h at 4°C and precipitated after incubation with 10 μl of protein G-Sepharose for 2 h at 4°C. Immunoprecipitates were washed three times with IP buffer or lysis buffer and analyzed by SDS-PAGE and Western blotting with the indicated antibodies.

**GST Pull Down**

1.5 mg nuclear extract (7 mg/ml) from 293 cells was adjusted to final 175 mM salt by BC-0 (BC buffer as described above, add NP-40 to finial 0.5%). 4 μg recombinant GST-FL, GST-ΔC and GST-C-tail were incubated respectively with nuclear extract over night at 4 degree. Resins were washed 3 times by BC-175, eluted by SDS-loading buffer and analyzed by SDS-PAGE following Western blot.

**Magnesium Agarose Gel Shift**

Magnesium Agarose Gel Shift was performed essentially as previously described (Liberman and Berk 1991; Emami et al., 1997). The plasmids TATA/INR and TATA were

used to amplified TATA/INR-probe and TATA-probe. (Forward: 5'-CCAGTCACGACGTTGTAAAAC-3' and Reverse: 5'-CAGCTATGACCATGATTACG-3'). 250 bps PCR products were precipitated with Ethanol and gel-purified by Qiagen Gel-extraction kit. PCR probes were labeled with T4 Polynucleotide Kinase using $P^{32}$ γ-ATP. The binding reaction (total 12.5 μL) contains 10 fmol probe (5000 cpm), 5 mM $MgCl_2$, 2.2 μg BSA, 12.5 mM Hepes pH7.9 at 4 ℃ , 12.5% Glycerol, 0.2 mM EDTA, 70 mM KCL, 60 mM β-ME, 40 μg/ml poly(dGdC):(dG-dC), add $H_2O$ to total 12.5μL if necessary. Mix proteins with binding buffers and probes, incubate at 30℃ for 60 mins. 1.4% agarose gel were made in 12.4*19 $cm^2$ cassette. Samples were run with 0.5*TBE at 45 V (20 mA) for 4-6 hours at room temeperature. Agarose gels were dried on DE-81 paper under vacuum and exposed to X-ray films.

### *In vitro* transcription assay

*In vitro* transcription assay was described previously in Materials and Methods of Chapter 2.

**Figure legends**

**Figure 2.2.1** HA-HMGA1 is stably transfected into 293 cells and expresses at a lower level than endogenous HMGA1. Western blot analyses of total HMGA1 in normal HEK293 cells (-) and a derivative clonal cell line that was stably transfected with HA-HMGA1b (+). Positions of endogenous HMGA1 and ectopic HA-HMGA1b proteins are indicated. An antibody to beta-actin was used as loading control.

**Figure 2.2.2** TFIID and CDK8-less Mediator co-immunoprecipitate with HA-HMGA1. Whole cell extracts from HEK293 cells and HEK293 cells stably expressing HA-HMGA1b (described above) were adjusted to 175 mM KCl and immunoprecipitated with anti-HA antibody resin in the presence or absence of ethidium bromide (EB), as indicated. Western blot was probed with antibodies to the indicated TFIID and Mediator subunits. NC2α served as negative control. On the buttom of panel A, a longer exposure film from the same experiment was used to show better TAF1 signal.

**Figure 2.2.3** Mediator interacts with endogenous HMGA1 in Hela nuclear extract. HeLa cell nuclear extracts were immunoprecipitated with a MED1 antibody, or mock immunoprecipitated with goat IgG (Mock), and associated proteins were analyzed by Western blot with the indicated antibodies.

**Figure 2.2.4** Mediator interacts with endogenous HMGA1 in HEK293 whole cell lysate. Whole cell extracts of HEK293 cells (input extract, lanes 2 and 3) were immunoprecipitated (at 175 mM NaCl) with a MED1 antibody or goat IgG (as negative control) and analyzed by Western blot with the indicated specific antibodies. HMGA1 in whole cell extracts is barely detectable under the conditions used but is highly enriched in the specific MED1 IP sample (lane 5). Input extracts and IP panels are from the same gel and film exposure. TBP is not specifically enriched in the MED1 IP. Lane 1 shows an extract of cells transfected with an HMGA1b expression vector. Arrowheads indicate the position of HMGA1.

**Figure 2.2.5** Acidic COOH-tail is required for HMGA1 interaction with TFIID and Mediator. Scheme of HMGA1b wild type structure, including AT-hooks (AT) and acidic tail (COOH-tail), and deletion mutants used for immunoprecipitation experiments below. HEK293 cells were mock transfected (-) or transiently transfected with HA-HMGA1b wild type (1-96) or the indicated deletion mutants. Whole cell extracts (input) were immunoprecipitated with anti-HA antibody (IP: HA) and analyzed by Western blot with the indicated antibodies.

**Figure 2.2.6** HMGA1 interaction with TFIID and Mediator *in vitro* requires both the $NH_2$-terminal AT-hook domains and the acidic COOH-tail. **(A)** Scheme of HMGA1b full-length (FL) and deletion mutants used as GST-fusion proteins for protein-protein interactions *in vitro*. **(B)** Coomassie stained SDS-PAGE analysis of purified GST and

GST-HMGA1b fusion proteins used for GST-pull down experiments as indicated. BSA served as protein standard to estimate protein concentrations. **(C)** Western blot analysis of a GST pull-down experiment with immobilized GST or GST-HMGA1b and HEK293 cell nuclear extract. Specific antibodies for the indicated subunits of TFIID and Mediator were used.

**Figure 2.2.7** SDS-PAGE analysis of recombinant HMGA1 wildtype (WT) and 15 COOH-terminal amino acids deletion mutant (ΔC). Analysis by SDS-PAGE and Coomassie blue staining of purified recombinant 6His-HMGA1b wild type protein (WT) and the ΔC mutant protein, which lacks the 15 amino acids COOH-terminal acidic tail. Both proteins were expressed in E. coli and purified in parallel from the insoluble inclusion bodies. On the left side, the molecular mass of protein markers is shown in kDa.

**Figure 2.2.8** HMGA1 acidic COOH-tail is required for HMGA1 to support maximal TATA/INR synergy. *In vitro* transcription was performed with the indicated core promoter constructs (linear form) in the purified TFIID-based system complemented with Mediator and different amounts of either wild type HMGA1b (wt) or a deletion mutant 1-81 (ΔC) that lacks the acidic COOH-tail domain (Fig. 2.2.7). Relative transcription levels were normalized to the signals in the absence of HMGA1/Mediator (lanes 1).

**Figure 2.2.9** TBP-TFIIA and TFIID bind TATA/INR and TATA core promoters in a similar pattern. **(Top)** Magnesium Agarose Gel Shift was performed with 2 ng recombinant TBP by

itself (lane 2 and 7) or 0.3 μl natural TFIIA together with different amounts of recombinant TBP (lanes 3-5 and 8-10). **(Bottom)** Magnesium Agarose Gel Shift was performed without TFIID (lane 1 and 6) or with different amount of TFIID on TATA/INR (lanes 2-5) and TATA (lanes 7-10).

**Figure 2.2.10** HMGA1 dose not facilitate TFIID binding to either TATA/INR or TATA core promoter in the absence of TFIIA. The binding of TFIID to TATA/INR or TATA was tested together with different amounts of natural TFIIA from 0.25 μl to 4 μl on (lanes 3-5 and lanes 13-15) or with different amounts of recombinant HMGA1 from 0.1 ng to 10 ng (lanes 6-10 and lanes 16-20).

**Figure 2.2.11** HMGA1 can not facilitate TFIID-IIA binding to core promoters. HMGA1 was titrated from 0.3 ng to 3 ng together with TFIID-IIA on TATA/INR (lanes 4-6) and TATA (lances 10-12). Gel was exposed to X-ray films.

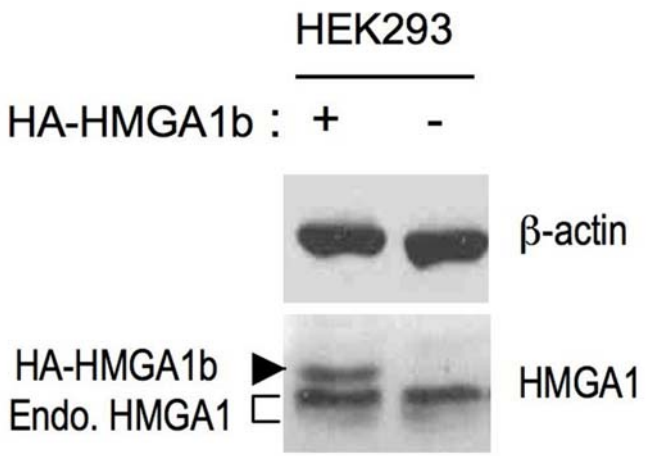



**Figure 2.2.1** HA-HMGA1 is stably transfected into 293 cells and expresses at a lower level than endogenous HMGA1.

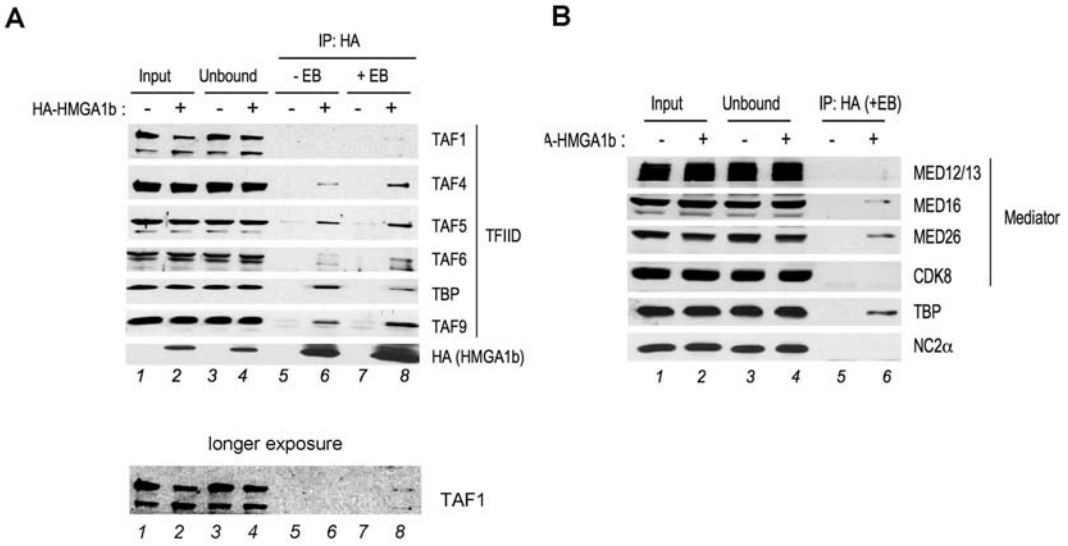



**Figure 2.2.2** TFIID and CDK8-less Mediator co-immunoprecipitate with HA-HMGA1.

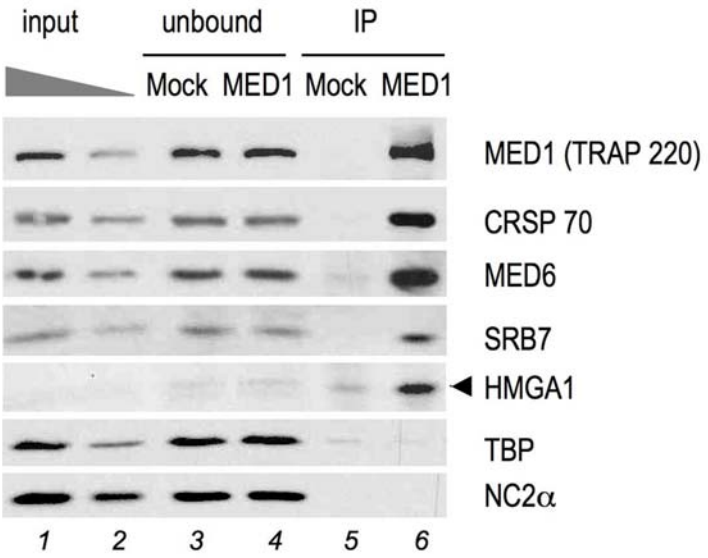



**Figure 2.2.3** Mediator interacts with endogenous HMGA1 in Hela nuclear extract.

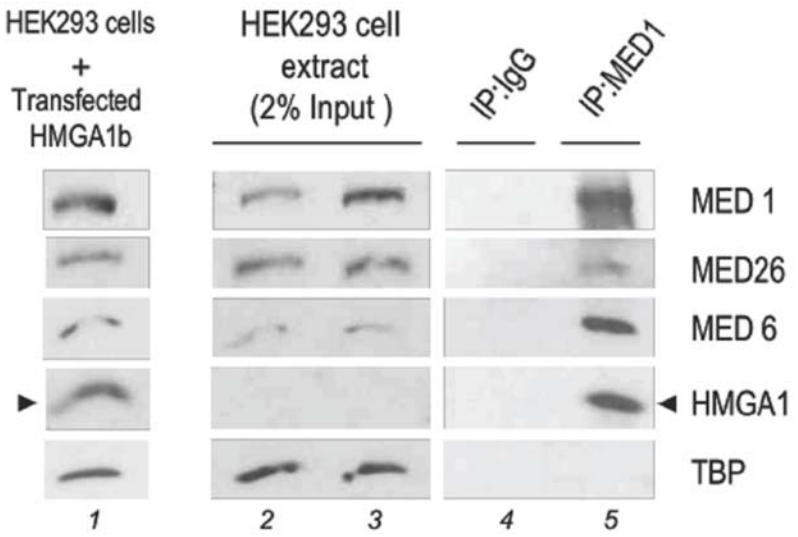



**Figure 2.2.4** Mediator interacts with endogenous HMGA1 in HEK293 whole cell lysate.

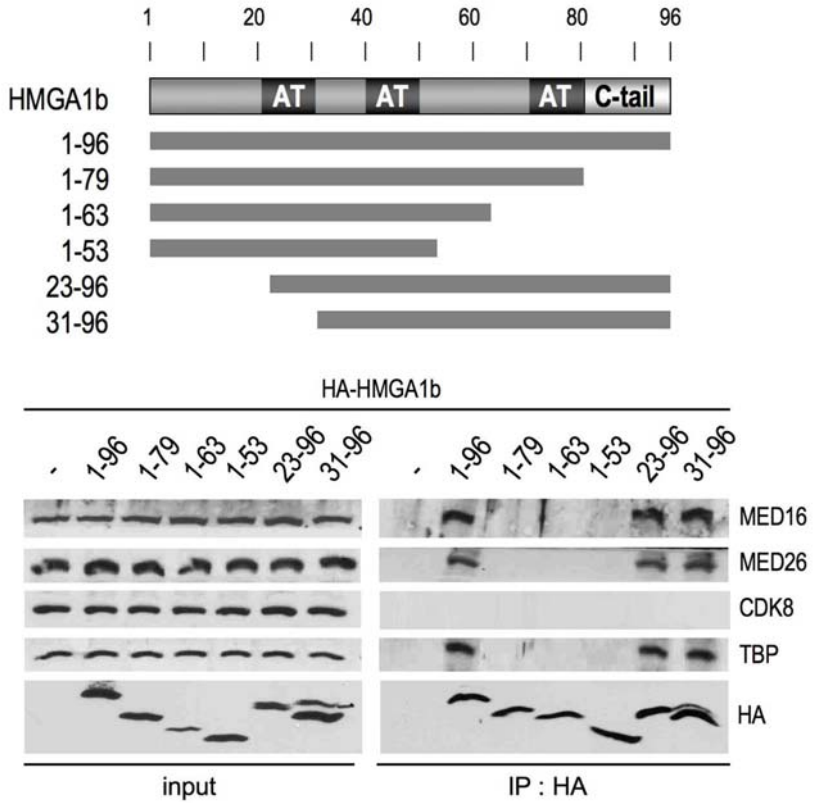



**Figure 2.2.5** Acidic COOH-tail is required for HMGA1 interaction with TFIID and Mediator (by Priyanka Sharma).

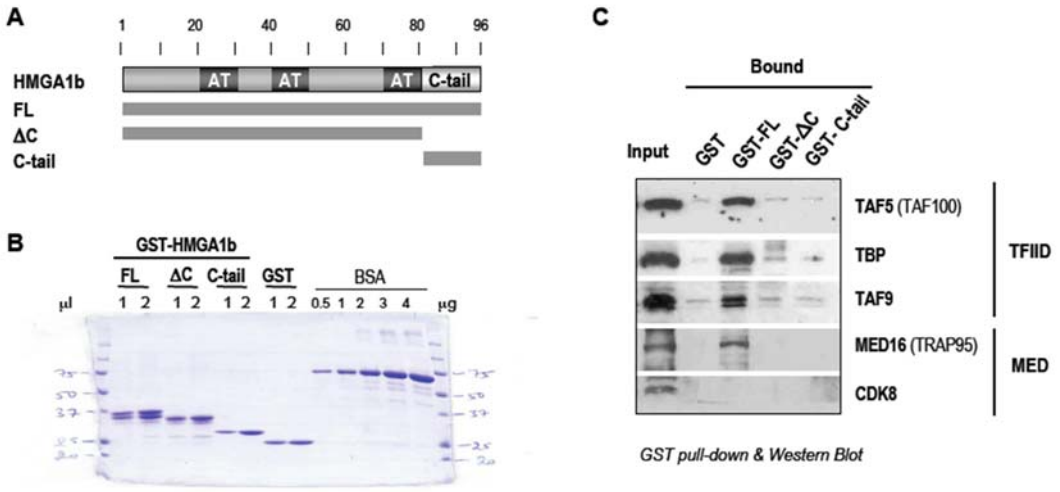



**Figure 2.2.6** HMGA1 interaction with TFIID and Mediator *in vitro* requires both the $NH_2$-terminal AT-hook domains and the acidic COOH-tail (by Priyanka Sharma).

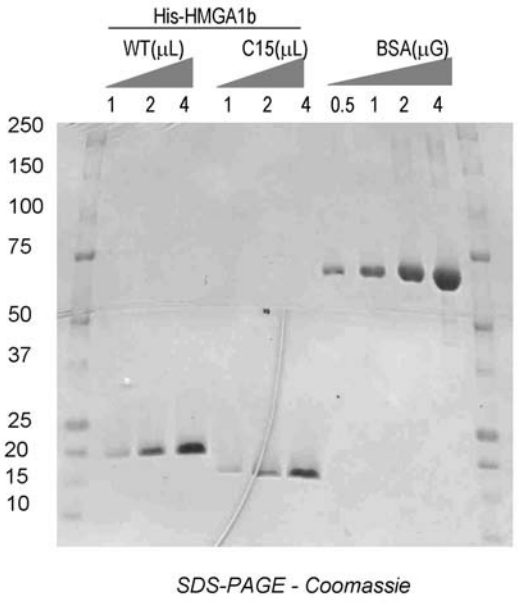



**Figure 2.2.7** SDS-PAGE analysis of recombinant HMGA1 wildtype (WT) and 15 COOH-terminal amino acids deletion mutant (ΔC).

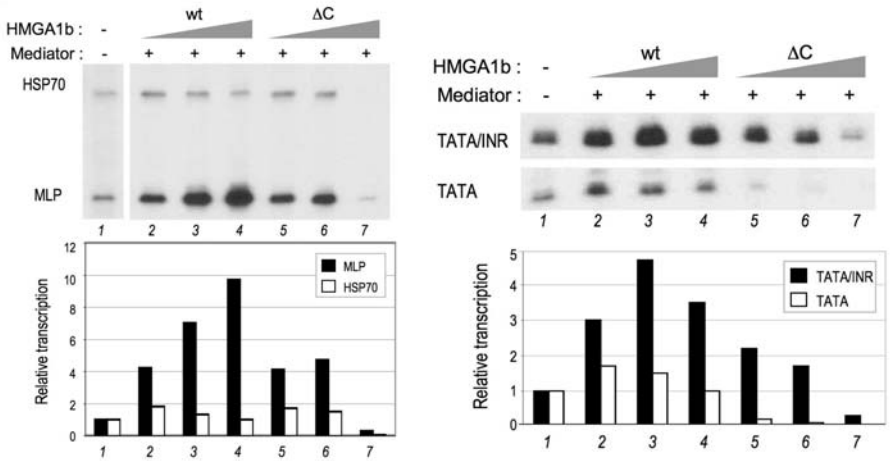



**Figure 2.2.8** HMGA1 acidic COOH-tail is required for HMGA1 to support maximal TATA/INR synergy.

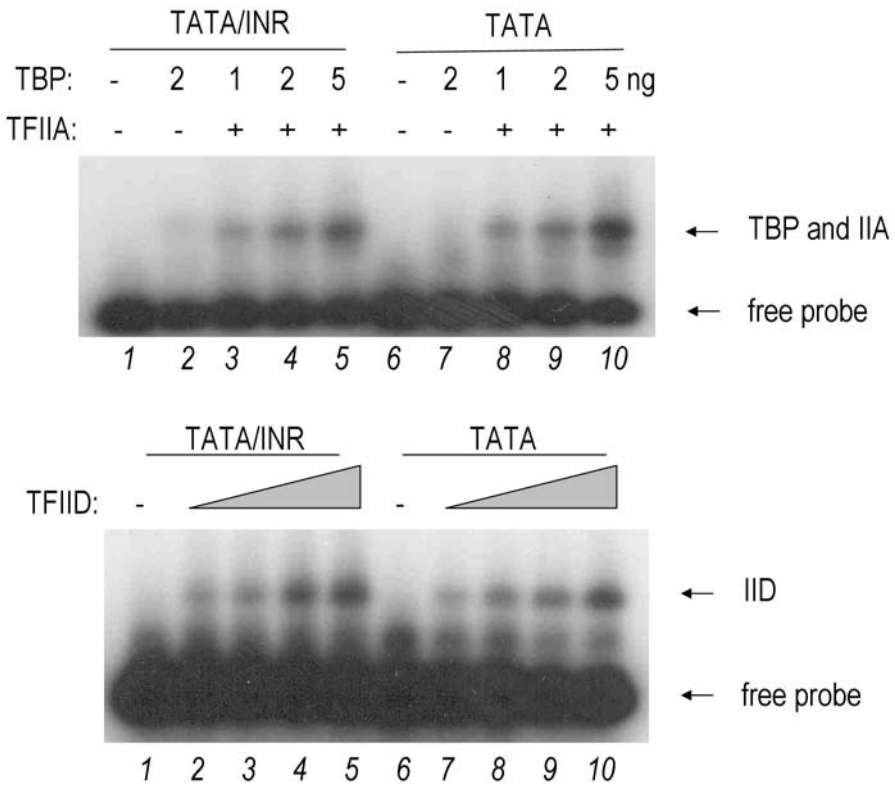



**Figure 2.2.9** TBP-TFIIA and TFIID bind TATA/INR and TATA core promoters in a similar pattern.

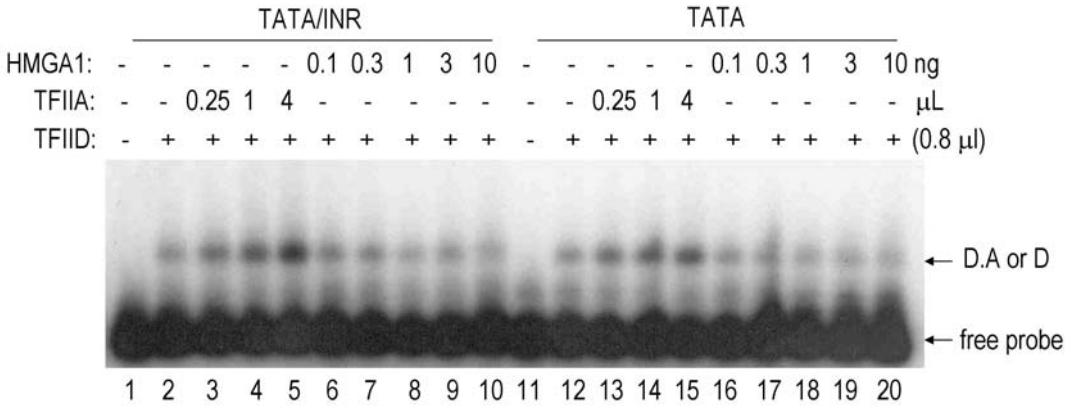



**Figure 2.2.10** HMGA1 does not facilitate TFIID binding to either TATA/INR or TATA core promoter in the absence of TFIIA.

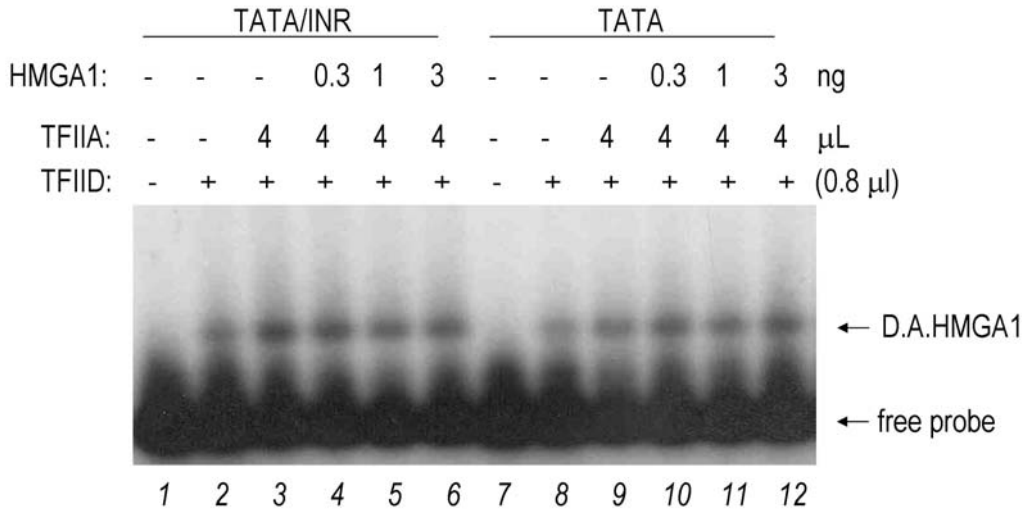



**Figure 2.2.11** HMGA1 can not facilitate TFIID-IIA binding to core promoters.

## Chapter 3

# Analysis of the core promoter preference of activators/activating sequences: a novel TATA-specificity by the *ACTB* gene upstream activating sequences

**Abstract**

Gene-specific activators possess both a DNA-binding domain (DBD) and a transcription activation domain. Activation domains were originally classified as acidic and Proline, Glutamine, Serine/Threonine, Isoleucine-rich groups. This chapter investigates the core promoter element preference of different Gal4-fusion activation domains, including the Proline-rich domain from CTF/NF1, the Glutamine-rich domains from Sp1 and Oct2, and the acidic domain from VP16 in the context of TATA, TATA/INR or INR core promoters. The results demonstrate that the core promoters regulate transcription from all Gal4-fusion activators. Gal4-Pro best supports synergistic activation from a TATA-box and an INR (highest synergy factor), followed by Gal4-Sp1, Gal4-Oct2 and Gal4-VP16 (strongest activation). In addition, this chapter also shows that YEATS2, a subunit of the newly identified ATAC complex, interacts directly with TBP and represses both TATA and TATA/INR core promoter directed transcription *in vitro*. Interestingly, mutagenesis and core promoter switching studies show for the first time that the *ACTB* proximal promoter possesses preference for TATA-box but not INR, providing the second example of a strict core promoter preference of proximal promoter in gene regulation.

## Introduction

It has been reported that the expression of specific gene depends not only on the upstream activators but also on the core promoter elements (Garrway et al., 1996; Bluter and Kadonaga 2001; Smale and Kadonaga 2003). The results from Chapter 2 together with research from other laboratories suggest that core promoter-specific coregulators such as HMGA1, NC2, CK2 and Mediator contribute to the enhancer-core promoter specificity in transcriptional activation (Willy et al., 2000; Levis et al., 2005; Chapter 2). Thus, the expression of specific gene might involve the combinatorial regulation by specific activators, coregulators and core promoters.

Transcriptional activation of the acidic trans-activation domain from VP16, the Q-rich domain of Oct2, and the P-rich domain of CTF/NF1 has been intensively investigated. Results demonstrate that VP16 strongly activates transcription when bound proximal as well as distal promoter DNA sequences; while Oct2 and CTF/NF1 only stimulate transcription when they bind proximal promoter DNA (Remacle et al., 1997). VP16 strongly activates transcription in yeast as well as in human systems, suggesting that the activation mechanism is conserved through out evolution (Babb et al., 2001). Activators recruit co-activators (e.g. TFIID, Mediator and SAGA), which in turn recruit general transcription machinery to the core promoter (Martinez 2002; Sohail and Roeder 2010).

Among the few studies of core promoter function in gene regulation by activators, VP16 and Sp1 have been investigated previously on different core promoters containing a TATA-box, TATA/INR or an INR (These core promoters contain different flanking sequences from the TATA, TATA/INR and INR core promoters used in this study). The results showed that VP16 strongly activated all core promoters and TATA/INR synergy. In contrast, Sp1 activated much weaker TATA/INR synergy. Interestingly, two domains of Sp1 possess preference for INR (Emami et al., 1995).

The myoglobin enhancer selectively activated its own consensus TATA-box but not the TATA-box from the *Simian virus 40* early promoter (Wafald et al., 1990). The murine lymphocyte-specific TdT promoter upstream activating sequences also stimulated transcription exclusively from an INR (Smale and Baltimore 1989; Garraway et al., 1996). Elf-1, which binds the DNA sequences 60 bp upstream of the TdT promoter transcription start site, was later reported to be the potential INR-specific activator (Ernst et al., 1996). The enhancer preference of the TATA/INR or INR/DPE core promoters has been analyzed by the Kadonaga laboratory (Bluter and Kadonaga 2001). However, it is still unclear whether activating sequences and cognate activators exist that have exclusive preference for TATA-box. Thus, the *ACTB* promoter has emerged as a good candidate to test the strict TATA-box preference because *ACTB* promoter contains a TATA-box and *ACTB* expression was not affected by a HMGA1 knockdown or knockout (Fig 2.1.4; Fig 2.1.5; Martinez-Hoyos et al., 2004).

The *ACTB* proximal promoter contains two proximal activating elements, the CCAAT box and the CCArGG box, which bind NF-Y and $p67^{SRF}$ (SRF), respectively. Mutagenesis studies demonstrate that both elements are required for *ACTB* transcription (Danalition et al., 1991). NF-Y (Gal4-NF-A, Gal4-NF-B and Gal4-NF-C trimeric factors) has been shown to activate both TATA and INR core promoters with no preference (Silvio et al., 1999). The core promoter preference of SRF has not been investigated.

Human ATAC complex was identified by Flag-GCN5 affinity purification. A 160-kDa protein was identified by mass spectrometry (LC-MS/MS) to be the "YNK7-ENL-AF9-TFIIF Small subunit" domain containing 2 (YEATS2) protein (Wang et al., 2008). Notably, the COOH-terminal of YEATS2 also contains a histone fold domain with 41% similarity (24% identity) to that of the NC2α/DRAP1 subunit of NC2 (Meisterernst and Roeder 1991; Wang et al., 2008).

Here I investigate the core promoter selectivities of the activation domains from VP16, Oct2, Sp1, CTF/NF1, the upstream *ACTB* activating sequences and the YEATS2 protein on different core promoters. Results demonstrate that Gal4-Pro supports the highest synergy factor from the TATA/INR core promoter and YEATS2 represses both TATA and TATA/INR core promoter directed transcription. Importantly, I also reveal a strict TATA-dependent activation by *ACTB* proximal promoter for the first time.

## Results

### Core promoter preference of different activation domains

In order to investigate the core promoter preference of different activation domains, the P-rich domain from CTF/NF1, the Q-rich domains from Sp1 and Oct2, and the acidic-rich domain from VP16 were fused to yeast Gal4-DBD (amino acids 1-147 from yeast Gal4 protein containing the DNA binding domain) to construct Gal4-fusion activators (Fig. 3.1). Five Gal4 binding sites were inserted into the upstream of TATA, TATA/INR and INR core promoters to construct luciferase reporters that could be used to analyze the core promoter preference of the Gal4-fusion activators (Fig. 3.2). Consistent with previous observation (Emami et al., 1995), Gal4-VP16 activates all three core promoters much stronger than the other tested activators (Fig. 3.3), but the synergy factor supported by Gal4-Pro is higher than that supported by the other tested activators, especially Gal4-VP16 (Fig. 3.3, synergy factor for Gal4-Pro (1.9) vs for Gal4-VP16 (1.2), synergy factor= (TATA/INR)/(TATA+INR), see legend). I note that the seven-fold basal transcription strength differences between the TATA/INR and TATA core promoters (Fig. 2.1.2) decrease to two-three fold when they are activated by Gal4-DBD (Fig 3.3, Gal4-DBD). Together these results show that core promoter regulates transcription from all the Gal4-fusion activators. Gal4-VP16 is the strongest activator on all three core promoters, but Gal4-Pro supports the highest synergy factor from TATA/INR core promoter.

**A novel TATA-specificity by *ACTB* upstream activating sequences**

The *ACTB* promoter contains a CCAAT box at -90 position, a CCArGG box at -60 position and a TATA-box at -30 position relative to the transcription start site (Fig. 3.4). The *ACTB* promoter directed transcription was not affected by a HMGA1 knockdown (Fig. 2.1.4, right panel). Initially, I intended to make the *ACTB* promoter HMGA1-responsive by mutating the sequences of the transcription start site region into the sequences of the TdT INR (Fig. 3.5, top). Unexpectedly, the INR mutation did not further increase the transcription from the *ACTB* promoter (Fig. 3.5, wildtype *ACTB* vs *ACTB*-Initiator), indicating that the *ACTB* upstream activating sequences do not activate through an INR or that there is already a functional INR in the *ACTB* promoter. Subsequently, the five Gal4 binding sites upstream of the TATA, TATA/INR and INR core promoters were replaced by the upstream activating sequences of *ACTB* promoter from -120 bp to -40 bp to construct pACTB-TATA, pACTB-TATA/INR and pACTB-INR luciferase reporters (Fig. 3.6, top). The results show that *ACTB* activating sequences activate through TATA and TATA/INR core promoters in a similar fashion but can not activate through the INR core promoter in the absence of a TATA-box (Fig. 3.6, (TATA and TATA/INR) vs INR, more than 10-fold difference).

Because NF-Y can activate an INR in the absence of a TATA-box (Silvio et al., 1999), the possibility of the TATA-specificity by SRF was further analyzed. Experimentally characterized CCArGG box-containing Actin cytoskeleton and contractile SRF target genes

are listed in Table 1 of the review by Miano JM et al., 2006. SRF target genes are sorted based on their core promoter sturcture defined by the bioinformatic study at Jin V et al., 2006 (Table 3.2). Chi square test shows that Actin cytoskeleton and contractile SRF target genes have a high tendency to possess a TATA-box as core promoter (Table 3.1, 37 out of 97 TATA-containing SRF target genes vs 1483 out of 9010 TATA-containing genes, $p < 0.001$). The bioinformatic analysis suggests that SRF could be the potential TATA-specific activator.

Finally, the requirement of HMGA1 for Ga4-Pro, Gal4-Oct2 and the *ACTB* upstream sequences activated transcription was investigated using a HMGA1 knockdown in HEK293 cells (Subchapter 2.2.1, Materials and Methods). The results show that HMGA1 is required for Gal4-Pro activated transcription but not for Gal4-Oct2 and the *ACTB* upstream activating sequences (Fig. 3.7). In summary, these results demonstrate that HMGA1 may function as a novel coactivator for Gal4-Pro and perhaps other P-rich activation domains. Futhermore, the activation of the *ACTB* proximal activating sequences depends on TATA-box but not INR, providing a novel example of core promoter functioning in the combinatorial transcription regulation.

**YEATS2 interacts with TBP and represses both TATA and TATA/INR core promoter directed transcription**

YEATS2 was first identified as a subunit of the ATAC complex with a molecular weight of 160 kD (Wang et al., 2008). YEATS2 contains a YEATS-domain at the $NH_2$-terminal region, a P-rich domain in the middle and a histone fold domain at the COOH-terminal (Fig 3.9). Interestingly, the YEATS2 histone fold domain is required for NC2β to associate with the ATAC complex (Wang et al., 2008). NC2 (NC2α and NC2β) interacts with TBP (Kamada et al., 2001), thus the possible direct interaction between TBP and YEATS2-NC2β is analyzed by Flag-pull down using Flag-NC2β. Results show that YEATS2-NC2β module interacts with recombinant TBP similarly to NC2 (Fig. 3.8, lanes7, 8). Gal4-YEATS2 is subsequently shown to repress transcription from G5-TK-Luc reporter in cells but Gal4-YEATS2-ΔC lacking the histone fold domain does not repress transcription (Fig. 3.9). Finally, Gal4-YEATS2 and Gal-YEATS-ΔC are affinity-purified (Flag-tag) from whole cell lysate and analyzed by *in vitro* transcription assay. Compared with a mock purified sample, Gal4-YEATS2 represses transcription from both G5-TK-Luc (TATA-only) and G5-TATA/INR templates, while Gal4-YEAST2-ΔC does not repress transcription (Fig. 3.10, WT vs ΔC). These results show that the histone fold domain of YEATS2 is required to repress transcription and YEATS2 might not be a core promoter-specific repressor.

## Discussion

It is known that core promoter plays an important role in gene regulation and the core promoter preference of certain activators such as VP16 and Sp1 have been studied before (Emami et al., 1995; Smale 2001; Smale and Kadonaga 2003). Here I show again that Gal4-VP16 strongly activates various core promoters, but the transcriptional synergy between the TATA-box and the INR is much weaker in this study (Fig. 3.3). One possible explanation is that the VP16 activation domains analyzed are different (413-454 amino acids in Emami et al., 1995 vs 411-490 amino acids in this study). The other possible reason is that the flanking sequences between the TATA-box and the INR are also different. In addition, I show for the first time that the activated transcription by Gal4-Pro and Gal4-Oct2 depends not only on the activation domains but also the core promoters (Fig 3.3). Notably, Gal4-Pro supports the highest TATA/INR synergy factor (Fig 3.3) and HMGA1 is required for Gal4-Pro activated TATA/INR transcription (most synergy) but not for Gal4-Oct2 (less synergy) and the *ACTB* proximal promoter (no synergy) (Fig 3.7). These results are consistent with results that HMGA1 functions as the TATA/INR-specific coregulator (Chapter 2), suggesting again the important role of core promoter-specific coregulator in gene regulation.

In contrast to Gal4-fusion activators, the *ACTB* upstream sequences demonstrate a strict TATA-specificity (Fig. 3.5; 3.6). The TATA-preference by c-fos was described before, however, when a TATA was replaced by an INR, c-fos also activated through an INR (Metz et

al., 1994). Thus, to our knowledge, this is the first time that reveals a strict TATA-sepcificity by the *ACTB* proximal promoter. In contrast, the strict INR-specificity by the murine lymphocyte-specific TdT promoter upstream activating sequences has been discovered (Smale and Baltimore 1989; Garraway et al., 1996). These studies of the selective communication between the core promoter elements and the upstream activating sequences demonstrate that the core promoter is not solely present to bind the basal transcription machinery, but also to regulate gene expression. Considering the diversity of the core promote elements (TATA, INR, DPE, MTE, DCE, BRE, and so on), the contribution of core promoter in gene regulation might be much greater than we thought.

Core promoter structure analyses of the SRF target gene promoters show that those gene promoters have a high tendency to possess a TATA-box as the core promoter (Table 3.1 and 3.2), suggesting that SRF might be TATA-specific activator. Importantly, some of the SRF-target genes contain more than one core promoters, which are all analyzed to avoid any bias (Table 3.2, e.g. *DMPK*, *FNBP3*). Another possibility is that neither NF-Y nor SRF is TATA-specific, but that the DNA-protein complex formed by NF-Y together with SRF activates preferentially through a TATA-box.

NC2 is considered to be a general repressor of TATA-directed transcription (Meisterernst and Roeder 1991; Kamada et al., 2001; Malecova et al., 2007). The crystal structure of NC2 demonstrates that NC2 blocks TFIIB and TFIIA binding TBP (Kamada et al.,

2001). Interestingly, YEATS2-NC2β module also interacts physically with TBP (Fig. 3.8). It will be interesting to crystalize YEATS2-NC2β and TBP on the core promoter to compare the structure with that formed by NC2-TBP-core prmoter. Functionally, YEATS2-NC2β represses TATA promoter directed transcription both *in vitro* and *in vivo* (Fig. 3.9 and Fig. 3.10). In contrast, GCN5, the other subunit of the ATAC complex, activates transcription through its "Histone Acetyl-transferase" activity. These functions suggest that the ATAC complex is a multi-functional complex that can activate as well as repress transcription *in vivo*. The SAGA complex directs transcription from the TATA genes but not the "TATA-less" genes in yeast (Rhee and Pugh 2012). The ATAC complex may also function as a core promoter-specific coregulator *in vivo*.

In summary, I demonstate that activators, core promoter-specific coregulators and the core promoters function together to regulate transcription. The initiation of gene-specific transcription might be a combinatorial effects from all of the three.

## Materials and Methods

### Plasmid constructs

The Gal4-DBD (1-147), Gal4-Pro (399-499), Gal4-Sp1 (132-243), Gal4-Oct2 (99-161) and Gal4-VP16 (411-490) are generous gifts from Dr. Yves Dusserre, Switzerland. The TATA/INR plasmid pG5TdT(-41TATA/+33), the TATA plasmid pG5TdT(-41TATA/Inr+33) and the INR plasmid pG5TdT(-41/Inr+33) have been described previously (Martinez et al., 1998). The pG5-TATA/INR-Luc, pG5-TATA-Luc and pG5-INR-Luc reporters were generated by cloning the Hind III fragments containing 5 Gal4-binding sites together with the core promoters from, respectively, TATA/INR, TATA and INR plasmids into the Hind III site of pGL3-Basic vector (Promega). The wild-type ACTB-Luc was described in the Materials and Methods of Subchapter 2.1. The ACTB-initiator was generated by Quick-change PCR from wild-type *ACTB* (primer: 5'-GGCGGCGCGACGCGCCCTCATTCTCGAGACCGCGTCCGCCCC-3'). The pACTB-TATA-Luc, pACTB-TATA/INR-Luc and pACTB-INR-Luc were constructed by inserting *ACTB* promoter activator sequences (-120 bp to -40 bp) into the Kpn I site of TATA-Luc, TATA/INR-Luc and INR-Luc, respectively. The primers with KpnI to PCR amplify -120 bp to -40 bp of *ACTB* upstream activating sequences are: Forward: 5'-TCTAGTggtaccGCGAAGCCGGTGAGTGAGCG-3'; Reverse: 5'-TACATAggtaccGCGGCCGCTCGAGCCATAAAAGGC-3'.

**Recombinant protein purifications and *in vitro* pull-down Assays**

Recombinant proteins were expressed and purified as described in Materials and Methods of Chapter 2. For TBP interaction assays (Fig 3.8), purified Flag-NC2α/His-NC2β or Flag-NC2β/His-YEATS2-HFD complexes were diluted 30-fold in BC100 containing 500 ng/μl bovine serum albumin (to reduce Flag peptide concentration) and immobilized on M2 resin for 15 h at 4 °C. The bound complexes were then washed twice with BC100 and incubated with purified TBP for 3 h at 4 °C. The resins were then washed extensively with BC100, eluted with 1X SDS-loading buffer, and analyzed by SDS-PAGE and Western blotting.

**Statistics analysis**

Chi square test refers to the "SISA" statistics website. Url: http://www.quantitativeskills.com/sisa/statistics/twoby2.htm. Chi square test: Row 1: 1483, 7527; Row 2: 37, 60. The degree of freedom is 1 and the p- value= $2.7E^{-8}$.

**Cell culture, Luciferase assay, RNA interference (RNAi) analysis and *in vitro* transcription assay**

Described in the Materials and Methods at Chapter 2 and Subchapter 2.1.

**Figure legends**

**Figure 3.1** Schematic diagram of the relative position of the ativation domains inside the activators and the Gal4-fusion activators. **(A)** The locations of Pro-domain (399-499) in CTF/NF1, Q1-domain (132-243) in Sp1, Q-domain (99-161) in Oct2 and the acidic-rich domain (411-490) in VP16, indicated by different colors. The DBD for CTF/NF1 is CCAAT box binding domain, for Sp1 is the $Zn^{+}$-finger domain, for Oct2 is the POU domain and for VP16 is the core domain. **(B)** Schematic diagram of Gal4-DBD (1-147), Gal4-Pro (399-499), Gal4-Sp1 (132-243), Gal4-Oct2 (99-161) and Gal4-VP16 (411-490). The activtion domains are colored the same way as in panel A.

**Figure 3.2** Schematic diagram of the reporters containing five Gal4 binding sites. The five Gal4-binding sites are colored in gray boxes upstream of the TATA-box.

**Figure 3.3** Analysis of Gal4-fusion activators on core promoter. The relative luciferase activity from pG5-TATA-Luc (blue), pG5-TATA/INR-Luc (black) and pG5-INR-Luc (white) activated by Gal4-DBD, Gal4-Pro, Gal4-Sp1, Gal4-Oct2 and Gal4-VP16 are normalized to pGL3 basic luciferase activity. β-Galactosidase activity is the internal control for transfection efficiency. Three independent triplicate experiments are averaged. The luciferase activity from pGL3 basic vector is arbitrarily set to 1. The number on the Y-axis represents luciferase activity relative to the activity from pGL3. Gal4-VP16 is separated in the right panel because

the activation effect is too strong to put in same graph together with the other activators. The synergy factors ([TATA/INR]/([TATA] + [INR]) ) for Gal4, Gal4-Pro, Gal4-Sp1, Gal4-Oct2 and Gal4-VP16 activated transcriptions are 1.1, 1.9, 1.7, 1.6 and 1.2, respectively.

**Figure 3.4** The *ACTB* promoter sequences analysis. The known functional elements are underlined. The -90 CCAAT box is labeled with blue color, the -60 CCArGG box is labeled with green color and the -30 TATA-box is labeled with red color. NF-Y binds CCAAT box and SRF binds CCArGG box.

**Figure 3.5** Mutating the transcription start site region of the *ACTB* promoter into the INR sequences does not further stimulate *ACTB* transcription. Top, the schematic diagram of the *ACTB* wildtype promoter and the *ACTB*-Initiator consturct. The sequences of INR are the same as the TdT INR and boxed. Bottom, the relative luciferase activitiy from three independent triplicate experiments. Luciferase activity from wildtype *ACTB* is arbitrarily set to 1 (blue color) and the luciferase activity from the *ACTB*-Initiator is normalized to wildtype *ACTB* luciferase activity.

**Figure 3.6** The *ACTB* proximal promoter functions through a TATA-box but not an INR. Top, the reporters pACTB-TATA-Luc, pACTB-TATA/INR-Luc and pACTB-INR-Luc are constructed by displacement of the five Gal4 binding sites from pG5-TATA-Luc, pG5-TATA/INR-Luc and pG5-INR-Luc reporters with the *ACTB* upstream activating

sequences from -120 bp to -40 bp. Bottom, the relative luciferase activity from pACTB-TATA-Luc (blue), pACTB-TATA/INR-Luc (black) and pACTB-INR-Luc (white) is averaged from two independent triplicate experiments. pACTB-TATA-directed luciferase activity is arbitrarily set to 1. Luciferase activity from pACTB-TATA/INR-Luc and pACTB-INR-Luc is normalized to the activity from pACTB-TATA-Luc.

**Figure 3.7** A HMGA1 knockdown decreases Gal4-Pro activated transcription but not Gal4-Oct2 or *ACTB* activating sequences. **(A)** Reporter G5-TATA/INR-Luc (co-transfected with either Gal4-Pro or Gal4-Oct2) or reporter pACTB-TATA/INR-Luc is co-transfected with either control siRNA (black) or HMGA1 siRNA (white) into HEK293 cells. Relative luciferase activity from samples co-transfected with control siRNA is arbitrarily set to 1. Two independent triplicate experiments are averaged for Gal4-Pro activated luciferase activity. One duplicate experiment with two samples is averaged for Gal4-Oct2 activated luciferase activity. Three independent triplicate experiments are averaged from pACTB-TATA/INR-Luc reporter luciferase activity. **(B)** Relative luciferase activity from Gal4-Pro activated G5-TATA/INR-Luc or G5-TATA-Luc reporters co-transfected with either control siRNA (black) or HMGA1 siRNA (white). One triplicate experiment is averaged.

**Figure 3.8** YEATS2-NC2β module directly interacts with TBP. *In vitro* Flag-pull down experiments are performed with the purified recombinant 6His-TBP and the indicated recombinant complexes (His-NC2α-Flag-NC2β or His-YEATS-HFD-Flag-NC2β)

immobilized on M2-agarose resin. Flag-tag bound proteins are pulled down by M2 agarose resin and analyzed by SDS-PAGE and Western blotting with the specific antibodies as indicated. Lanes 1-2: TBP input (7 ng, 20 ng); lanes 3-4 : His-NC2α-Flag-NC2β input; lanes 5-6: His-YEATS-HFD-Flag-NC2β input; lanes 7-9: Flag-pull down using His-NC2α-Flag-NC2β, His-YEATS-HFD-Flag-NC2β or resin-only, respectively.

**Figure 3.9** Gal4-YEATS2 represses TK promoter transcription. Top, the diagram of Gal4-DBD, Gal4-YEATS2, Gal4-YEATS2-ΔC and the reporter G5-TK-Luc. Bottom, relative luciferase activity from G5-TK-Luc co-transfected either with Gal4-DBD, Gal4-YEATS2 or Gal4-YEATS2-ΔC, respectively, is averaged from three independent triplicate experiments. G5-TK luciferase activity co-transfected with Gal4-YEATS2-ΔC is set to arbitrarily 1.

**Figure 3.10** Gal4-YEATS2-ΔC lacking histone fold domain does not repress transcription *in vitro*. *In vitro* transcription experiments are performed with either nuclear extract or purified system (Chapter 2, Materials and Methods). GAL4-Flag-YEATS2 or Gal4-Flag-YEATS2-ΔC complexes are purified with Flag-affinity from HEK293 whole cell lysates. The indicated promoter templates G5-TK (TATA-type) and G5-TATA/INR are used in the transcription assay.

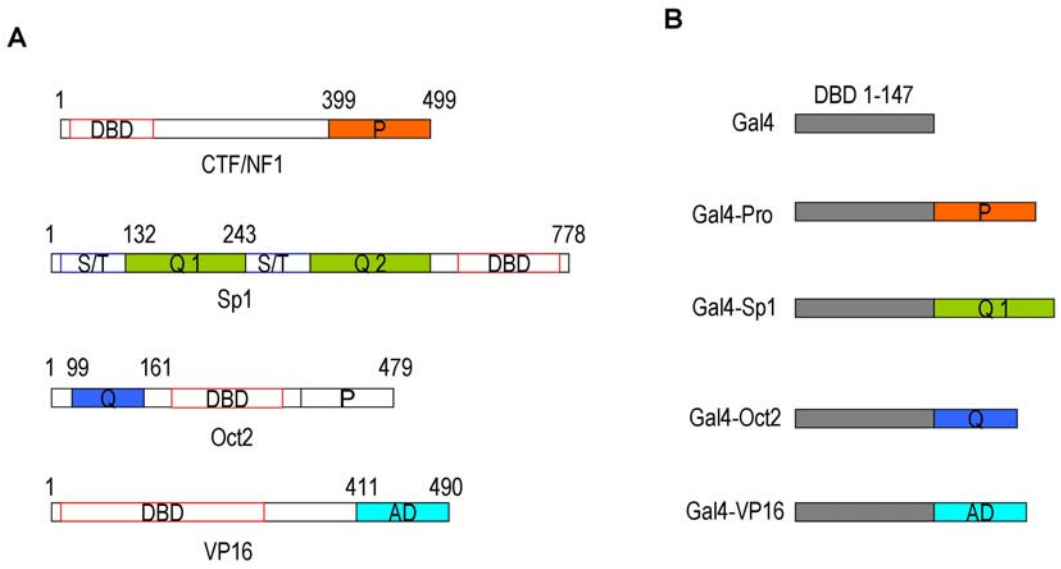



**Figure 3.1** Schematic diagram of the relative position of the ativation domains inside the activators and the Gal4-fusion activators.

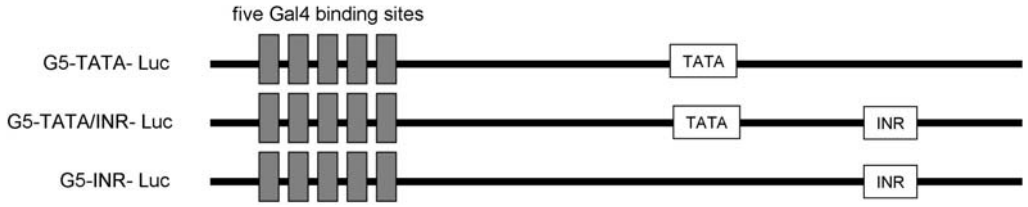



**Figure 3.2** Schematic diagram of the reporters containing five Gal4 binding sites.

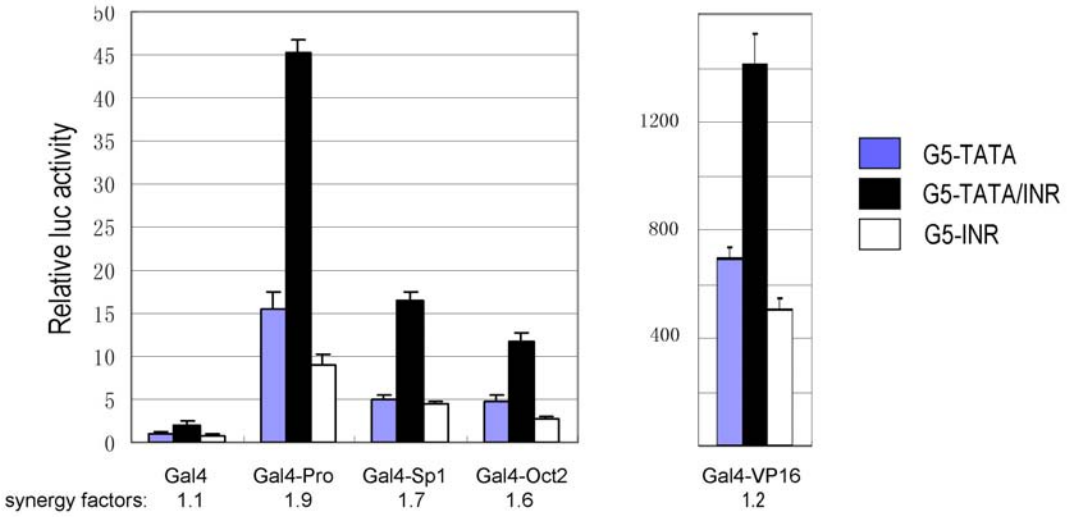



**Figure 3.3** Analysis of Gal4-fusion activators on core promoter.

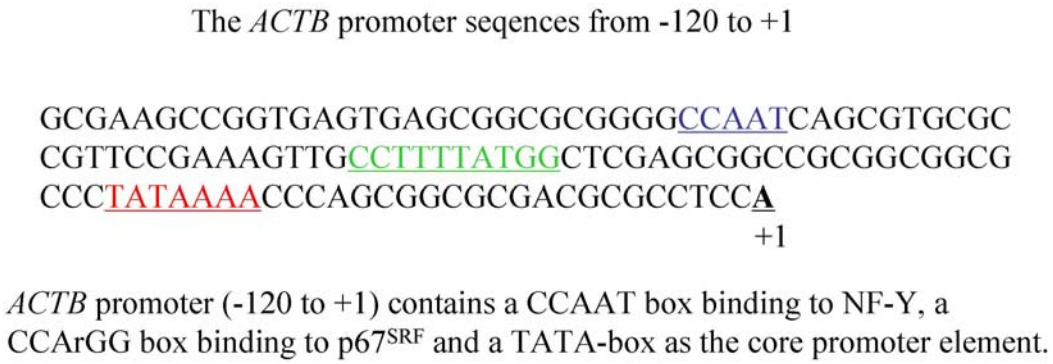



**Figure 3.4** The *ACTB* promoter sequences analysis.

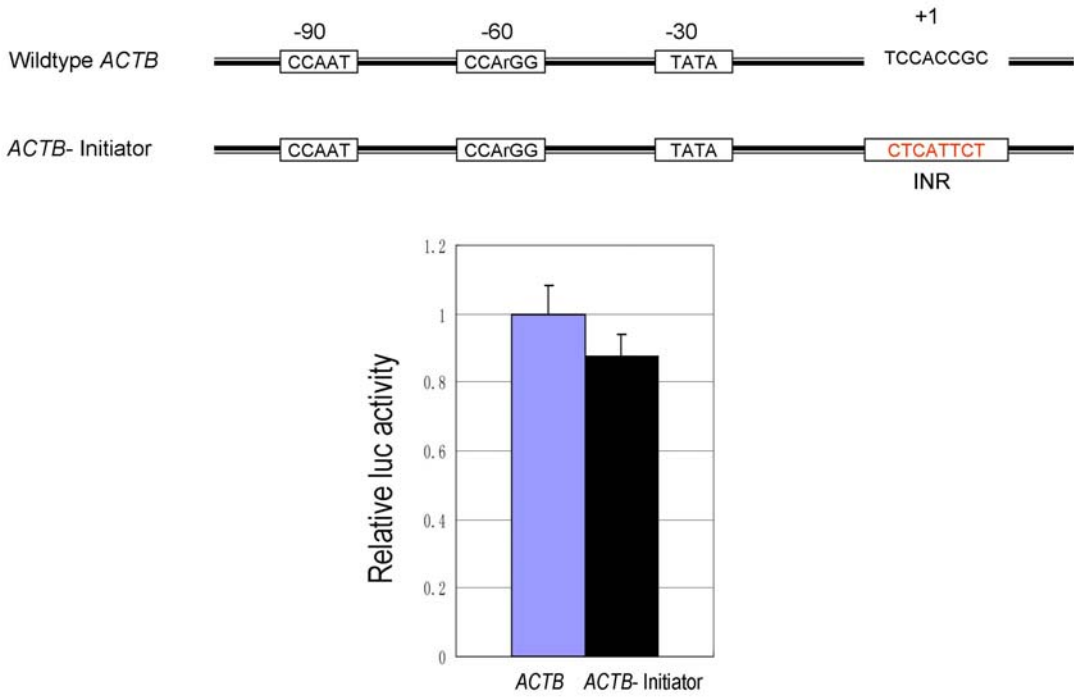



**Figure 3.5** Mutating the transcription start site region of the *ACTB* promoter into the INR sequences does not further stimulate *ACTB* transcription.

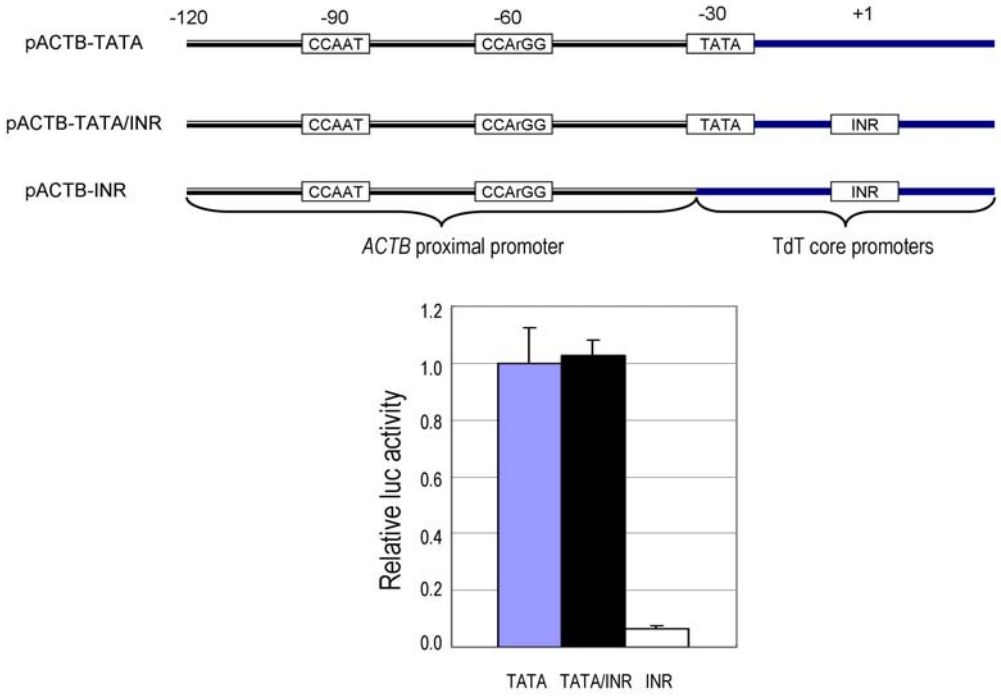



**Figure 3.6** The *ACTB* proximal promoter functions through a TATA-box but not an INR.

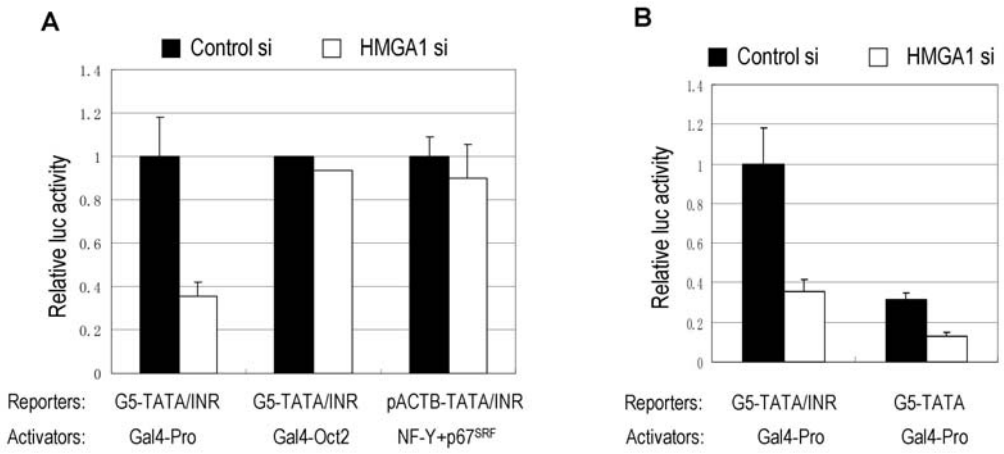



**Figure 3.7** A HMGA1 knockdown decreases Gal4-Pro activated transcription but not Gal4-Oct2 or *ACTB* activating sequences.

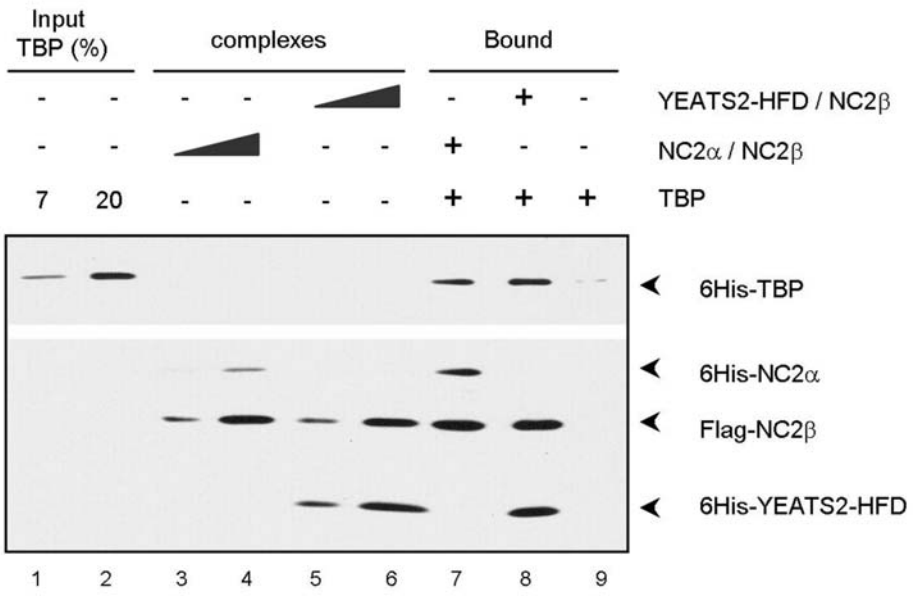



**Figure 3.8** YEATS2-NC2β module directly interacts with TBP (by Yuan-Liang Wang).

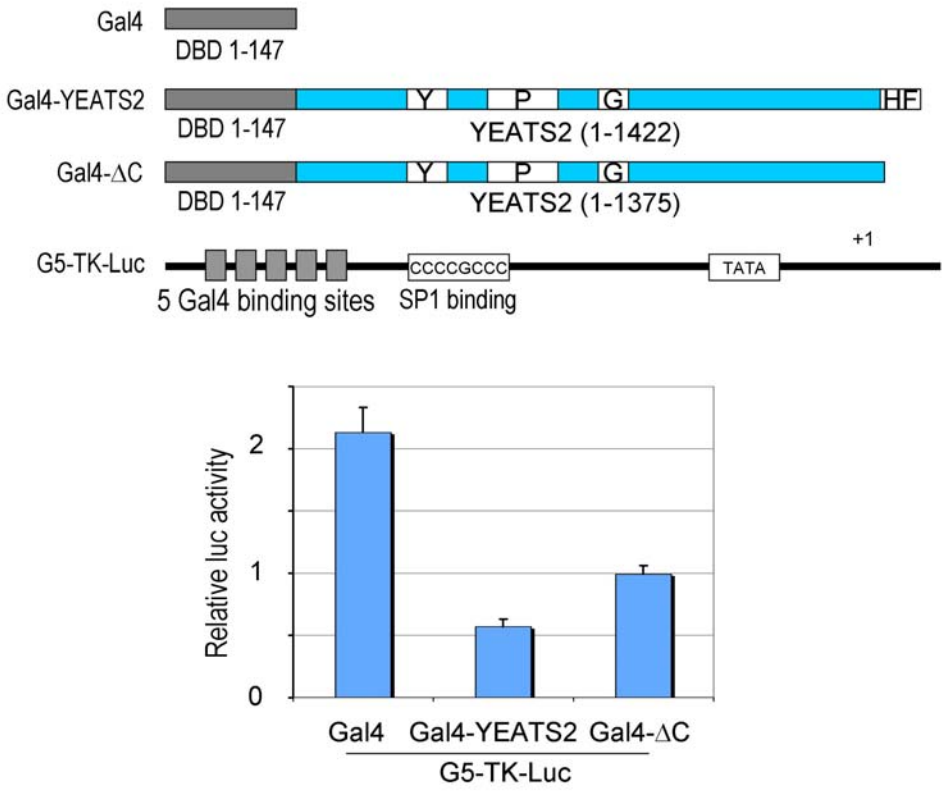



**Figure 3.9** Gal4-YEATS2 represses TK promoter transcription (by Francesco Faiola).

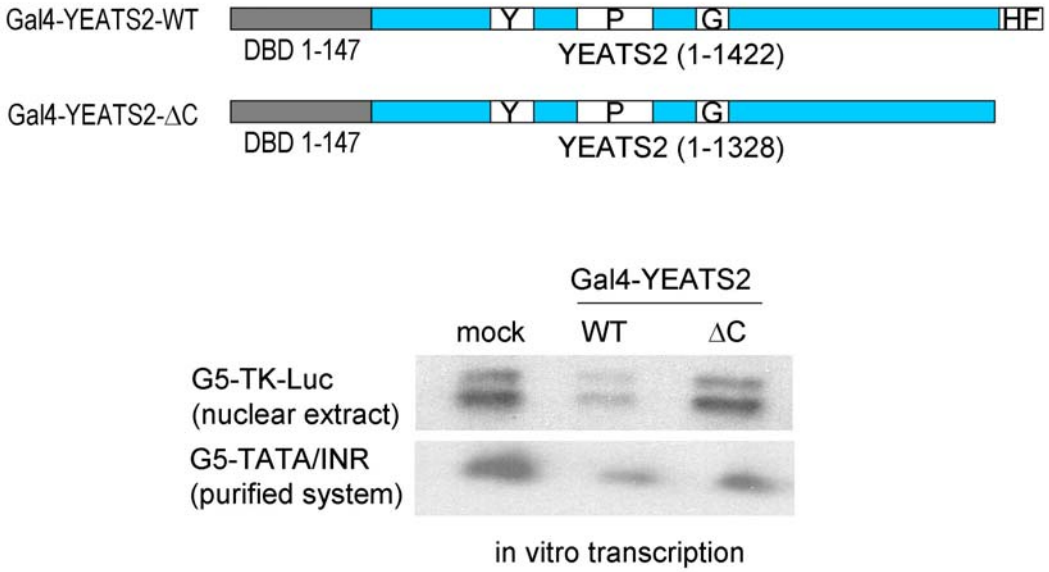



**Figure 3.10** Gal4-YEATS2-ΔC lacking histone fold domain dose not repress transcription *in vitro*.

| Core promoter type | All | TATA-containing | TATA-less |
|---|---|---|---|
| Genes Characterized[a] | 9010 (100%) | 1483 (16.5%) | 7527 (83.5%) |
| SRF target genes[b] | 97 (100%) | 37 (38%) | 60 (62%) |

Comparisons of TATA-box distributions in SRF target genes with its global distributions. Chi square test shows they are significant different ($p<0.001$, $p= 2.7E^{-8}$).

a. Frequencies of TATA and TATA-less type core promoters among 9010 human and mouse orthologous genes with experimentally validated TSS (Jin V et al., 2006);

b. Frequencies of TATA-box among cytoskeleton-contractile SRF target genes that were validated at least through two methods including ChIP, EMSA, luciferase assay, SRF knock down with RT-PCR or expression profiling (Table 1, Miano JM et al., 2006).

**Table 3.1** TATA-box is preferentially present in cytoskeleton-contractile SRF target genes

| Symbol | ID, Accession | Structure | Symbol | ID, Accession | Structure |
|---|---|---|---|---|---|
| *ACTB* | 60, M10277 | TATA | *ACTG1* | 71, M19283 | TATA/INR |
| *CSRP1* | 1465, M33146 | | *CFL1* | 1072, X95404 | |
| *CSRP2* | 1466, U95017 | | *KRT17* | 3872, AK095342 | |
| *DES* | 1674, AK097038 | | *TPM1* | 7168, CR610855 | |
| *TTN* | 7273, NM_133379 | | *AKAP12* | 9590, NM_005100 | |
| *DMD* | 1756, AF213401 | | *ITGB1BP2* | 26548, NM_012278 | |
| *DTNA* | 1837, NM_001392 | | *SDC4* | 6385, D79206 | |
| *TLN1* | 7094, AB028950 | | *ACTA2* | 59, CR597033 | |
| *VIL1* | 7429, A07400 | | *ACTG2* | 72, NM_001615 | |
| *ACTA1* | 58, NM_001100 | | *MYH6* | 4624, Z20656 | |
| *MYH11* | 4629, BC031040 | | *MYH7* | 4625, NM_000257 | |
| *TPM2* | 7169, NM_003289 | | *MYL3* | 4634, M76408 | |
| *CA3* | 761, NM_005181 | | *MYL4* | 4635, M37069 | |
| *CALB1* | 793, BC020864 | | *TNNT2* | 7139, NM_000364 | |
| *NPPA* | 4878, BC005893 | | *TPM2* | 7169, CR615839 | |
| *ANXA2* | 302, NM_001002858 | | *SLC8A1* | 6546, Y12885 | |
| *KRT14* | 3861, CR613487 | | *MSN* | 4478, NM_002444 | |
| *KRT19* | 3880, NM_002276 | | | | |
| *KRT7* | 3855, AF509891 | | | | |
| *VASP* | 7408, Z46389 | | | | |
| Symbol | ID, Accession | Structure | Symbol | ID, Accession | Structure |
| *ACTR3* | 10096, NM_005721 | INR | *CFL1* | 1072, AK097690 | None |
| *CAPZA3* | 93661, BC016745 | | *CNN1* | 1264, D85611 | |
| *CNN1* | 1264, NM_001299 | | *DSTN* | 11034, CR591834 | |
| *CORO1A* | 11151, NM_007074 | | *GSN* | 2934, AK096280 | |
| *FLNC* | 2318, NM_001458 | | *HSPB7* | 27129, NM_014424 | |
| *HSPB7* | 27129, AK057295 | | *SNX2* | 6643, NM_003100 | |
| *PDLIM5* | 10611, BC017902 | | *TRIP6* | 7205, AF093834 | |
| Continue in next page | | | | | |

**Table 3.2** Classification of the core promoters of Cytoskeleton-contractile SRF target genes.

| Symbol | ID, Accession | Structure | Symbol | ID, Accession | Structure |
|---|---|---|---|---|---|
| *PFN1* | 5216, CR593169 | INR | *DTNA* | 1837, NM_032978 | None |
| *AOC3* | 8639, NM_003734 | | *TJP1* | 7082, NM_003257 | |
| *DMD* | 1756, M84768 | | *VCL* | 7414, M33308 | |
| *ITGA5* | 3678, NM_002205 | | *MYL3* | 4634, M24242 | |
| *SDC2* | 6383, BC030133 | | *TNNC1* | 7134, NM_003280 | |
| *TJP1* | 7082, NM_003257 | | *TNNT2* | 7139, X79859 | |
| *CRYAB* | 1410, NM_001885 | | *ATP2A2* | 488, M23114 | |
| *MYH6* | 4624, X56181 | | *ATP2A2* | 488, M23115 | |
| *CA3* | 761, AJ006473 | | *BIN1* | 274, U83999 | |
| *CASQ1* | 844, NM_001231 | | *CKM* | 1158, NM_001824 | |
| *DMPK* | 1760, L00727 | | *CPT1B* | 1375, NM_004377 | |
| *FHL2* | 2274, BC012742 | | *CPT1B* | 1375, NM_004377 | |
| *MYOD1* | 4654, AF027148 | | *DMPK* | 1760, BC026328 | |
| *NKX2-5* | 1482, NM_004387 | | *DMPK* | 1760, L08835 | |
| *ARPC4* | 10093, AK024110 | | *NPPA* | 4878, AL021155 | |
| *FNBP3* | 55660, AF049528 | | *PLN* | 5350, AF177763 | |
| *FNBP3* | 55660, BC029414 | | *SLC2A1* | 6513, NM_006516 | |
| *JPH2* | 57158, AL132999 | | *ANLN* | 54443, NM_018685 | |
| *VASP* | 7408, Z46389 | | *ANXA3* | 306, CR601701 | |
| | | | *CDV1* | 28981, NM_031473 | |
| | | | *EPLIN* | 51474, AF245391 | |
| | | | *KRT19* | 3880, Y00503 | |
| | | | *PFN2* | 5217, CR615830 | |
| | | | *PFN2* | 5217, BC043646 | |
| | | | *RHOQ* | 23433, CR624597 | |
| | | | *TCAP* | 8557, NM_003673 | |
| | | | *TPM4* | 7171, BC037576 | |

**Table 3.2** Classification of the core promoters of Cytoskeleton-contractile SRF target genes.

# Chapter 4

# Conclusions

**Core promoter-specific function of HMGA1 and Mediator**

Historically, all core promoters were thought to contain a TATA-box as core promoter element (Smale and Kadonaga 2003). The sequencing of multiple genomes including human results in the discovery of new additional core promoter elements. Consequently, the emerging diversity of core promoter elements vanished the importance of the TATA-box (Smale and Kadonaga 2003; Yang et al., 2007). Early research suggests that GTFs together with RNA Pol II are sufficient to execute transcription from distinct core promoters. However, as a result of the identification of core promoter-specific factors, it is now apparent that additional coregulators are required to support transcription from different core promoters. For exmaple, NC2, a general repressor of TATA/TBP-directed transcription, supports the INR/DPE synergy in *Drosophila* cell free system and *in vivo* (Willy et al., 2000; Hsu et al., 2008). In contrast, PC4, Mediator and Casein Kinase 2 (CK2) support the INR/DPE synergy (instead of NC2) in mammalian cell free transcription system (Lewis et al., 2005). TIC1 and TIC2 fractions from Hela nuclear extract can support TATA/INR synergy and TATA-less core promoter transcription, respectively (Martinez et al., 1998). In this work, factors responsible for the TATA/INR synergy in the TIC1 fraction are further purified. HMGA1 and Mediator are indentified as the TATA/INR-specific factors in the presence of TAFs (Chapter 2 and summarized in Fig. 4.1), providing additional evidence that transcription from different core promoters involves specific coregulators (Martinez 2012).

The Adenovirus major late promoter (MLP), which contains a TATA, INR and DCE, has been proposed to wrap around several subunits of TFIID (Oelgeschlager et al., 1996). TAF1 and TAF2 bind INR (Chalkley and Verrijzer 1999); TAF6 and TAF9 interact with DPE (Burke and Kadonaga 1996; 1997). Furthermore, A histone octamer-like structure consisting of TAF6 (H4-like), TAF9 (H3-like) and TAF12 (H2A-H2B-like) has been uncovered in TFIID (Hoffmann et al., 1996). The histone-like TAFs complex has been shown to stabilize TAF6-TAF9-DPE binding (Shao et al., 2005). Therefore, the TFIID might adapt different conformation on different groups of core promoters.

HMGA1can facilitate IFN-beta enhanceosome formation through the interactions with NFκB as well as promoter DNA (Thanos and Maniatis 1992; Yie et al., 1999). HMGA1 can bind AT-rich DNA sequences (e.g. TATA-box and INR, Huth et al., 1997). HMGA1 can also compete with histone H1 binding nucleosomes, results in activation of transcription (Catez et al., 2004). Interestingly, in this study, HMGA1 is shown to interact with TFIID (Subchapter 2.2). Thus, HMGA1 might act as molecular glue that stabilizes the interaction of TFIID with DNA. In contrast, the interaction of TATA promoters with TFIID is restricted to TATA-box region (Emami et al., 1997), suggesting that TAFs do not interact with the DNA sequences downstream of the TATA-box. In this scenario, HMGA1 might compete with TBP to bind TATA-box and therefore repress TATA-only promoter transcription at high concentrations (Fig. 2.6; Fig. 2.2.8). Alternatively, the HMGA1 interaction with TFIID might induce a conformational change in TFIID, allowing TFIID to recruit subsequent GTFs and RNA Pol II

more efficiently. However, HMGA1 might not be the only protein factor that can facilitate or stabilize the TFIID-core promoter binding. Topo I and the other "HMG" architectural proteins may substitute for HMGA1 in a gene context and cell type-specific manner. Given the diverse chromatin and gene regulatory roles of HMGA1 (Reeves 2003), including the novel core promoter-specific function described in this study, the transcriptional contribution of HMGA1 to the regulation of specific genes is likely to be cell type- and context-dependent. A hypothesis that is consistent with the observed tissue specificity of HMGA1-dependent gene regulation (Martinez Hoyos et al. 2004).

Mediator is considered to be a general coactivator that bridges numerous activators with the general transcription machinery. Mediator is essential for transcription and associates with RNA Pol II in a holo-enzyme, suggesting that Mediator could function as a general transcription factor in yeast (Kelleher et al., 1990; Kornberg 2005; Malik and Roeder 2010; Kagey et al. 2010). It was therefore surprising to find that Mediator can function as a core promoter-specific coregulator in the presence of HMGA1 (Chapter 2). The interaction of HMGA1 with CDK8-less Mediator (Fig. 2.2.2-B; 2.2.5) suggests that HMGA1 selectively interacts with a CDK8-less Mediator; or that the HMGA1 interaction with Mediator causes a conformational change, leading to the disassociation of CDK8. Notably, CDK8 (together with CycC and MED12/13) can block the RNA Pol II-Mediator interaction, resulting in repression of transcription (Wang et al., 2001; Elmlund et al., 2006). Therefore, HMGA1 displacement of CDK8 from Mediator facilitates the Mediator-RNA Pol II interaction result in

transcriptional activation. The recruitment of both HMGA1 and Mediator to regulatory DNA/enhancer sequences and the reported involvement of both HMGA1 and Mediator in long-range enhancer function via DNA/chromatin looping (Bagga et al., 2000; Kagey et al., 2010) and in core promoter-selective stimulation (as shown in Chapter 2) suggest possible concerted functions of HMGA1 and Mediator at the interface between distal regulatory elements and specific core promoters. The HMGA1-Mediator interaction could not only facilitate the long-range communication between activators/enhancers and the basal transcription machinery, but also mediate the core promoter-selective function of certain activators and enhancers.

**HMGA1 interaction with TFIID could provide AT-hooks for human TFIID**

AT-hooks are AT-rich DNA minor groove binding domains and have been discovered in a large number of chromatin and DNA-binding proteins, including *Drosophila* TAF1 (Aravind and Landsman 1998). HMGA1 was speculated to serve as an accessory DNA-binding domain for several transcription factors, presumably by anchoring them to particular DNA structures (e.g., AT-rich DNA minor groove and four-way junction) (Aravind and Landsman 1998; Reeves 2003). Interestingly, the comparisons of the amino acid sequences from human and *Drosophila* TAF1 revealed two surprising differences. The first difference is that *Drosophila* TAF1 contains AT-hook domains (isoforms B and C contain two AT-hooks and isoform A contains one), while human TAF1 has no AT-hook (Metcalf and

Wassarman 2006; Fig. 4.2). The other difference is that *Drosophila* TAF1 HMG-box contains a 35 amino acid-insertion, which may disrupt its ability to bind DNA (Weinzierl et al., 1993). Therefore *Drosophila* TAF1 AT-hooks may substitute for its inactive HMG-box in DNA and core promoter binding. In contrast, the DNA-binding activity of human TAF1 HMG-box is regulated through phosphorylation by CK2 (Lewis et al., 2005), thus the requirement of HMGA1 AT-hooks for human TFIID is much more flexible.

It has been shown that a heterodimeric complex composed of human TAF1 and *Drosophila* TAF2 subunits of TFIID can bind specifically to the INR sequences (Chalkley and Verrijzer 1999). In contrast, *Drosophila* TAF1 isoforms with two AT hooks can directly bind the transcription start site of several *Drosophila* core promoters independently of TAF2 (Metcalf and Wassarman 2006). Furthermore, TAF1 is the only subunit within *Drosophila* TFIID that contacts the INR, as indicated by short-range protein-DNA cross-linking (Wu et al., 2001). This study shows that the HMGA1 interacts with human TFIID (Fig. 2.2.5; Fig. 2.2.6). Thus, it is tempting to speculate that the interaction of HMGA1 COOH-terminal with TFIID complexes that lack one or both AT hooks could compensate for the missing AT hooks and facilitate specific TFIID-INR interactions or stereo-specific conformations at TATA/INR core promoters (Fig. 4.2). It appears possible that the interaction of HMGA1 COOH-tail with TFIID and Mediator is regulated through the phosphorylation by CK2.

**Activators, coregulators and the core promoters regulate gene expression**

Transcriptional initiation is regulated through multiple steps (e.g. activator binding, coregulator recruitment, core promoter recognition). Most genes have multiple promoters and one promoter can have multiple transcription factor binding sites, including distal enhancers, proximal regulatory elements and core promoter elements (Sandelin et al., 2007). It is well known that combinatorial binding of activators to enhancers/proximal promoters can regulate gene expression, but little is known about the contribution of core promoters and core promoter-specific coregulators in gene regulation (Smale et al., 2001, Smale and Kadonaga 2003; this study). My second part study focuses on the function of the core promoters and the core promoter-specific coregulators in combinatorial gene regulation (Fig 4.4).

Most viral promoters contain multiple core promoter elements and most viral activators (eg. VP16) can function through diverse core promoter elements (Emami et al., 1995; Juven-Gershon et al., 2006). The mechanisms enhance the ability of the virus to survive in host cells. In eukaryotes, it is rare to find a gene core promoter that contains all known core promoter elements (Jin et al., 2006). In addition, some of the eukaryotic enhancers/activators can only function through specific core promoter elements. A test of 18 *Drosophila* lines demonstrates that four enhancer/promoter sequences out of 18 tested possess preferential selection with core promoters (Chapter 1, Introduction; Butler and Kadonaga 2001). Recently, Elf-1 binding sequences preference for the INR is revealed (Ernst et al., 1996). In addition,

the TATA-INR core promoter competes with INR-DPE core promoter to preferentially use *Drosophila* AE1 and IAB5 enhancers (Ohtsuki et al., 1998).

In the second part of this study of activated transcription, the expression of luciferase gene was shown to be regulated by different activators as well as different core promoters (Fig. 3.3). More extraordinarily, a strict TATA-specificity of the *ACTB* upstream activating sequences was revealed for the first time (Fig 3.6), which is analogous to the INR-specificity of the TdT promoter upstream activating sequences (summarized in Fig. 4.3). Subsquent statistical analysis of the SRF-target genes suggests that SRF might function selectively through a TATA-box (Table 3.1, this table might contain bias because only the Actin cytoskeleton-contractile SRF target gene groups are analyzed). The discovery of the preference of the *ACTB* proximal promoter for the TATA-box enlarges the list of examples for the selective communication between enhancers/proximal promoters and core promoter elements, and supports the importance of the core promoter elements in combinatorial gene regulation (summarized in Fig. 4.3). The RNAi knock down of HMGA1 shows that HMGA1 supports Gal4-Pro but not Gal4-Oct2 activated transcription and is not required for the *ACTB* proximal promoter dependent transcription (Chapter 3, Fig 3.7), suggesting that the coregulator-HMGA1 might also regulate gene expression in an activator specific fashion (Gal4-Pro). This part of study shows again the important role of the core promoter in gene regulation.

In summary, this study has elucidated a novel core promoter-selective cooperative function of HMGA1 and Mediator at TATA/INR core promoter. The HMGA1/Mediator interaction involves TAFs and the COOH-terminal acidic tail domain of HMGA1. From the dual roles of HMGA1 and Mediator at enhancers and core promoters, I propose that distal regulators and their recruited coregulators could coordinate the assembly of stereo-specific pre-initiation complexes at the core promoters and dictate the productive utilization of specific core promoter DNA elements by the general transcription machinery. Thus, the expression of a specific gene might be the combinatorial regulation event involving activators, coregulators (the general and the core promoter-specific coregulators) and core promoters (Fig 4.4). The reconstitution of distinct core promoter sequence-dependent transcription pathways *in vitro* with purified factors is an important first step to test this model.

## Figure legends

**Figure 4.1** Core promoter-specific factors are required to regulate transcription from different core promoters. Core promoter specific factors for different groups of core promoter (from top to bottom): GTFs and RNA Pol II are sufficient to support TATA promoter-directed transcription, while HMGA1, Topo I and Mediator are additionally required to support the TATA/INR synergy (this study). In mammalian system, TIC2 fraction is shown to support the INR/DPE promoter transcription by Martinez laboratory (Martinez et al., 1998); PC4, CK2 and Mediator are shown to be required for the Sp1-activated INR/DPE promoter transcription by Reinberg laboratory (Lewis et al., 2005). In *Drosophila* system, NC2 and Mot1 are shown to be the INR/DPE-specific factors to stimulate the INR/DPE driven transcription as well as to repress the TATA-directed transcription by Kadonaga laboratory (Willy et al., 2000; Hsu et al., 2008). The core promoter elements are boxed. GTFs and RNA Pol II are indiacated with white color and the core promoter-specific factors are indicated with different colors.

**Figure 4.2** The hypothesis of HMGA1 interaction with TFIID to provide AT-hooks for human TFIID. The AT-hooks are colored with light green, the HMG-boxes are colored blue and the HMGA1 COOH-tail is colored yellow. The small red box inside the *Drosophila* HMG-box indicates the 35 amino acid-insertion. Human TAF1 HMG-box has no insertion. HMGA1 C-terminal domain might interact with TAF1 or the other TAFs to provide AT-hooks for human TFIID to bind AT-rich core promoter elements.

**Figure 4.3** The studies of TATA-box and INR in combinatorial gene regulation. Many activators, such as VP16 and Sp1, can function through both a TATA and an INR (known). The murine lymphocyte-specific TdT promoter upstream sequences function only via the INR. Elf-1 was reported to be the potential activator in this scenario (Smale laboratory, known). The *ACTB* proximal activating sequences preferentailly function through the TATA-box (this study). SRF could be the potential activator in this scenario, which need to be further examined.

**Figure 4.4** Activators, coregulators and core promoters regulate gene expression. The gene-specific activators are recruited to the specific upstream activating sequences. The activators recruit general coregulators and/or the core promoter-specific coregulators (depending on the core promoter structure and the activators). The core promoter-specific coregulators could potentially coordinate the assembly of stereo-specific PICs at the core promoters and dictate the productive utilization of specific core promoter DNA elements by the basal transcription machinery.

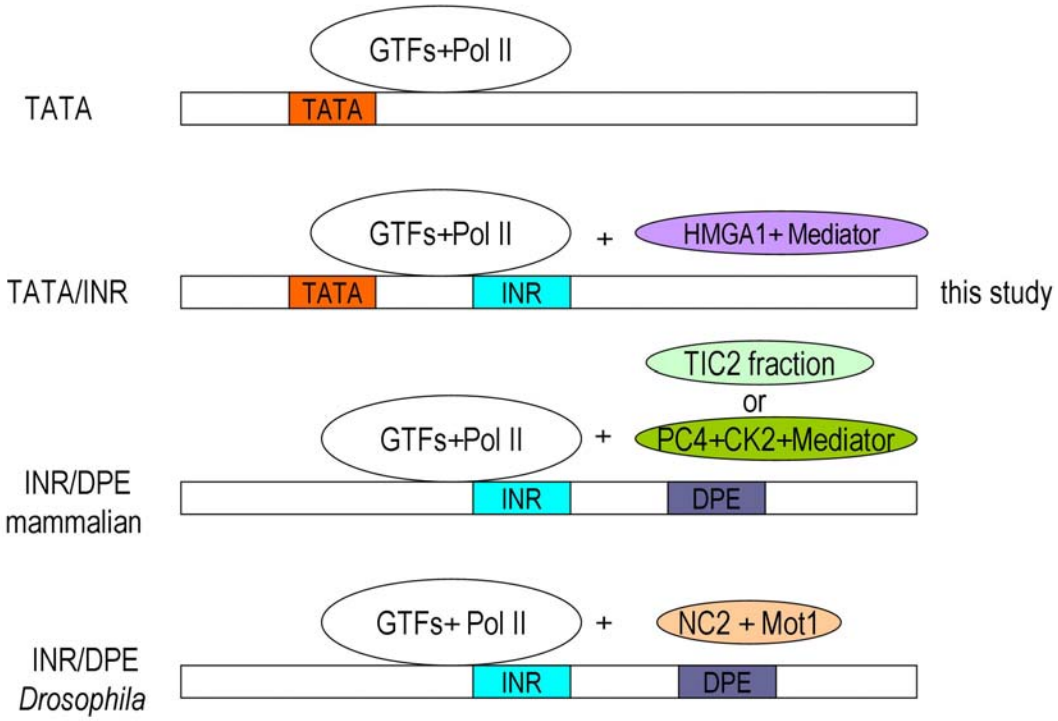



**Figure 4.1** Core promoter-specific factors are required to regulate transcription from different core promoters.

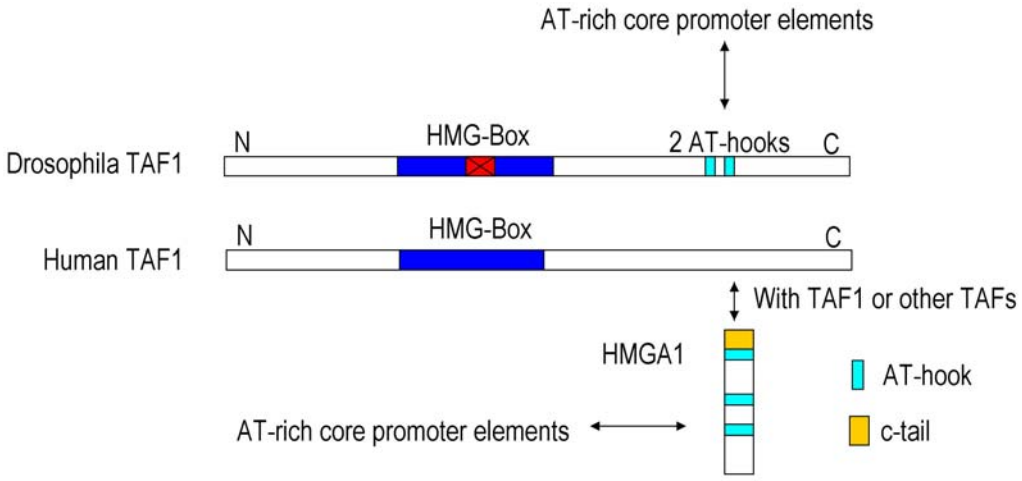



**Figure 4.2** The hypothesis of HMGA1 interaction with TFIID to provide AT-hooks for human TFIID

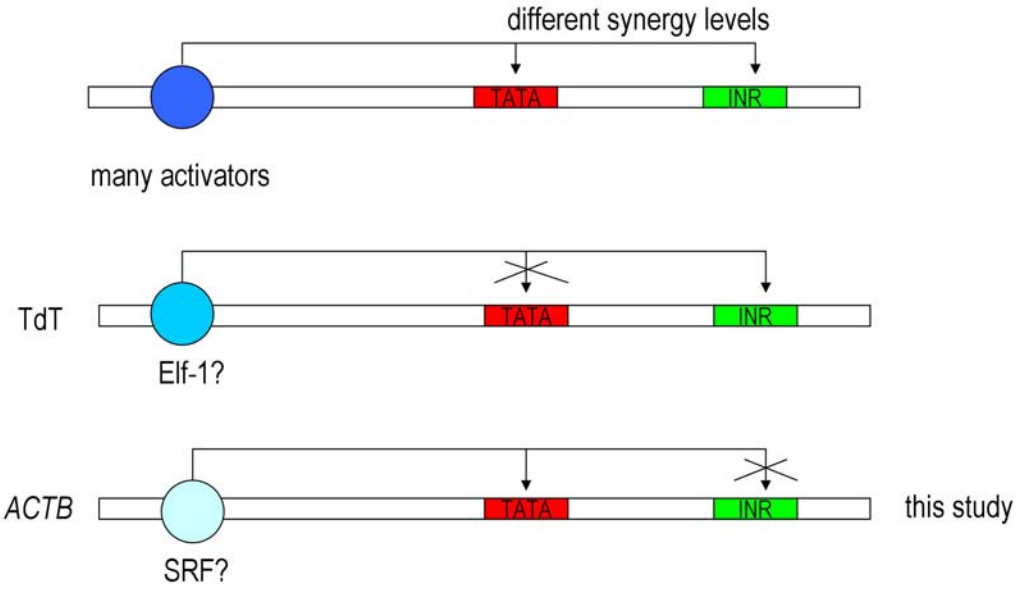



**Figure 4.3** The studies of TATA-box and INR in combinatorial gene regulation.

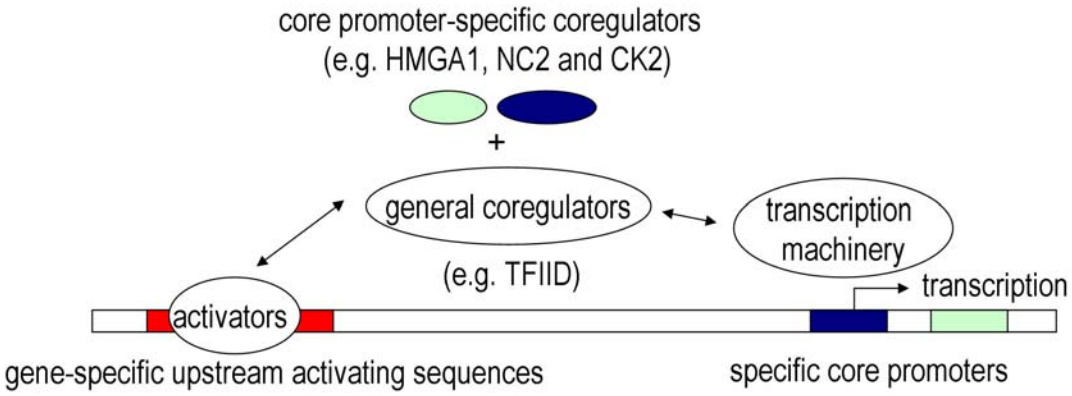



**Figure 4.4** Activators, coregulators and core promoters regulate gene expression.

**References:**


1. Andel F et al., 1999. Three dimensional structures of human TFIID-IIA-IIB complex. Science **286**: 2153-2155.

2. Anish R et al., 2009. Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. PloS One **4**: e5103.

3. Aravind L and Landsman D 1998. AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Res **26**: 4413-4421.

4. Asturias F et al., 1999. Conserved structures of Mediator and RNA Polymerase II Holoenzyme. Science **283**: 985-987.

5. Babb R et al., 2001. DNA recognition by the Herpes Simplex Virus transactivator VP16: a novel DNA-binding structure. Mol Cell Biol **21**: 4700-4712.

6. Baek HJ et al., 2006. Human Mediator enhances basal transcription by facilitating recruitment of TFIIB during Preinitiation complex assembly. J Biol Chem **281**: 15172-15181.

7. Baek HJ et al., 2002. Requirement of TRAP/Mediator for both activator-independent and activator-dependent transcription in conjunction with TFIID-associated TAFIIs. Mol Cell Biol **22**: 2842-2852.

8. Barlev NA et al., 2003. A novel human Ada2 homologue functions with Gcn5 or Brg1 to coactivate transcription. Mol Cell Biol **23**: 6944-6957.

9. Bjornsdottir G and Myers LC 2008. Minimal components of the RNA polymerase II transcription apparatus determine the consensus TATA box. Nucleic Acids Res **36**: 2906-2916.

10. Brand M et al., 1999. Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. Science **286**: 2151-2153.

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

12. Brownell JE and Allis CD 1995. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proc Natl Acda Sci USA **92**: 6364-6368.

13. Brownell, J.E et al., 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell **84**: 843-851.

14. Buratowski S et al., 1988. Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature **334**: 37-43.

15. Burke T and Kadonaga J 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to human and is recognized by TAFII60 of *Drosophila*. Genes Dev **11**: 3020-3031.

16. Burley SK and Roeder RG 1996. Biochemistry and structural biology of transcription factor IID (TFIID). Annu Rev Biochem **65**: 769-799.

17. Cai G et al., 2010. Mediator Head module structure and functional interactions. Nat Struct Mol Biol **17**: 273-279.

18. Carcamc J et al., 1991. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc Natl Acda Sci USA **88**: 8052-8056.

19. Catez F et al., 2004. Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. Mol Cell Biol **24**: 4321-4328.

20. Chai J and Tarnawski AS 2002. Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. J Physiol Pharmacol **53**: 147-157.

21. Chalkley G and Verrijzer P 1999. DNA binding site selection by RNA polymerase II TAFs: a TAFII250-TAFII150 complex recognizes the Initiator. The EMBO Journal **18**: 4835-4845.

22. Chen JL et al., 1994. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. Cell **79**: 93-105.

23. Chi T and Carey M 1996. Assembly of the isomerized TFIIA-TFIID-TATA ternary complex is necessary and sufficient for gene activation. Genes Dev **10**: 2540-2550.

24. Chin MT et al., 1998. Enhancement of serum-response factor-dependent transcription and DNA binding by the architectural transcription factor HMG-I(Y). J Biol Chem **273**: 9755-9760.

25. Corden J et al., 1980. Promoter sequences of eukaryotic protein-coding genes. Science **4463**: 1406-1414.

26. Danilition SL et al., 1991. Transcription factor binding and spacing constraints in the human beta-actin proximal promoter. Nucleic Acids Res **19**: 6913-6922.

27. de Silvio A et al., 1999. Dissection of the NF-Y transcriptional activation potential. Nucleic Acids Res **27**: 2578-2584.

28. Elmlund H et al., 2006. The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. Proc Natl Acda Sci USA **103**: 15788-15793.

29. Emami KH et al., 1995. Core romoter specificities of the Sp1 and VP16 transcriptional activation domains. Mol Cell Biol **15**: 5906-5916.

30. Emami KH et al., 1997. Mechanism of synergy between TATA and initiator: synergistic binding of TFIID following a putative TFIIA-induced isomerization. Genes Dev **11**: 3007-3019.

31. Ernst P et al., 1996. A potential role for Elf-1 in terminal transferase gene regulation. Mol Cell Biol **16**: 6121-6131.

32. Faiola F et al., 2005. Dual regulation of c-Myc by p300 via acetylation dependent control of Myc protein turnover and coactivation of Myc-induced transcription. Mol Cell Biol **25**: 10220-10234.

33. Flores O et al., 1989. Factors involved in specific transcription by mammalian RNA polymerase II: Purification and subunit composition of transcription factor IIF. J Biol Chem **265**: 5629-5634.

34. Flores O et al., 1992. Factors involved in specific transcription by mammalian RNA polymerase II: Identification and characterization of transcription factor IIH. J Biol Chem **267**: 2786-2793.

35. Fondell et al., 1996. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. Proc Natl Acda Sci USA **96**: 8329-8333.

36. Fusco A and Fedele M 2007. Role of HMGA1 in cancer. Nature Reviews **7**: 899-910.

37. Garraway IP et al., 1996. Transcription of the lymphocyte-specific terminal deoxynucleotidyltransferase gene requires a specific core promoter structure. Proc Natl Acda Sci USA **93**: 4336-4341.

38. Ge H et al., 1996. Activator-dependent transcription by mammalian RNA Polymerase II: *In vitro* reconstitution with General Transcription Factors and Cofactors. Methods in Enzymology **274**: 57-71.

39. Gershenzon N and Ioshikhes I 2004. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics **21**: 1295-1300.

40. Goldberg ML 1979. Ph. D. Thesis, Stanford University, Stanford, USA.

41. Goppelt A and Meisterernst M 1996. Characterization of the basal inhibitor of class II transcription NC2 from Saccharomyces cerevisiae. Nuclear Acid Research **24**: 4450-4455.

42. Grant PA et al., 1997. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev **11**: 1640-1650.

43. Guelman S et al., 2006. Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC: a novel dAda2A/dGcn5-containing histone acetyltransferase complex in *Drosophila*. Mol Cell Biol **26**: 871-882.

44. Hahn S 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol **11**: 394-403.

45. Häkkinen A et al., 2011. Genome wide study of NF-Y type CCAAT boxes in unidirectional and bidirectional promoters in human and mouse. J Theor Biol **281**: 74-83.

46. Hans E et al., 2006. The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. Proc Natl Acad Sci USA **103**: 15788-15793.

47. Heintzman ND et al., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature **459**: 108-112.

48. Hill D et al., 1999. Direct binding of HMG-I(Y) on four-way junction DNA and the molecular basis for competitive binding with HMG-1 and histone H1. Nucleic Acids Research **27**: 2135-2144.

49. Huth J et al., 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. Nature **4**: 657-665.

50. Hoffmann A et al., 1990. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). Nature **346**: 387-390.

51. Hoffmann A et al., 1996. A histone octamer-like structure within TFIID. Nature **380**: 356-359.

52. Hsu JY, et al., 2008. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATAdependent transcription. Genes Dev **22**: 2353-2358.

53. Ito M et al., 1999. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. Molecular Cell **3**: 361-370.

54. Jean O et al., 2001. HMG1 and 2, and reated 'architectural' DNA-binding proteis. Trends Biochem SCI **26**: 167-74.

55. Jin VX et al., 2006. Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. BMC Bioinformatics **7**:114.

56. Juven-Gershon T et al., 2006. Rational design of a super core promoter that enhances gene expression. Nat Methods **3**: 917-922.

57. Juven-Gershon T et al., 2008. The RNA polymerase II core promoter-The gateway to transcription. Curr Opin Cell Biol **20**: 253-259.

58. Kagey MH et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature **467**: 430-435.

59. Kamada K et al., 2001. Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell **106**: 71-81.

60. Kelleher RJ et al., 1990. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell **61**: 1209-1215.

61. Kim JL et al., 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature **365**: 520-527.

62. Kim YC et al., 1993. Crystal structure of a yeast TBP/TATA-box complex. Nature **365**: 512-520.

63. Kimura A et al., 2005. A decade of histone acetylation: marking eukaryotic chromosomes with specific codes. Journal of Biochemistry **138**: 647-662.

64. Koch HG et al., 2003. Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. Rev Physiol Biochem Pharmacol **146**: 55-94.

65. Kolb S et al., 2007. HMGA1 controls transcription of Insulin Receptor to regulate Cyclin D1 translation in Pancreatic Cancer cells. Cancer Research **67**: 4679-4686.

66. Kornberg RD 2007. The molecular basis of eukaryotic transcription. Proc Natl Acad Sci USA **104**: 12955-12961.

67. Kretzschmar M et al., 1993. Identification of human DNA topoisomerase I as a cofactor for activator-dependent transcription by RNA polymerase II. Proc Natl Acad Sci USA **90**: 11508-11512.

68. Kuo MH et al., 1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature **383**: 269-272.

69. Kusch T et al., 2003. Two *Drosophila* Ada2 homologues function in different multiprotein complexes. Mol Cell Biol **23**: 3305-3319.

70. Largrange T et al., 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev **12**: 34-44.

71. Leppard JB and Champoux JJ 2005. Human DNA topoisomerase I: relaxation, roles, and damage control. Chromosoma **114**: 75-85.

72. Lewis B et al., 2000. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. Proc Natl Acad Sci USA **97**: 7172-7177.

73. Lewis B et al., 2005. Functional characterization of core promoter elements: DPE-Specific transcription requires the Protein Kinase CK2 and the PC4 coactivator. Molecular Cell **18**: 471-481.

74. Li Y et al., 2007. Dual role for SUMO E2 conjugase Ubc9 in modulating the transforming and growthpromoting properties of the HMGA1b architectural transcription factor. J Biol Chem **282**: 13363-13371.

75. Liberman PM and Berk AJ 1991. The Zta trans-activator protein stabilizes TFIID association with promoter DNA by direct protein-protein interaction. Genes Dev **5**: 2441-2454.

76. Lieberman PM and Berk AJ 1994. A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. Genes Dev **8**: 995-1006.

77. Lim CY et al., 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev **18**: 1606-1617.

78. Lin J and Carey M 2012. *In vitro* transcription and immobilized template analysis of Preinitiation complexes. Current protocols in Molecular Biology Chapter **12**: Unit 12.14.

79. Malecova B et al., 2007. The Initiator core promoter element antagonizes repressionof TATA-directed transcription by Negative Cofactor NC2. J Biol Chem **282**: 24767-24776.

80. Malini E et al., 2011. HMGA interactone: New insights from Phage Display Technology. Biochemisty **50**: 3462-3468.

81. Malik S and Roeder RG 2003. Isolation and functional characterization of the TRAP/Mediator complex. Methods in Enzymology **364**: 257-284.

82. Malik S and Roeder RG 2005. Dynamic regulation of pol II transcription by Mammalian Mediator complex. Trends in Biochemical Sciences **30**: 256-263.

83. Malik S and Roeder RG 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet **11**: 761-772

84. Martinez E et al., 1994. TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. The EMBO Journal **13**: 3115-3126.

85. Martinez E et al., 1995. Core promoter-specific function of a mutant transcription factor TFIID defective in TATA box-binding. Proc Natl Acda Sci USA **92**: 11864-11868.

86. Martinez E et al., 1998a. Novel cofactors and TFIIA mediate functional core promoter selectivity by the Human TAFII150-containing TFIID complex. Mol Cell Biol **18**: 6571-6583.

87. Martinez E et al., 1998b. A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. J Biol Chem **273**: 23781-23785.

88. Martinez E 2002. Multi-protein complexes in eukaryotic gene transcription. Plant Molecular Biology **50**: 925-947.

89. Martinez E 2012. Core promoter-selective coregulators of transcription by RNA polymerase II. Transcription **3**: 1-6, in press.

90. Martinez Hoyos et al., 2004. Identification of the genes up- and down- regulated by the High Mobility Group A1 (HMGA1) proteins: tissue specificity of the HMGA1-dependent gene regulation. Cancer Reserch **64**: 5728-5735.

91. Matsui T et al., 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem **255**: 11992-11996.

92. Merino A et al., 1993. DNA topoisomerase I is involved in both repression and activation of transcription. Nature **365**: 8854-5635.

93. Metcalf C and Wassarman D 2006. DNA-binding properties of TAF1 isoforms with two AT-hooks. J Biol Chem **281**: 30015-30023.

94. Metz R et al., 1994. c-Fos-induced activation of a TATA-box-containing promoter involves direct contact with TATA-box-binding protein. Mol Cell Biol **14**: 6021-6029.

95. Meyer KD et al., 2010. p53 activates transcription by directing structural shifts in Mediator. Nat Struct Mol Biol **17**: 753-760.

96. Miano JM et al., 2007. Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. Am J Physiol Cell Physiol **292**: C70-81.

97. Muratoglu S et al., 2003. Two different *drosophila* ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. Mol Cell Biol **23**: 306-321.

98. Myers L and Kornberg RD 2000. Mediator of Transcription Regulation. Annu Rev Biochem **69**: 729-749.

99. Nicolov DB et al., 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature **377**: 119-128.

100. Nissen MS and Reeves R 1995. Changes in superhelicity are introduced into closed circular DNA by binding of high mobility group protein I/Y. J Biol Chem **270**: 4355-4360.

101. Noro B et al., 2003. Molecular dissection of the architectural transcription factor HMGA2. Biochemistry **42**: 4569-4577.

102. Oelgeschlager T et al., 1996. Topology and reorganization of a human TFIID-promoter complex. Nature **382**: 735-738.

103. Ogryzko VV et al., 1998. Histone-like TAFs within the PCAF histone acetylase complex. Cell **94**: 35-44.

104. Ohler U et al., 2002. Computational analysis of core promoters in the *Drosophila* genome. Genome Biol **3**: Research0087.

105. Ohtsuki S et al., 1998. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. Genes Dev **12**: 547-556.

106. O'Shea-Greenfield A and Smale ST 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J Biol Chem **267**: 1391-1402

107. Pankiewicz R et al., 2005. Reversal of the silencing of tetracycline-controlled genes requires the coordinate action of distinctly acting transcription factors. J Gene Med **7**: 117-132.

108. Panne D 2008. The enhanceosome. Curr Opin Struct Biol **18**: 236-242.

109. Parker CS and Topol J 1984. A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell **36**: 357-369

110. Parry et al., 2010. The TCT motif, a key component of a RNA polymerase II transcription system for the translational machinery. Genes Dev **24**: 2013-2018

111. Pierantoni GM et al., 2003. A truncated HMGA1 gene induces proliferation of the 3T3-L1 pre-adipocytic cells: A model of human lipomas. Carcinogenesis **24**: 1861-1869.

112. Pohler R et al., 1998. HMG box proteins bind to four-way DNA junctions in their open conformation. The EMBO Journal **17**: 817-826.

113. Rachez C et al., 1998. A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system gene. Genes Dev **12**: 1787-1800.

114. Reeves R and Beckerbauer L 2001. HMGI/Y proteins: flexible regulators of transcription and chromatin structure. Biochim Biophys Acta **1519**: 13-29.

115. Reeves R 2003. HMGA proteins: Flexibility finds a nuclear niche? Biochem Cell Biol **81**: 185-195.

116. Reinberg D and Roeder RG 1987. Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. J Biol Chem **262**: 3310-3321.

117. Remacle JE et al., 1997. Three classes of mammalian transcription activation domain stimulate transcription in Schizosaccharomyces pombe. The EMBO J **16**: 5722-5729.

118. Rhee HS and Pugh BF 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature **483**: 295-301.

119. Roeder RG and Rutter WJ 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature **224**: 234-237.

120. Roeder RG 2005. Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett **579**: 909-915.

121. Sandelin et al., 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nature Reviews Genetics **8**: 424-436

122. Sawadogo M and Roeder RG 1985. Factors involved in specific transcription by human RNA polymerase II: Analysis by a rapid and quantitative *in vitro* assay. Proc Natl Acda Sci USA **82**: 4394-4398.

123. Seto E et al., 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. Nature **354**: 241-245.

124. Sgarra R et al., 2008. Interaction proteomics of the HMGA chromatin architectural factors. Proteomics **8**: 4721-4732.

125. Shao H et al., 2005. Core promoter binding by Histone-like TAF complexes. Mol Cell Biol **1**: 206-219

126. Shykind BM et al., 1997. Topoisomerase I enhances TFIID-TFIIA complex assembly during activation of transcription. Genes Dev **11**: 397-407

127. Smale ST and Baltimore D 1989. The “Initiator” as a transcription control element. Cell **57**: 103-113

128. Smale ST et al., 1990. Transcriptional activation by SP1 as directed through TATA or initiator: Specific requirement for mammalian transcription factor IID. Proc Natl Acda Sci USA **87**: 4509-4513

129. Smale ST and Kadonaga J 2003. The RNA Polymerase II core promoter. Annu Rev Biochem **72**: 449-479

130. Sterner DE and Berger SL 2000. Acetylation of histones and transcription-related factors. Microbiology and Molecular Biology Reviews **64**: 435-459.

131. Stros M et al., 2007. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol Life Sci **64**: 2590-2606.

132. Struhl K 1986. Constitutive and inducible Saccharomyces cerevisiae promoters: evidence for two distinct molecular mechanisms. Mol Cell Biol **6**:3847-3853.

133. Taatjes DJ et al., 2002. Structure, function, and activator-induced conformations of the CRSP coactivator. Science **295**: 1058-1062.

134. Taatjes DJ 2010. The human Mediator complex: A versatile, genome-wide regulator of transcription. Trends Biochem Sci **35**: 315-322.

135. Takagi Y and Kornberg R 2006. Mediator as a General Transcription Factor. J Biol Chem **281**: 80-89.

136. Takahashi et al., 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell **146**: 92-104.

137. Tanaka et al., 1994. The Oct2 glutamine-rich and Proline-rich domains can synergize with each other or duplicate themselves to activate transcription. Mol Cell Biol **14**: 6046-6055.

138. Theisen JW et al., 2010. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. Mol Cell Biol **30**: 3471-3479

139. Thomas M and Chiang CM 2006. The general transcription machinery and general cofactors. Critical Reviews in Biochemistry and Molecular Biology **41**: 105-178.

140. Thanos D and Maniatis T 1992. The high mobility group protein HMG I(Y) is required for NF-kB-dependent virus induction of the human IFN-b gene. Cell **71**: 777-789.

141. Tokusumi Y et al., 2007. The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. Mol Cell Biol **27**: 1844-1858.

142. Wang G et al., 2001. Characterization of Mediator Complexes from HeLa Cell Nuclear Extract. Mol Cell Biol **21**: 4604–4613.

143. Wang JC 2002. Cellular roles of DNA topoisomerases: a molecular perspective. Nature Rev Mol Cell Biol **3**: 430-440.

144. Wang YL et al., 2008. Human ATAC is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. J Biol Chem **283**: 33808-33815.

145. Wassarman D et al., 1999. TAF250 is required for multiple developmental events in *Drosophila*. Proc Natl Acda Sci USA **97**: 1154-1159.

146. Wefald et al, 1990. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. Nature 344: 260-262.

147. Weiss S and Gladstone L 1959. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J Am Chem Soc **81**: 4118–4119

148. Weil PA et al., 1979. Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell **2**: 469-484

149. Weinzierl R et al., 1993. Largest subunit of *Drosophila* transcription factor IID directs assembly of a complex containing TBP and a coactivator. Nature **362**: 511-517.

150. Willy P et al., 2000. A basal transcription factor that activates or represses transcription. Science **290**: 5493-982.

151. Wu CH et al., 2001. Analysis of core promoter sequences located downstream from the TATA element in the hsp70 promoter from *Drosophila* melanogaster. Mol Cell Biol **21**: 1593-1602.

152. Xiao H et al., 1994. The upstream activator CTF/NF1 and RNA polymerase II share a common element involved in transcriptional activation. Nucleic Acids Res **22**: 1966-1973.

153. Xu M et al., 2011. Core promoter-selective function of HMGA1 and Mediator in Initiator- dependent transcription. Genes Dev **25**: 2513-2524.

154. Yang C et al., 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene **389**: 52-65.

155. Yean D and Gralla J 1997. Transcription reinitiation rate: a special role for the TATA box. Mol Cell Biol **17**: 3809-3816.

156. Yie J et al., 1997. Intra- and intermolecular cooperative binding of high-mobility-group protein I(Y) to the β-interferon promoter. Mol Cell Biol **17**: 3649-3662.

157. Yie J et al., 1999. The role of HMG I(Y) in the assembly and function of IFN-b enhanceosome. The EMBO Journal **18**: 3074-3089.

158. Venter JC et al., 2001. The sequence of Human genome. Science **291**: 1304-1351.

159. Verrijzer P et al., 1995. Binding of TAFs to core elements directs promoter selectivity by RNA Polymerase II. Cell **81**: 1115-1125.

160. Zhang Q and Wang Y 2008. High mobility group proteins and their post-translational modifications. Biochim Biophys Acta **1784**: 1159-1166.

161. Zhang Q and Wang Y 2010. HMG modifications and nuclear function. Biochim Biophys Acta **1799**: 28-36.

162. Zhou Q et al., 1992. Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. Genes Dev **10**: 1964-1974

163. Zhou X et al., 1996. Genomic structure and expression of the murine Hmgi-c gene. Nucleic Acids Res **24**: 4071-4077.