UNIVERSITY OF CALIFORNIA SAN DIEGO

Control of Transcription by Core Promoter Elements and Methyl-CpG-binding Protein 2

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Joshua W. M. Theisen

Committee in charge
- Professor James T. Kadonaga, Chair
- Professor Christopher K. Glass, Co-chair
- Professor Bruce A. Hamilton
- Professor Katherine A. Jones
- Professor Bing Ren

2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

The Dissertation of Joshua W. M. Theisen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-chair

_____

Chair

University of California San Diego

2010

DEDICATION

This work is dedicated to the young women with Rett syndrome

who so generously allowed me to learn from them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thank you, first and foremost, to Jim. Under your guidance, I have grown so much as a scientist and as a person over the past 4+ years. Thank you for supporting me at every turn, for trusting me to explore on my own, and for always bringing me back on course when it was necessary. I am indebted to you always.

Thank you also to all of my lab mates. You gave me so much over the years. To Chin, for handing me your second-born. To Timur, for bringing me onto a project I could talk about in public. To Arthur, for being such a good neighbor. To Trevor, for being my first guinea pig. To Sam and Deb, for Shakespeare's and for always keeping things light. And, last but certainly not least, to Tammy, for being so much more than a mentor and friend. I will miss all of you.

To my friends, roommates and classmates, thank you for understanding. For knowing when to study, and when to party. When to read papers on the beach and when to play Trivial Pursuit instead of going to class. For San Felipe, Big Bear, Tahoe, Scripps, The Shack, Cass St. Bar & Grill and every other place we raised hell or got lost in conversation. Fer, Jason, Justin, Neekesh, Scott, Todd and Vieng, I will never forget what you have done and been for me. Thank you for everything.

To my siblings, words will never be able to express what we share. Even after all these years apart, being with you is so easy and natural. Thank you for loving me.

Finally, Mom and Dad, you have given me more than I can say, more even than I can know. I am who I am because of your love and guidance, and all my best qualities come from you. A million times over, thank you. I love you.

Chapter 2, in full, has been accepted for publication as: Theisen, J. W. M., C. Y. Lim, and J. T. Kadonaga. 2010. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Molecular and Cellular Biology*, in press.

VITA

2002	Bachelor of Science University of Wisconsin-Madison
2005-2010	Research Assistant University of California San Diego
2010	Doctor of Philosophy University of California San Diego

PUBLICATIONS

Juven-Gershon, T., J. Y. Hsu, J. W. M. Theisen, and J. T. Kadonaga. 2008. The RNA polymerase II core promoter - the gateway to transcription. *Current Opinions in Cell Biology* **20:**253-9.

Theisen, J. W. M., C. Y. Lim, and J. T. Kadonaga. 2010. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Molecular and Cellular Biology*, in press.

Parry, T. J., J. W. M. Theisen, J.-Y. Hsu, M. Eustice, D. L. Corcoran, U. Ohler, and J. T. Kadonaga. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. In submission

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Developmental Biology
Professor Thaddeus G. Golos, University of Wisconsin-Madison

Studies in Virology
Professor Paul D. Friesen, University of Wisconsin-Madison

Studies in Biochemistry
Professor James T. Kadonaga, University of California San Deigo

ABSTRACT OF THE DISSERTATION

Control of Transcription by Core Promoter Elements

and Methyl-CpG-binding Protein 2

by

Joshua W. M. Theisen

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2010

Professor James T. Kadonaga, Chair

Professor Christopher K. Glass, Co-chair

Transcription is the process of copying DNA into RNA. This process is critical for transmitting the information stored in DNA to the rest of the cell. The timing and amount of transcription must be regulated in order for cells and organisms to function

normally. This regulation is achieved through many levels of control. This dissertation explores regulation of transcription by core promoter elements and methyl-CpG-binding protein 2.

The core promoter is a diverse and complex entity. To gain a better understanding of the core promoter, I examined the function of the motif ten element (MTE). I found that the MTE recruits TFIID to the core promoter via interactions with TAF6 and TAF9. I also performed a detailed mutational analysis of the MTE and demonstrated the importance of nucleotides in the 27-29 subregion. This analysis identified three downstream subregions (18-22, 27-29, 30-33) that contribute to core promoter activity and led to the discovery of the novel 'Bridge' (18-22 and 30-33) core promoter motif.

MeCP2 is a methyl-CpG-binding protein that is required for normal neurodevelopment. Mutations in MeCP2 cause the neurologic disorder Rett syndrome. I examined the mechanism of transcriptional repression by MeCP2 and discovered a novel, histone deacetylase-independent mechanism of repression. This mechanism involves inhibition of pre-initiation complex formation and is not dependent on a previously identified transcription repression domain. Analysis of transcription templates with specifically placed CpG dinucleotides reveals that templates with more methylation sites or with sites closer to the transcription start site are subject to stronger repression of transcription by MeCP2. These findings show that MeCP2 has a second mechanism of transcriptional repression and highlight the importance of methylation patterns in directing MeCP2 function.

Chapter 1

Introduction

In a multicellular, eukaryotic organism most cells have the same DNA genome as every other cell. Yet, skin cells, muscle fibers and neurons look and behave in dramatically different ways. These cells also modify their behavior over time as part of normal development or in response to changing environmental conditions. The proximate source of these changes in behavior can be traced to changes in the cellular machinery that performs the tasks necessary for each cell to do what it does. For example, exposure of skin to sunlight causes increased production of the protein melanin by melanocytes, resulting in a tan. However, while the immediate source of the tan is increased melanin, this increase is just the final result of a complex regulatory hierarchy. Though every step in the production of melanin, or any other protein, is exquisitely regulated, the first step is in many ways the most important. Once the first step is taken, the cell has committed energy and resources to either make the protein or degrade the intermediate products. This first step is transcription.

Transcription is the process of copying DNA, the chemical repository of all the information needed for the development, survival and reproduction of an organism, into RNA, the chemical conveyance of this information. The information in DNA is encoded by the arrangement of four chemical structure, adenine, cytosine, guanine and thymine. This arrangement an be represented by a sequence of the letters A, C, G and T. Together, these letters spell the instructions for every process that occurs in all cells.

In eukaryotic cells, three distinct RNA polymerases are responsible for all nuclear transcription (68). RNA polymerase I and III transcribe rRNAs, tRNAs, other

RNA species (65). RNA polymerase II (RNAP II) is responsible for transcription of all known protein coding genes (73), and it is this class of transcription that is the focus of this dissertation. However, while transcription occurs when RNAP II binds to DNA and begins to polymerize ribonucleotides, many layers of control and organization must be navigated before this occurs.

DNA can be conceived of as a linear string. Each human cell has approximately 2 meters of total DNA packed into a nucleus with a diameter of only 6 micrometers. This feat requires 300,000-fold compaction of the DNA molecules. To accomplish this compaction and to maintain the structural and organizational integrity of the DNA, it is packaged into a nucleoprotein complex called chromatin (44). The basic repeating unit of chromatin is the nucleosome, which consists of approximately 148 bp of DNA wrapped around a core histone octamer (56). The core histone octamer is made up of two histone H2A-histone H2B dimers and a histone H3-histone H4 tetramer.

The histone proteins serve a number of vital purposes in the nucleus. In addition to aiding in the compaction and organization of DNA, histones also contribute to regulation of DNA-directed processes, including transcription (53). Histones can physically block sequence-specific DNA-binding transcription factors from interacting with their binding sites (described later), thereby inhibiting the function of these factors. Alternatively, histones can be chemically modified in ways that serve as signals to regulators of transcription (29). Known modifications of histone proteins include acetylation (69), methylation (67), SUMOylation (72),

phosphorylation (8) and ubiquitination (82). These posttranslational modifications change the biophysical properties of chromatin and are detected and interpreted by various nuclear proteins. Both of these features can result in increased or decreased transcription of the genes located near the modified histones. For example, acetylation of histones by histone acetyltransferases generally results in increased transcription of nearby genes (70), while deacetylation by histone deacetylases (HDACs) is associated with decreased transcription (2, 26, 46).

In addition to chemical modification of histones, chemical modification of DNA can also be used to regulate transcriptional activity. The most common DNA modification in mammalian genomes is CpG methylation (9). CpG methylation involves the substitution of a methyl group for a hydrogen atom at the 5 position of the cytosine ring. This substitution is catalyzed by DNA methyltransferases, which are required for embryonic development (49, 54, 64).

CpG methylation of DNA is recognized by the methyl-CpG-binding domain (MBD) family of proteins (10). Proteins in this family recognize and bind to methyl-CpG dinucleotides and recruit co-regulators to regions of DNA methylation. These co-regulators include Sin3A (30, 62), N-CoR (41), Dnmt1 (36), CoREST (4, 57) and others. In this way, the methylation signal is transmitted into some change in transcriptional activity, usually transcriptional repression.

The founding member of the MBD family of proteins is methyl-CpG-binding protein 2 (MeCP2) (58). This protein was first identified by its ability to selectively bind to CpG-methylated DNA. Later, it was shown that MeCP2 can repress

transcription, in part through recruitment of a class of histone modifying enzymes, the HDACs (30, 61, 62). Mutations in MeCP2 cause Rett syndrome, a severe neurodevelopmental disorder primarily affecting young girls (3). The devastating nature of this disease highlights the importance of MeCP2 in particular and transcription in general. My studies of the function of MeCP2 in transcriptional repression are detailed in Chapter 3 of this dissertation.

Even when the appropriate histone modifications and DNA methylation patterns required to allow transcription are present, the molecular machinery necessary for transcription must be recruited to the DNA. This recruitment is facilitated by sequence-specific DNA-binding transcription factors (50). These transcription factors generally have two functions (59): (1) they recognize and bind to short stretches of DNA with specific sequences, called binding sites, and (2) they recruit transcriptional co-regulators to increase or decrease transcription. Binding sites for transcription factors are often found in clusters, called enhancers, which can be located tens or even hundreds of kilobases away from the genes that they regulate (5, 83). Alternatively, these binding sites can be located in the proximal promoter (21). The proximal promoter is the region of DNA from approximately 50 bp to 250 bp upstream of the point at which transcription starts. After binding to the enhancer or the proximal promoter, sequence-specific DNA-binding transcription factors can then recruit co-regulators that will further modify nearby histones (48), to facilitate or inhibit transcription, or they can directly recruit the basal transcription factors necessary to accomplish transcription .

Biochemical studies of the basal machinery involved in transcription have identified five basal transcription factors required for accurate initiation of transcription, in addition to RNAP II (66). These are: transcription factor (TF) IID, TFIIB, TFIIF, TFIIH and TFIIE. While all of these factors are required for accurate transcription, TFIID is particularly important for two reasons. First, TFIID is the factor primarily responsible for recognizing and binding to the core promoter (78). The core promoter is the region of DNA extending approximately 40 bp upstream and downstream of the transcription start site, the exact position at which RNAP II initiates transcription. Second, the strength of TFIID binding to the core promoter plays a significant role in determining how much transcription occurs from a given gene (31).

Binding of TFIID to the core promoter is directed, in large part, through direct interaction with core promoter elements. Core promoter elements are DNA sequences in the core promoter (33, 76). A number of core promoter elements have been identified, including: the TATA box (22), the BRE (18, 19, 45), the initiator (75), the MTE (55), the DPE (12, 13), the XCPE1 (79) and the DCE (47, 51). The combination of core promoter elements in a given promoter plays a role in determining where transcription starts and how strong transcription is at that promoter. In addition, some sequence-specific DNA-binding transcription factors can only regulate transcription from promoters with certain core promoter elements (32). My studies of the function of downstream core promoter elements in recruiting TFIID and regulating transcription are detailed in Chapter 2 of this dissertation.

Eukaryotic cells are constantly bombarded with stimuli from an ever-changing environment. These signals must be detected and interpreted, and the appropriate responses must be made. Between stimulus and response exist many levels of control and regulation. One of the most important levels of control is the decision to increase or decrease transcription of protein coding genes. My work presented here examines two aspects of the regulation of transcription: the role of core promoter elements in activating transcription and mechanisms of transcriptional repression by MeCP2.

Chapter 2

Three Key Subregions Contribute to the Function of the Downstream RNA Polymerase II Core Promoter

ABSTRACT

The RNA polymerase II core promoter is a diverse and complex regulatory element. To gain a better understanding of the core promoter, we examined the motif ten element (MTE), which is located downstream of the transcription start site and acts in conjunction with the initiator (Inr). We found that the MTE promotes the binding of purified TFIID to the core promoter, and that the TAF6 and TAF9 subunits of TFIID appear to be in close proximity to the MTE. To identify the specific nucleotides that contribute to MTE activity, we performed a detailed mutational analysis and determined a functional MTE consensus sequence. These studies identified favored as well as disfavored nucleotides and demonstrated the previously unrecognized importance of nucleotides in the 27-29 subregion (+27 to +29 relative to A+1 in the Inr consensus) for MTE function. Further analysis led to the identification of three downstream subregions (18-22, 27-29, 30-33) that contribute to core promoter activity. The three binary combinations of these subregions leads to the MTE (18-22 and 27-29), DPE (27-29 and 30-33), and novel 'Bridge' (18-22 and 30-33) core promoter motifs. These studies have thus revealed a tripartite organization of key subregions in the downstream core promoter.

INTRODUCTION

The expression of the tens of thousands of genes within a cell is regulated during growth, development, and response to environmental stimuli. In eukaryotes, transcription of protein-coding genes is mediated by RNA polymerase II.

Transcription by RNA polymerase II is regulated by a wide variety of factors that include the basal transcription factors, sequence-specific enhancer- and promoter-binding proteins, co-regulatory factors, and other chromatin remodeling and modifying factors. The signals from these factors ultimately converge at the core promoter during the process of transcription initiation (for reviews, see: 17, 33, 76, 78).

The RNA polymerase II core promoter is the region of DNA that directs the initiation of transcription, and generally spans from about −40 to +40 nucleotides relative to the transcription start site. The core promoter is diverse in terms of its structure and function, as there are different mechanisms by which RNA polymerase II can be recruited to the promoter. For transcription that is directed by the TFIID transcription factor, there are several known sequences, termed core promoter elements, that mediate the recruitment of TFIID as well as other basal transcription factors to the DNA template. These core promoter elements include the TATA box, the initiator (Inr), the motif ten element (MTE), the downstream core promoter element (DPE), the TFIIB recognition elements (BRE^u and BRE^d), the downstream core element (DCE), and the X core promoter element 1 (XCPE1) (for recent review, see Ref. 33). In this work, we focus on the analysis of the MTE.

The study of the MTE began with the identification of motif 10 as an overrepresented sequence in a computational analysis of nearly 2000 *Drosophila* promoters (63). The motif 10 consensus sequence (CSARCSSAACGS) was then found to be a functional core promoter element, termed the motif ten element (MTE)

(55). The MTE functions cooperatively with the Inr, and there is a strict spacing requirement between the Inr and MTE motifs, as the insertion or deletion of a single nucleotide between the Inr and MTE was observed to result in a three- to 30-fold decrease in transcriptional activity. The addition of an MTE was found to compensate for the loss of basal transcription activity upon mutation of the TATA box or the DPE. Because MTE sequences from +18 to +27 were sufficient to confer MTE activity to heterologous promoters, the location of the MTE was designated to be from +18 to +27. In addition, there is synergism between the MTE and the TATA box as well as between the MTE and the DPE. This transcriptional synergism was exploited in the design of an unusually strong core promoter that contains TATA, Inr, MTE, and DPE motifs (31).

In the present study, we initially sought to gain a better understanding of the sequences that constitute the MTE as well as the role of the MTE in the binding of TFIID to the core promoter. In the earlier work, the mutation of the MTE was found to result in an alteration of the weak binding of a partially-purified preparation of *Drosophila* TFIID (55). Hence, it was necessary to examine interaction of TFIID with the MTE. To this end, we developed a new method for the purification of TFIID and performed DNase I footprinting and photocrosslinking experiments with the purified protein complex. It was also important to identify the specific sequences that are important for MTE function. We therefore carried out a systematic mutational analysis of the region encompassing the MTE. These studies unexpectedly led to a broader understanding of the downstream core promoter region. Specifically, we identified

three key subregions that, in different binary combinations, yield the MTE, DPE, and novel 'Bridge' core promoter motifs.

MATERIALS AND METHODS

Purification of *Drosophila* TFIID. The pCaSpeR-FLAG-TBP plasmid used in stable transfection of *Drosophila* S2 cells was constructed by PCR of the *TBP* gene with *Drosophila* genomic DNA sequences followed by subcloning into the pCaSpeR4 vector. The PCR was carried out in two steps and resulted in the introduction of sequences that add a double FLAG tag to the N-terminus of TBP. Aside from the sequences encoding the N-terminal FLAG tag, pCaSpeR-FLAG contains wild-type *TBP* sequences from 650 bp upstream of the initiating Met codon to downstream of the polyadenylation site, 460 bp downstream of the translation termination codon. Stably transfected *Drosophila* S2 cell lines were established by co-transfecting 5 μg of pCaSpeR-Flag-TBP with 1 μg of the selection plasmid, pCoHygro (Invitrogen). The cells were cultured in Schneider's *Drosophila* Medium (Gibco) supplemented with 300 μg/mL hygromycin B for the selection and maintenance of hygromycin resistance. For the purification of TFIID, the S2 cells expressing *Flag-TBP* were cultured in suspension in 4 x 400 mL of Schneider's *Drosophila* Medium supplemented with 100 μg/ml hygromycin B and 0.05% Pluronic F68 (Invitrogen). The cells were harvested (at > 6 x 10^6 cells/mL) by centrifugation in a Sorvall GSA rotor at 2000 rpm for 5 min and yielded a cell pellet of approximately 20 g.

TFIID containing FLAG-TBP was purified from ~20 g of cells as follows. The

cells were washed twice with PBS; pelleted by centrifugation in a Sorvall GSA rotor at 2000 rpm for 5 min; resuspended in 5 cell-pellet volumes (~100 mL) of Buffer H (10 mM Tris-HCl, pH 7.9, 10 mM KCl, 0.75 mM spermidine, 0.15 mM spermine, 0.1 mM EDTA, 0.1 mM EGTA, 2 mM DTT, 1 mM benzamidine-HCl, 0.2 mM PMSF, and 2 μg/mL each of aprotinin, leupeptin, and pepstatin); incubated on ice for 15 min; and lysed with 12 strokes of a 40 mL Dounce homogenizer with a loose (B) pestle. The resulting nuclei were pelleted by centrifugation in a GSA rotor at 7000 rpm for 10 min; suspended in 15 mL of Buffer H; repelleted by centrifugation in a Sorvall SS-34 rotor at 8000 rpm for 10 min; and resuspended in one nuclei-pellet volume (~5 mL) of Buffer E [50 mM Tris-HCl, pH 7.5, 0.6M KCl, 20% (v/v) glycerol, 10% (w/v) sucrose, 5 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 1 mM benzamidine-HCl, 0.2 mM PMSF]. The mixture was stirred at 4°C for 30 min, and then subjected to centrifugation in a Beckman SW28.1 rotor at 25,000 rpm for 1 h. The supernatant was dialyzed for 3 x 50 min against 2 L (for each dialysis) TM buffer [50 mM Tris-HCl (pH 7.9), 10% (v/v) glycerol, 12.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1 mM benzamidine-HCl, 0.1 mM PMSF, and 1 mM sodium metabisulfite] containing 0.1 M KCl. (When necessary, the extract was diluted with TM buffer without KCl until the mixture had the same conductivity as TM buffer with 0.1 M KCl.) The mixture was subjected to centrifugation in an SS-34 rotor at 10,000 rpm for 10 min. The supernatant was incubated with 0.01 volume of 1 mg/mL sonicated calf thymus DNA (to a final concentration of 10 μg/mL DNA) at 4 °C for 30 min, and then centrifuged in a SS-34 rotor at 10,000 rpm for 10 min to remove insoluble material. The

supernatant, which was typically about 10 to 15 mL in volume, was applied to five parallel 1 mL sequence-specific DNA affinity columns containing the *Drosophila* G core promoter sequence (see, for example, Ref. 43) from −5 to +40 by using methods described previously (35). The 0.3-0.4 M KCl fractions were pooled and incubated with 150 μL (bed volume) of anti-FLAG M2 agarose resin (Sigma) for 8 to 10 h with rotation. The unbound proteins were removed by centrifugation at 2000 rpm for 3 min in a clinical centrifuge. After five sequential washes in HEGN buffer [25 mM HEPES (pH 7.6), 0.1 mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) NP-40] containing 0.1 M KCl and 5 mM β-mercaptoethanol, the agarose resin was transferred to an Amicon Ultrafree-MC spin column (0.45 μm), and residual buffer was removed by microcentrifugation at 3000 rpm for 10 sec. The bound TFIID complex was eluted from the anti-FLAG M2 resin by incubation with 100 μL of HEGN buffer containing 0.1 mg/mL of the 3 x FLAG peptide (Sigma), rotation of the sample at 4 °C for 20 min, and collection by microcentrifugation at 3000 rpm for 10 sec. The elution procedure was repeated two more times. The concentration of purified TFIID was approximately 50 nM.

Core promoter constructs. Plasmids containing minimal core promoters for the in vitro transcription and DNase I footprinting experiments were constructed by the insertion of double-stranded oligonucleotides into the *Xba* I and *Pst* I sites of the multiple cloning region of pUC119. The sequences of the oligonucleotides used to construct the core promoter templates are shown in Table 2.1

DNase I footprinting. DNase I footprinting probes were prepared by PCR

amplification of core promoter constructs with 5'-^{32}P-labeled M13 reverse sequencing primer and unlabeled M13 forward primer. The PCR amplification products were purified by polyacrylamide gel electrophoresis. DNase I footprinting reactions were carried out as described previously (12).

In vitro transcription. In vitro transcription reactions were carried out as described previously (81) by using 250 ng of supercoiled DNA templates with *Drosophila* high salt nuclear extracts (77). The resulting transcripts were subjected to primer extension analysis with 5'-^{32}P-labeled M13 reverse sequencing primer (AGCGGATAACAATTTCACACAGGA). Quantitation of reverse transcription products was carried out with a phosphorimager (GE Health Sciences). All experiments were carried out a minimum of three independent times to ensure reproducibility of the data.

Photocrosslinking analysis. The photoaffinity DNA probes were synthesized by minor modification of previously-described methods (6, 7). Synthetic oligonucleotides with promoter sequences from −35 to + 45 were used as templates instead of single-stranded M13 DNA. The 5-[*N*-(*p*-azidobenzoyl)-3-aminoallyl]-dUTP (AB-dUTP, also known as N_3RdUTP) was generously provided by Drs. George Kassavetis and E. Peter Geiduschek (University of California, San Diego). The synthesized probes were subjected to 5% polyacrylamide gel electrophoresis, and eluted passively (overnight at room temperature) into 0.5 mL elution buffer [20 mM Tris-HCl (pH 8.0), 0.2 mM EDTA, 0.2% (w/v) SDS, 1 M LiCl]. The DNA probes were precipitated with ethanol and resuspended in 30 μL of TE buffer. In a typical

crosslinking experiment, photoaffinity DNA probe (50 fmol) was incubated with purified TFIID (approximately 80 ng) in a 10 μL reaction containing 12.5 mM HEPES, (pH 7.6), 50 mM KCl, 6.75 mM $MgCl_2$, 0.05 mM EDTA, 2.5 mM 2-mercaptoethanol, 5% (v/v) glycerol, 0.05% (v/v) NP-40 and 2% (v/v) polyvinyl alcohol. The binding reactions were incubated on ice for 15 min and then at 25°C for 15 min. The samples were subjected to short-wavelength UV irradiation at 380 $\mu W/cm^2$ for 5 min. Irradiated protein-DNA complexes were digested with DNase I at 37°C for 10 min. A solution of 10% (w/v) SDS was added to a final concentration of 0.5%, and the samples were heated at 90 °C for 3 min. The DNA-protein complexes were additionally treated at 37 °C for 10 min with S1 nuclease (Invitrogen). The crosslinked proteins were resolved by 10% polyacrylamide-SDS gel electrophoresis. The gel was dried and subjected to autoradiography. The identity of the TAF6 and TAF9 subunits was confirmed by western blot analysis.

RESULTS

A new method for the purification of *Drosophila* TFIID. To gain a better understanding of the role of the MTE in the binding of TFIID to the core promoter, we sought to develop a more rapid and reliable method for the generation and purification of TFIID. To this end, we stably transfected *Drosophila* S2 cells with a modified *Drosophila TBP* gene that encodes N-terminally FLAG-tagged TBP (Figure 2.1A). The TFIID complex containing FLAG-TBP was purified from the stably transfected cells as outlined in Figure 2.1B. To enrich for TFIID relative to other TBP-containing

complexes as well as free TBP, we partially purified TFIID by sequence-specific DNA-affinity chromatography with the core promoter sequence of the TATA-less, DPE-containing *Drosophila G* promoter. By western blot analysis with antibodies against *Drosophila* TBP and TAF1, we determined that most of the TFIID complex elutes from the DNA affinity column in the 0.3-0.4 M KCl fraction. The TFIID in the 0.3-0.4 M KCl fraction was then purified to near homogeneity by anti-FLAG immunoaffinity chromatography (Figure 2.1C). The highly purified TFIID is similar to other preparations of TFIID obtained from *Drosophila* embryos (see, for example: 13, 20, 27, 40).

The MTE promotes transcription by increasing the affinity of TFIID for the core promoter. The purification of TFIID enabled us to carry out DNase I footprinting experiments with the MTE-containing *Drosophila Tollo* and CG10479 core promoters (Figure 2.2). The binding of TFIID to the *Tollo* promoter is much more distinct than that observed previously (55), possibly due to increased purity and activity of the TFIID. The wild-type versions of the *Tollo* and CG10479 promoters exhibit related patterns of alternating DNase I protection and hypersensitivity that extend from approximately –25 to +35 relative to the A+1 in the Inr consensus sequence. The most prominent DNase I protection and hypersensitivity span from the Inr to about the +15 position. These results indicate that TFIID binds throughout the core promoter region of MTE-containing promoters.

To examine the role of the MTE in the binding of TFIID to the core promoter, we performed DNase I footprinting experiments with a series of core promoters that

contain mutations in the MTE, the DPE, or both the MTE and DPE motifs. We employed mutations that had been previously demonstrated to inactivate the MTE and DPE motifs (55). Specifically, m18-22 inactivates the MTE via mutation of the sequences from +18 to +22 (relative to the A+1 in the Inr) to ATCCA, whereas m30-33 inactivates the DPE by mutation of the sequences from +30 to +33 to CATA.

Mutation of the MTE or the DPE results in a reduction in the interaction of TFIID with the *Tollo* and CG10479 promoters, and mutation of both the MTE and the DPE further decreases the binding of TFIID (Figure 2.2). The strength of TFIID binding to DNA observed in this work correlates with the efficiency of transcription that had been previously determined with the same promoter constructs (55). Thus, the MTE contributes to the affinity for TFIID and the transcriptional activity of the *Tollo* and CG10479 core promoters.

We additionally sought to determine whether the addition of MTE sequences to an MTE-deficient core promoter increases the binding of TFIID. To this end, we used the *Drosophila E74B* and *Doc* core promoters, both of which contain Inr and DPE motifs but lack an MTE (55). The addition of *Tollo* MTE sequences to the *E74B* and *Doc* promoters results in stronger binding of TFIID to the promoter (Figure 2.3), which correlates with the increase in transcriptional activity that was observed with the same promoter constructs (55). Hence, studies involving either the loss or the gain of the MTE at the core promoter indicate that the MTE promotes transcription by increasing the affinity of the binding of TFIID to the core promoter.

The DNase I footprinting patterns with the MTE-containing promoters are

similar to those seen with DPE-containing promoters (see, for example: 12, 13, 43). There is alternating DNase I protection and hypersensitivity, which suggest a close interaction of TFIID with one face of the DNA helix. The upstream boundaries of the protected and/or hypersensitive sites are from about −15 to −24. When it is considered that each boundary of DNase I protection and/or hypersensitivity is several nucleotides or larger than the actual site of protein binding, these data are consistent with the interaction of TFIID with the promoter from the Inr though the downstream promoter region, and are consistent with cooperativity between the MTE and Inr for transcriptional activity (55).

Photocrosslinking analysis of purified TFIID with MTE-containing promoters. To ascertain which subunit or subunits of TFIID are in close proximity to the MTE, we performed photocrosslinking studies of TFIID bound to MTE-containing promoters. In previous work on the DPE (12), we generated photoaffinity probes by the incorporation of 5-[*N*-(*p*-azidobenzoyl)-3-aminoallyl]-dUTP (AB-dUTP, also known as N3RdUTP; see Refs. 6, 7), a photoreactive TTP analogue, into the DNA. The resulting photocrosslinking experiments revealed that TAF6 and TAF9 (also known as $TAF_{II}60$ and $TAF_{II}40$, respectively) are in close proximity to the DPE. To enable the comparison of the new MTE data with the previous results on the DPE (12), we also used AB-dUTP to examine the interaction of TFIID with the MTE.

To investigate the interaction of TFIID subunits with the downstream core promoter region, we generated photoaffinity probes with the *Tollo* promoter and the *E74B*+MTE hybrid promoter. The *Tollo* promoter is a natural core promoter that

contains both MTE and DPE motifs (55). The *E74B*+MTE hybrid promoter also contains both MTE and DPE motifs, and is identical to the construct used in the DNase I footprinting analysis shown in Figure 2.3A. The MTE contributes to the binding of TFIID to each of these promoters (Figs. 2A and 3A). As depicted in Figure 2.4A, a photoreactive AB-dUMP nucleotide was introduced at the +20, +25, or +30 position (relative to the +1 start site) of each DNA probe alongside a radiolabeled nucleotide that enabled detection of the crosslinked polypeptides. Experiments were performed in the presence or absence of TFIID as well as with or without UV irradiation to allow the identification of polypeptides to which photocrosslinking is dependent upon both TFIID and UV activation.

With the *Tollo* promoter, we observed photocrosslinking of a 60 kDa band at +30 (Figure 2.4B). By western blot analysis, this band was found to co-migrate with TAF6, which is consistent with previous results on the DPE (12). We also observed a weak 60 kDa signal at +20 but not at +25. The 60 kDa signal at +20 is suggestive of an interaction of TAF6 with the MTE region. In addition, a weak band at 40 kDa, which corresponds to TAF9 by western blot analysis, can be seen. The 40 kDa band is, however, not sufficiently strong for a conclusion to be drawn regarding the interaction of TAF9 with the *Tollo* promoter.

With the *E74B*+MTE promoter, we observed strong photocrosslinking of the 60 kDa TAF6 band at +30 as well as a distinct TAF6 band at +20 but not at +25 (Figure 2.4C). We also saw crosslinking of TAF9 at +20 and +30 but not at +25. The crosslinking of both TAF6 and TAF9 at +20 and +30 but not at +25 suggests that these

proteins are in close proximity to one face of the DNA helix in this region. In this regard, it is relevant to note that TAF6 and TAF9 are related to histones H4 and H3 (see, for example, Ref. 14). It is therefore possible that a TAF6- and TAF9-containing subcomplex of TFIID interacts with the downstream core promoter region with key contacts at +20 and +30.

Single nucleotide mutational analysis reveals sequences that are important for MTE activity. To determine the key sequences that contribute to MTE function, we carried out a saturating single nucleotide mutational analysis of the MTE region of the *Tollo* and CG10479 promoters. We performed these studies with two different core promoters because we sought to identify sequences that are important for MTE function in multiple sequence contexts. For each promoter, we generated a set of constructs containing A, C, G, or T at each position from +15 to +29 relative to the A+1 in the Inr. To study the function of the MTE in the absence of the DPE, all constructs contained the m30-33 mutation, which inactivates the DPE motif.

In vitro transcription analysis of the *Tollo* and CG10479 promoters revealed both similarities and differences in their responses to specific mutations (Figure 2.5). For instance, at +17, both promoters exhibit a preference for C or T relative to A or G, whereas at +28, there is a preference for A or G relative to C or T. In contrast, at +22, +23, and +27, we observed different responses to mutations in the *Tollo* versus CG10479 promoters, even though the wild-type nucleotides of the two promoters are the same at those positions. These findings indicate that the preferred nucleotides in the MTE are influenced by the surrounding sequence.

As a guide for the analysis of the MTE, we categorized the single nucleotide substitution data in Figure 2.5 as follows. For each position, nucleotides that result in >90% of the activity of the wild-type nucleotide (which is defined to be 100%) in both the *Tollo* and CG10479 promoters are designated as 'Favored Nucleotides', whereas nucleotides that result in <60% of the wild-type nucleotide in both promoters are designated as 'Disfavored Nucleotides' (Figure 2.6). The favored nucleotide sequence is a generally more restrictive subset of the computationally-derived motif 10 consensus (63). It is also useful to note that the m18-22 sequence, which is ATCCA from +18 to +22, correlates well with the disfavored nucleotides at those positions.

Three downstream subregions are important for MTE and DPE activity. Inspection of the single nucleotide substitution data (Figure 2.5) reveals that there is a strong nucleotide preference at positions +27 to +29 of both the *Tollo* and CG10479 promoters containing the m30-33 mutation. The +28 and +29 positions overlap with the DPE consensus (33). Because the single nucleotide mutational analysis was performed with constructs containing the m30-33 mutation, which inactivates the DPE, it appeared that the +28 and +29 positions are important for both the MTE and DPE motifs.

We therefore examined the role of the 27-29 subregion in the function of the MTE and DPE motifs. To this end, we constructed and analyzed three series of core promoter constructs that contain all possible combinations of mutations in the 18-22, 27-29, and 30-33 subregions. The promoters used in this analysis each contain Inr, MTE, and DPE motifs and are as follows: (1) the wild-type *Tollo* promoter, which

contains Inr, MTE, and DPE motifs; (2) an *E74B*+MTE hybrid promoter that contains the Inr and DPE motifs from *E74B* and the *Tollo* MTE (from +18 to +27); and (3) a *Doc*+MTE hybrid promoter that contains the Inr and DPE motifs from *Doc* and the *Tollo* MTE (from +18 to +27). The *E74B*+MTE and *Doc*+MTE hybrid promoters are identical to those used in Figure 2.3. We employed the m18-22 and m30-33 mutations to inactivate the MTE and DPE, as in Figure 2.2 and Ref. 55, and used the disfavored sequence of GTA from +27 to +29 (Figure 2.6) to inactivate the 27-29 subregion.

In vitro transcription analysis of the three promoter series revealed that mutation of any one of the three downstream subregions (18-22, 27-29, 30-33) results in a substantial reduction in transcription relative to the reference promoters containing optimal sequences in the three subregions (Figure 2.7). Mutation of any two of the three subregions leads to a further decrease in transcriptional activity. Finally, constructs lacking all three subregions are essentially inactive.

Even though the 27-29 subregion is shared by the MTE and DPE motifs, the loss of the 27-29 subregion is not more deleterious to promoter activity than the loss of the 18-22 or the 30-33 subregions. Instead, it appears that the 18-22, 27-29, and 30-33 subregions each make roughly equivalent contributions to core promoter activity. These findings, in combination with the single nucleotide mutational data (Figure 2.5) and our current understanding of the MTE and DPE, suggest that the MTE comprises both the 18-22 and 27-29 subregions, whereas the DPE contains the 27-29 and 30-33 subregions. It is also interesting to note that core promoter constructs containing a novel configuration of the 18-22 and 30-33 subregions (with the m27-29 mutation)

have comparable activity to those containing MTE (18-22 and 27-29) or DPE (27-29 and 30-33) motifs (Figure 2.7).

Use of the downstream subregions in natural core promoters. The results of the mutational analysis of the three downstream subregions led us to consider whether all possible binary combinations of these subregions occur in natural core promoters. In previous studies, we found that the *Drosophila E74B* and *Doc* core promoters have a DPE (27-29 and 30-33) and lack substantial MTE (18-22) activity (55). On the other hand, we have not yet identified MTE-containing promoters that lack substantial DPE (30-33) activity. In addition, core promoters that are predominantly driven by the novel combination of the 18-22 and 30-33 subregions, which we term the 'Bridge' configuration, have not yet been characterized.

We therefore sought to identify naturally occurring core promoters that are predominantly driven by MTE (18-22 and 27-29) or Bridge (18-22 and 30-33) sequences in conjunction with an Inr. To this end, we searched a database of predicted *Drosophila melanogaster* transcription start sites (kindly provided by Drs. C. Benner and C. K. Glass, UCSD, La Jolla, CA). We found two Inr+MTE core promoters that are more sensitive to mutation of the 18-22 subregion than the 30-33 subregion (Figure 2.8A and 1.8B) as well as an Inr+Bridge core promoter that is more sensitive to mutation of the 18-22 or 30-33 subregions than the 27-29 subregion (Figure 2.8C). For the Inr+MTE promoters (CG5397 and CG6980), we further confirmed the absence of a strong 30-33 subregion by converting it to a consensus 30-33 sequence, and found that the addition of the consensus sequences yielded a 2.5- to 3.7-fold increase in

transcriptional activity (Figure 2.8A and 1.8B). Analogously, we constructed a version of the CG15253 promoter that contains a favorable 27-29 sequence, and observed a two-fold increase in transcription. It thus appears that the Inr+MTE and Inr+Bridge structures are used in natural core promoters.

These findings led us to consider the general co-occurrence of the DPE (27-29 + 30-33), MTE (18-22 + 27-29), and Bridge (18-22 + 30-33) sequences with the Inr in natural core promoters. To this end, we analyzed core promoter sequences derived from the transcription start site data in the *Drosophila* MachiBase database (1). As shown in Table 2.2, the Inr is present in about 27% of all core promoters, but is seen in only about 21% of core promoters that lack all three downstream subregions. On the other hand, the Inr is present in about two-thirds of core promoters that contain MTE (67%), DPE (65%), or Bridge (63%) sequences. In this analysis, we used somewhat stringent criteria for the definition of the Inr, 18-22, 27-28, and 30-33 sequences (see Table 2.2), but similar relative values are observed if the criteria are relaxed. Hence, the presence versus the absence of any of the three binary combinations of the three downstream subregions results in about a threefold (~65% versus ~21%) increase in the occurrence of the Inr. These observations support the notion that the MTE, DPE, and Bridge sequences are functionally important core promoter motifs.

DISCUSSION

In this work, we have analyzed the interaction of TFIID with the MTE and investigated the sequences that are important for MTE function. These studies led to

the identification of three downstream subregions (18-22, 27-29, 30-33) that contribute to core promoter activity. Notably, the three different binary combinations of these subregions create the MTE (18-22 and 27-29), DPE (27-29 and 30-33), and novel 'Bridge' (18-22 and 30-33) core promoter motifs. Thus, these findings have resulted in a broader understanding of the downstream core promoter region (Figure 2.9).

Interaction of TFIID with the downstream core promoter. To study the binding of TFIID to the MTE, we developed a new and reliable method for the purification of *Drosophila* TFIID (Figure 2.1). One key feature of this method is the use of sequence-specific DNA affinity chromatography with the TATA-less, DPE-containing core promoter sequence from the *G* long interspersed nuclear element (LINE). We had previously found that the *G* core promoter has an unusually high affinity for TFIID (43). Therefore, the use of this strong TATA-less core promoter for affinity chromatography enabled the purification of TFIID relative to free TBP and other TBP-containing species.

The DNase I footprinting (Figs. 2 and 3) and photocrosslinking (Figure 2.4) data of this study, combined with the previous analysis of the binding of TFIID to the DPE (12, 71), suggest that the TAF6 and TAF9 subunits of TFIID are in close proximity to the downstream core promoter region. As noted above, TAF6 and TAF9 are related to histones H4 and H3, respectively; thus, TAF6 and TAF9 may form a subcomplex that interacts with downstream core promoter sequences. In addition, examination of the DNase I footprinting data reveals an extended alternating pattern of

DNase I protection and hypersensitivity that suggests a close interaction between promoter DNA and TFIID. A schematic diagram of this model is depicted in Figure 2.9.

Tripartite organization of the downstream core promoter. In the single nucleotide mutational analysis of the MTE, we found that the +27 to +29 region was particularly important for transcriptional activity (Figure 2.5). These experiments were performed in constructs that contained the m30-33 (DPE-inactivating) mutation; thus, the 27-29 region is important for MTE-driven transcription in the absence of a DPE. We then further examined the relation between the 18-22, 27-29, and 30-33 subregions and the MTE and DPE motifs by analyzing all possible combinations of mutations of the subregions in three different promoters (Figure 2.7). These experiments led to the model that the MTE comprises the 18-22 and 27-29 subregions, whereas the DPE contains the 27-29 and 30-33 subregions (Figure 2.9). The *Drosophila E74B* and *Doc* core promoters are examples of natural downstream core promoters that predominantly use the DPE motif (55). The *Drosophila* CG5397 and CG6980 promoters are predominantly MTE-driven downstream core promoters (Figure 2.8A and 1.8B). We also found that the 'Bridge' combination of the 18-22 and 30-33 subregions can yield core promoters with comparable activity to MTE-driven (18-22 and 27-29) or DPE-driven (27-29 and 30-33) core promoters (Figure 2.7). The *Drosophila* CG15253 core promoter is a natural core promoter that is driven predominantly by a Bridge (18-22 and 30-33) motif (Figure 2.8C). These observations suggest that an Inr along with any binary combination of the 18-22, 27-29, and 30-33

subregions can yield a strong core promoter. We additionally determined the general co-occurrence of the Inr with core promoters containing MTE, DPE, or Bridge motifs (Table 2.1). This analysis revealed that the presence of an MTE, DPE, or Bridge motif results in a substantial increase in the frequency of occurrence of an Inr in the core promoter. It is, moreover, interesting to note that the novel Bridge motif appears to occur more frequently than the MTE.

We had previously found that the addition of *Tollo* MTE sequences from +18 to +27 can compensate for the loss of DPE activity (55), and had therefore designated the MTE to be sequences from +18 to +27. In those experiments, however, the constructs contained the favored nucleotide G at position +29 downstream of the *Tollo* MTE sequences. The G+29 probably contributed to the activity of the *Tollo* MTE sequences from +18 to +27. Based on our more extensive analysis of this motif and the importance of the +27 to +29 region (Figure 2.7), it is probably more accurate to describe the MTE as encompassing the sequence from +18 (or +17) to +29 (Figure 2.9).

It is also relevant to compare the MTE and DPE subregions with those of the downstream core element (DCE), which is another downstream core promoter element (47, 51). The DCE was found to consist of three subelements: subelement I is CTTC in the region from +6 to +11; subelement II is CTGT located from +16 to +21; and subelement III is AGC in the region from +30 to +34 (47). Comparison of the DCE sequences with the Motif 10 consensus (63), the MTE favored nucleotides (Figure 2.6), and the DPE consensus (33) reveals no relation between the DCE and the

downstream sequences associated with the MTE and DPE.

The results presented here suggest a tripartite organization of the downstream sequences that contribute to the binding of TFIID to the core promoter (Figure 2.9). These studies clarify the downstream interactions of TFIID with the core promoter. These interactions may be an important feature of the mechanisms by which sequence-specific activators, such as Caudal protein, activate transcription in a core-promoter-specific manner (32). In the future, it is likely that many additional dimensions of complexity will be uncovered in the core promoter. For instance, recent data have revealed transcription systems that do not involve TFIID (for review, see: Ref. 17). These other transcription systems may have their own variations in core promoter sequences and functions, and therefore suggest the existence of a vast range of regulatory phenomena that are mediated via the core promoter.

Table 2.1: Oligonucleotides used to prepare core promoter constructs. Core promoter sequences are represented by uppercase lettering, whereas the sequences used for subcloning of the oligonucleotides into the *Xba* I and *Pst* I restriction sites of pUC119 are denoted by lowercase lettering. The A+1 position in the Inr is shown in bold red lettering. Mutant sequences are indicated by bold black lettering. *Tollo* MTE sequences (+MTE; *Tollo* sequences from +18 to +27) that were used to replace downstream sequences in *E74B* and *Doc* are denoted by blue lettering. The replacement of wild-type sequences with optimal or consensus sequences for the +27-29 or +30-33 regions is depicted in bold fuchsia lettering.

Construct	Sequence
Tollo wild type	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTTCGAGCCGAGCGGACGCGCGTTTGctgca
Tollo m18-22	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTT**ATCCA**CGAGCGGACGCGCGTTTGctgca
Tollo m27-29	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTTCGAGCCGAG**GTA**ACGCGCGTTTGctgca
Tollo m30-33	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTTCGAGCCGAGCGG**CATA**GCGTTTGctgca
Tollo m18-22 m27-29	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTT**ATCCA**CGAG**GTA**ACGCGCGTTTGctgca
Tollo m18-22 m30-33	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTT**ATCCA**CGAGCGG**CATA**GCGTTTGctgca
Tollo m27-29 m30-33	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTTCGAGCCGAG**GTACATA**GCGTTTGctgca
Tollo m18-22 m27-29 m30-33	ctagaGAGAGAGCGATC**A**GTTTTTGAGTTGACTT**ATCCA**CGAG**GTACATA**GCGTTTGctgca
CG10479 wild type	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTTCGAACCGATCGCTCGTATCGCTCTgca
CG10479 m18-22	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTT**ATCCA**CGATCGCTCGTATCGCTCTgca
CG10479 m27-29	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTTCGAACCGAT**GTA**TCGTATCGCTCTgca
CG10479 m30-33	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTTCGAACCGATCGC**CATA**ATCGCTCTgca
CG10479 m18-22 m27-29	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTT**ATCCA**CGAT**GTA**TCGTATCGCTCTgca
CG10479 m18-22 m30-33	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTT**ATCCA**CGATCGC**CATA**ATCGCTCTgca
CG10479 m27-29 m30-33	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTTCGAACCGAT**GTACATA**ATCGCTCTgca
CG10479 m18-22 m27-29 m30-33	ctagaCCCGTGGATTTC**A**TTCCCTCTGCGCACTT**ATCCA**CGAT**GTACATA**ATCGCTCTgca
E74B wild type	ctagaATTTT**A**GTTTTGCTGATACCGTAAGAGATAAATGACGTGCCGCGGctgca
E74B+MTE	ctagaATTTT**A**GTTTTGCTGATACCGTCGAGCCGAGCTGACGTGCCGCGGctgca
E74B+MTE m18-22	ctagaATTTT**A**GTTTTGCTGATACCGT**ATCCA**CGAGCTGACGTGCCGCGGctgca
E74B+MTE m27-29	ctagaATTTT**A**GTTTTGCTGATACCGTCGAGCCGAG**GTA**ACGTGCCGCGGctgca

Table 2.1 (cont.)

E74B+MTE m30-33	ctagaATTTT**A**GTTTTGCTGATACCGTCGAGCCGAGCTG**CATA**GCCGCGGctgca
E74B+MTE m18-22 m27-29	ctagaATTTT**A**GTTTTGCTGATACCGT**ATCCA**CGAG**GTA**ACGTGCCGCGGctgca
E74B+MTE m18-22 m30-33	ctagaATTTT**A**GTTTTGCTGATACCGT**ATCCA**CGAGCTG**CATA**GCCGCGGctgca
E74B+MTE m27-29 m30-33	ctagaATTTT**A**GTTTTGCTGATACCGTCGAGCCGAG**GTACATA**GCCGCGGctgca
E74B+MTE m18-22 m27-29 m30-33	ctagaATTTT**A**GTTTTGCTGATACCGT**ATCCA**CGAG**GTACATA**GCCGCGGctgca
Doc wild type	ctagaGAC**A**TTCGGCATTCCACAGTCTTCGGGTGGAGACGTGTTTCTTctgca
Doc+MTE	ctagaGAC**A**TTCGGCATTCCACAGTCGAGCCGAGCAGACGTGTTTCTTctgca
Doc+MTE m18-22	ctagaGAC**A**TTCGGCATTCCACAGT**ATCCA**CGAGCAGACGTGTTTCTTctgca
Doc+MTE m27-29	ctagaGAC**A**TTCGGCATTCCACAGTCGAGCCGAG**GTA**ACGTGTTTCTTctgca
Doc+MTE m30-33	ctagaGAC**A**TTCGGCATTCCACAGTCGAGCCGAGCAG**CATA**GTTTCTTctgca
Doc+MTE m18-22 m27-29	ctagaGAC**A**TTCGGCATTCCACAGT**ATCCA**CGAG**GTA**ACGTGTTTCTTctgca
Doc+MTE m18-22 m30-33	ctagaGAC**A**TTCGGCATTCCACAGT**ATCCA**CGAGCAG**CATA**GTTTCTTctgca
Doc+MTE m27-29 m30-33	ctagaGAC**A**TTCGGCATTCCACAGTCGAGCCGAG**GTACATA**GTTTCTTctgca
Doc+MTE m18-22 m27-29 m30-33	ctagaGAC**A**TTCGGCATTCCACAGT**ATCCA**CGAG**GTACATA**GTTTCTTctgca
CG15253 wild type	ctagaCTGCCCAAAC**A**GTATTGAATGGGAAGTCTAGCAGCCAATATGTGCTATTActgca
CG15253 m18-22	ctagaCTGCCCAAAC**A**GTATTGAATGGGAAGT**ATCCA**AGCCAATATGTGCTATTActgca
CG15253 m27-27	ctagaCTGCCCAAAC**A**GTATTGAATGGGAAGTCTAGCAGCC**GTA**ATGTGCTATTActgca
CG15253 m30-33	ctagaCTGCCCAAAC**A**GTATTGAATGGGAAGTCTAGCAGCCAAT**CATA**GCTATTActgca
CG15253 optimal 27-29	ctagaCTGCCCAAAC**A**GTATTGAATGGGAAGTCTAGCAGCC**CGG**ATGTGCTATTActgca
CG5397 wild type	ctagaGCCAGAGCTC**A**GTTCCCAGTTGACAGTCGAGCGAGGCGAAATGCAATCTCctgca
CG5397 m18-22	ctagaGCCAGAGCTC**A**GTTCCCAGTTGACAGT**ATCCA**GAGGCGAAATGCAATCTCctgca
CG5397 m30-33	ctagaGCCAGAGCTC**A**GTTCCCAGTTGACAGTCGAGCGAGGCGA**CATA**CAATCTCctgca
CG5397 consensus 30-33	ctagaGCCAGAGCTC**A**GTTCCCAGTTGACAGTCGAGCGAGGCGA**ACGT**CAATCTCctgca
CG6980 wild type	ctagaCTTAAAAGAC**A**TTCTGTTCTGAGCAGTCGAACCAGTCGGTATCTTAACTActgca
CG6980 m18-22	ctagaCTTAAAAGAC**A**TTCTGTTCTGAGCAGT**ATCCA**CAGTCGGTATCTTAACTActgca
CG6980 m30-33	ctagaCTTAAAAGAC**A**TTCTGTTCTGAGCAGTCGAACCAGTCGG**CATA**TTAACTActgca
CG6980 consensus 30-33	ctagaCTTAAAAGAC**A**TTCTGTTCTGAGCAGTCGAACCAGTCGG**ACGT**TTAACTActgca

Figure 2.1: Purification of TFIID from *Drosophila* S2 cells containing FLAG-tagged TBP. (A) Synthesis of FLAG-tagged TBP in S2 cells. Whole cell lysates derived from S2 cells and two different FLAG-TBP-containing S2 cell lines were subjected to western blot analysis with antibodies against *Drosophila* TBP. (B) Scheme for the purification of TFIID from S2 cells containing FLAG-tagged TBP. (C) Purification of TFIID containing FLAG-tagged TBP. The polypeptides were resolved by 10% polyacrylamide-SDS gel electrophoresis and visualized by silver staining. In addition, by western blot and mass spectrometry, the purified TFIID was found to contain FLAG-tagged TBP as well as all TAFs from TAF1 through TAF14.

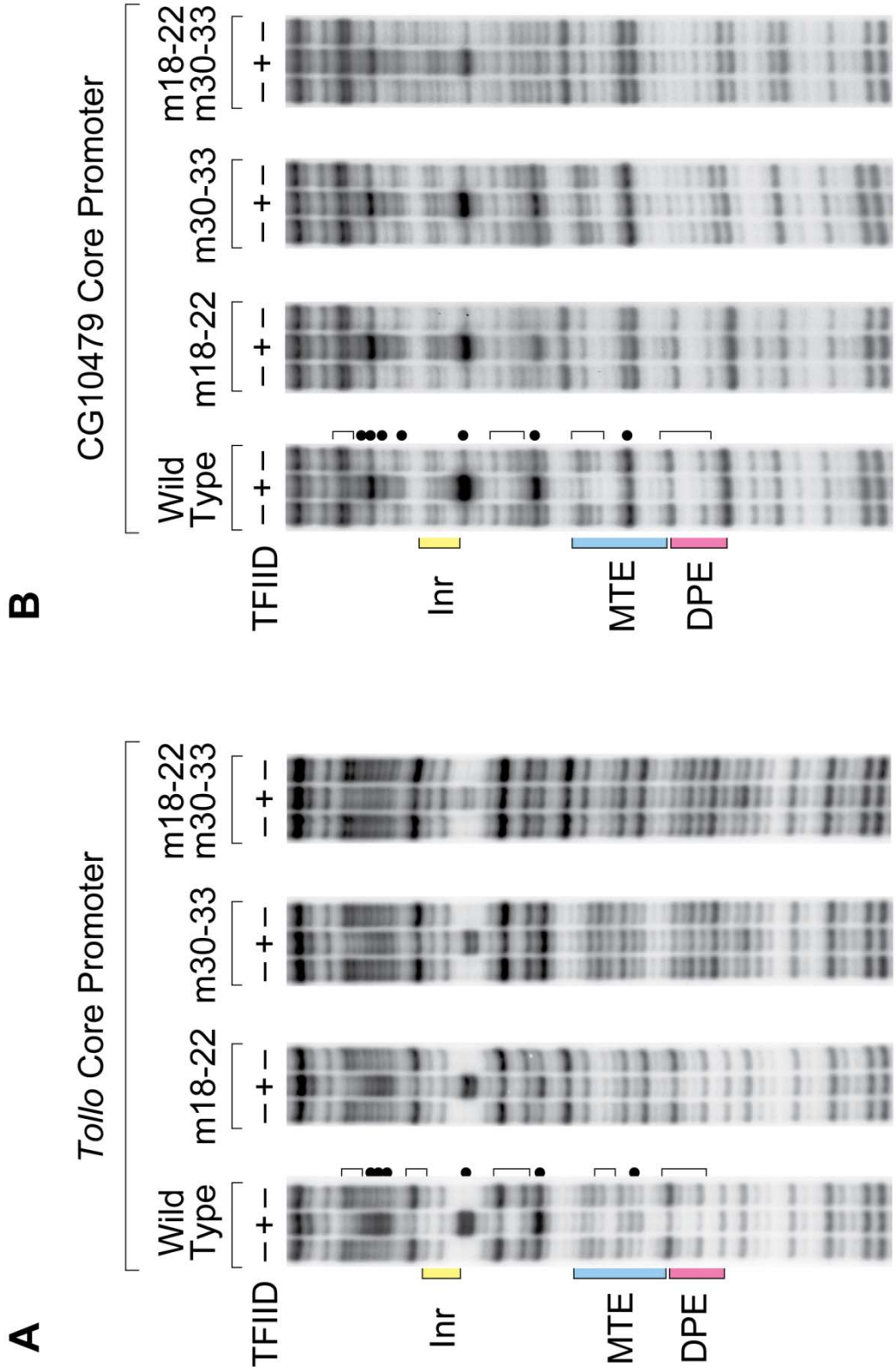
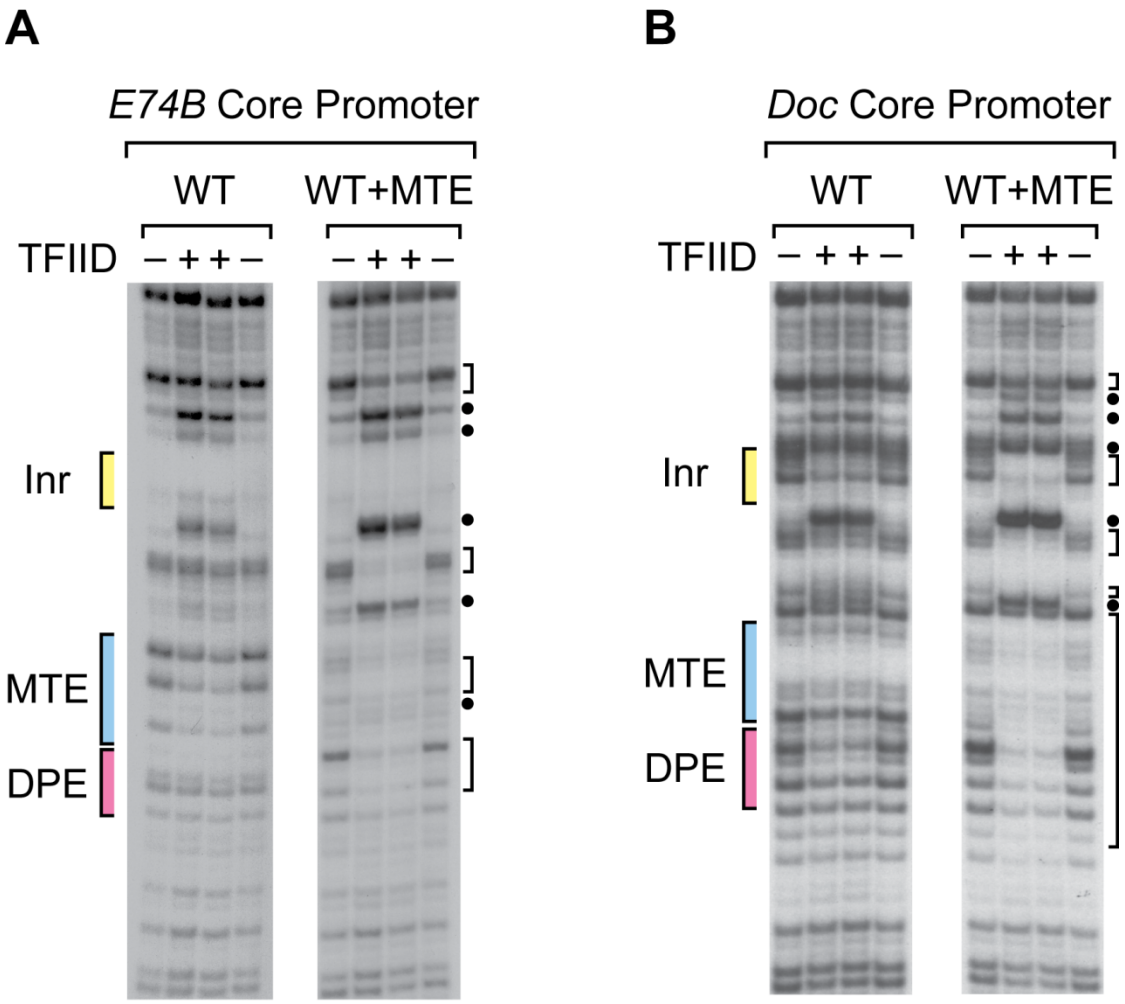

A
Tollo Core Promoter
Wild Type
m18-22
m30-33
m18-22 m30-33
TFIID
– + – – + – – + – – + –
Inr
MTE
DPE
B
CG10479 Core Promoter
Wild Type
m18-22
m30-33
m18-22 m30-33
TFIID
– + – – + – – + – – + –
Inr
MTE
DPE

Figure 2.3: The addition of an MTE increases the affinity of TFIID for the core promoter. The wild-type (WT) and WT+MTE (containing the *Tollo* MTE sequence from +18 to +27, relative to the A+1 position in the Inr) versions of the *Drosophila E74B* (A) and *Doc* (B) core promoters were subjected to DNase I footprinting analysis with purified *Drosophila* TFIID. The positions of the Inr (−2 to +4), MTE (*Tollo* core promoter sequence from +18 to +27), and DPE (+28 to +33) motifs are indicated. In the WT+MTE promoters, regions of DNase I protection and hypersensitivity are indicated by brackets and solid dots, respectively. (C) Schematic of core promoter constructs.


A
E74B Core Promoter
WT
WT+MTE
TFIID
– + + –
– + + –
Inr
MTE
DPE
B
Doc Core Promoter
WT
WT+MTE
TFIID
– + + –
– + + –
Inr
MTE
DPE

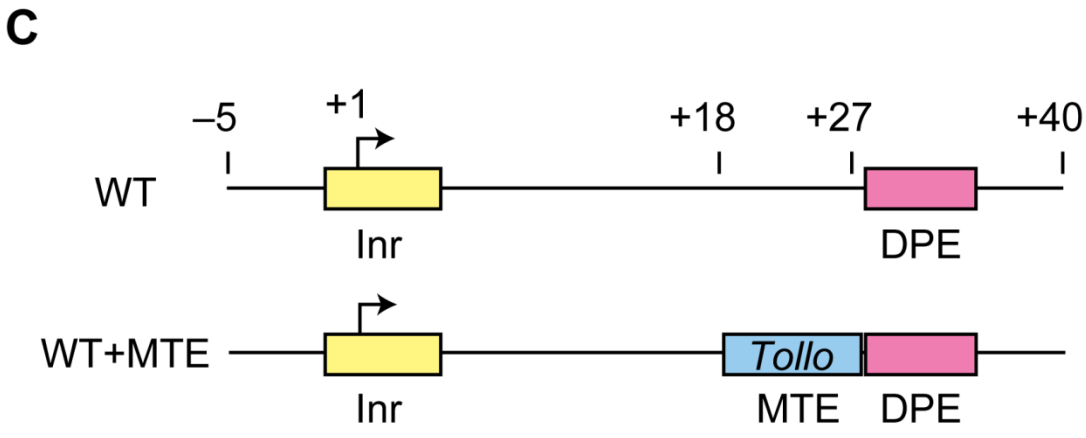

C
–5
+1
+18
+27
+40
WT
Inr
DPE
WT+MTE
Tollo
Inr
MTE
DPE

Figure 2.4: The TAF6 and TAF9 subunits of TFIID appear to be in close proximity to the MTE. (A) Diagram of photoaffinity probes containing AB-dUMP at the +20, +25, or +30 position relative to the +1 transcription start site. The *Tollo* core promoter sequences are shown. The diagram is roughly to scale. (B) Photocrosslinking of purified TFIID with the *Tollo* core promoter. Reactions were performed in the presence or absence of TFIID as well as with or without UV irradiation, as indicated. (C) Photocrosslinking of purified TFIID with the *E74B* core promoter containing the *Tollo* MTE. The *E74B*+MTE core promoter is identical to the construct that was used in the DNase I footprinting analysis shown in Figure 2.3A. Reactions were performed as in (B).

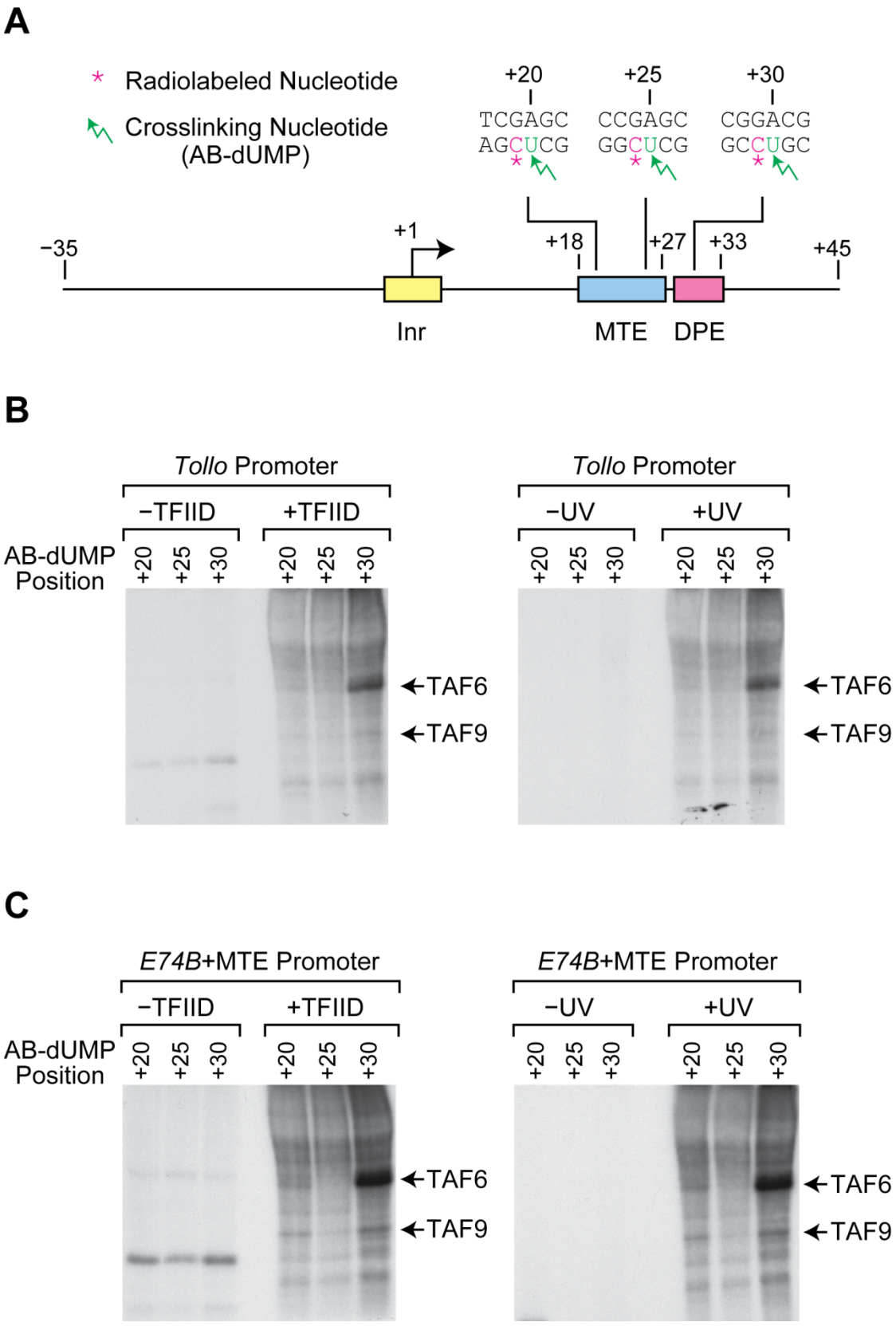
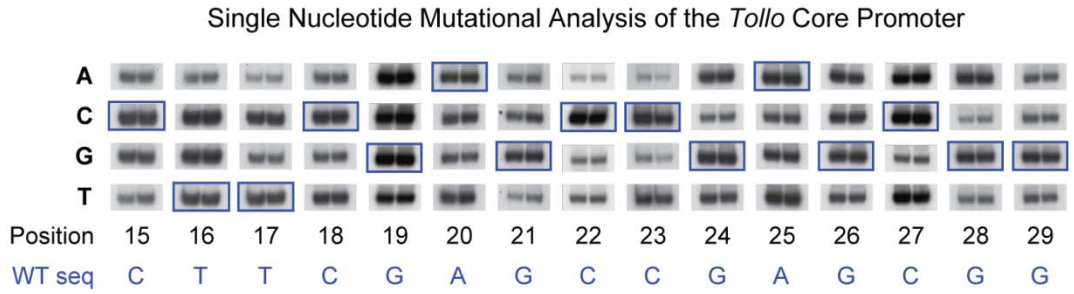
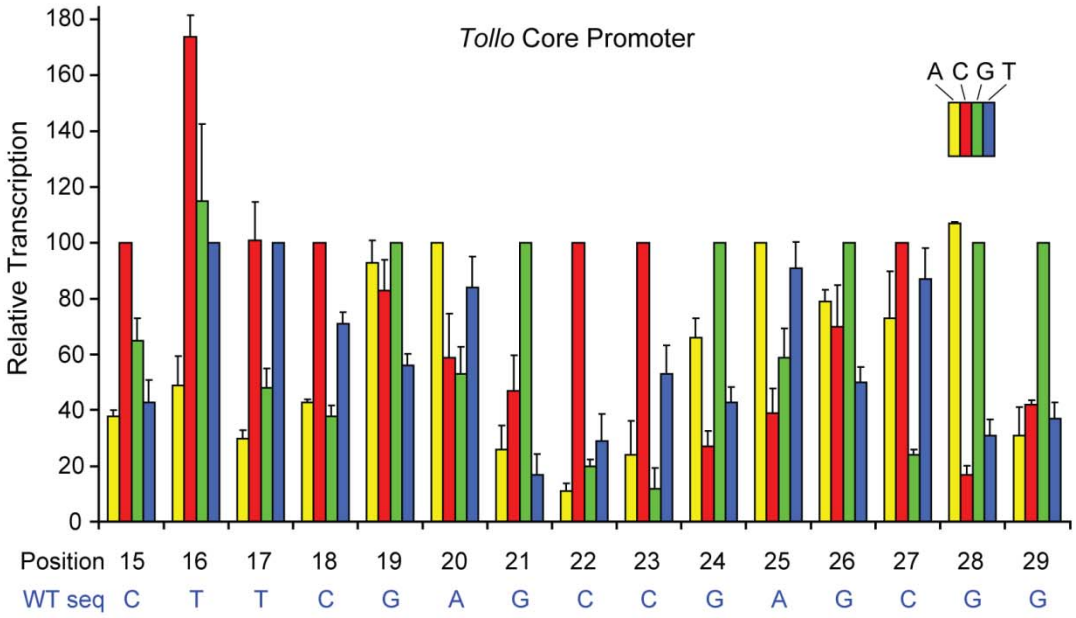

A
Radiolabeled Nucleotide
Crosslinking Nucleotide
(AB-dUMP)
+20
+25
+30
TCGAGC
AGCUCG
CCGAGC
GGCUCG
CGGACG
GCCUGC
−35
+1
+18
+27
+33
+45
Inr
MTE
DPE
B
Tollo Promoter
−TFIID
+TFIID
AB-dUMP Position
+20
+25
+30
TAF6
TAF9
Tollo Promoter
−UV
+UV
C
E74B+MTE Promoter
−TFIID
+TFIID
E74B+MTE Promoter
−UV
+UV

Figure 2.5: Single nucleotide substitution analysis reveals sequences that are important for MTE activity. (A,C) Single nucleotide substitution analysis of the MTE in the *Drosophila Tollo* (A) and CG10479 (C) core promoters. Mutant promoters containing every possible single nucleotide substitution from +15 to +29 (relative to the A+1 in the Inr) were generated. To eliminate the contribution of the DPE in these experiments, all of the constructs contain the m30-33 mutation (CATA from +30 to +33), which inactivates DPE function. The promoters were subjected to in vitro transcription analysis with a *Drosophila* embryo nuclear extract, and the resulting transcripts were detected by primer extension-reverse transcription analysis. The primer extension data that correspond to the wild-type promoter are boxed. (B,D) The relative transcriptional activities of wild-type and mutant MTE sequences in the *Tollo* (B) and CG10479 (D) core promoters. Quantitation of the data from three independent experiments is shown. The data are normalized to promoters containing the wild-type MTE sequence. The error bars represent the standard deviation.

A


Single Nucleotide Mutational Analysis of the *Tollo* Core Promoter
A
C
G
T
Position 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
WT seq C T T C G A G C C G A G C G G

B


Tollo Core Promoter
A C G T
Relative Transcription
180
160
140
120
100
80
60
40
20
0
Position 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
WT seq C T T C G A G C C G A G C G G

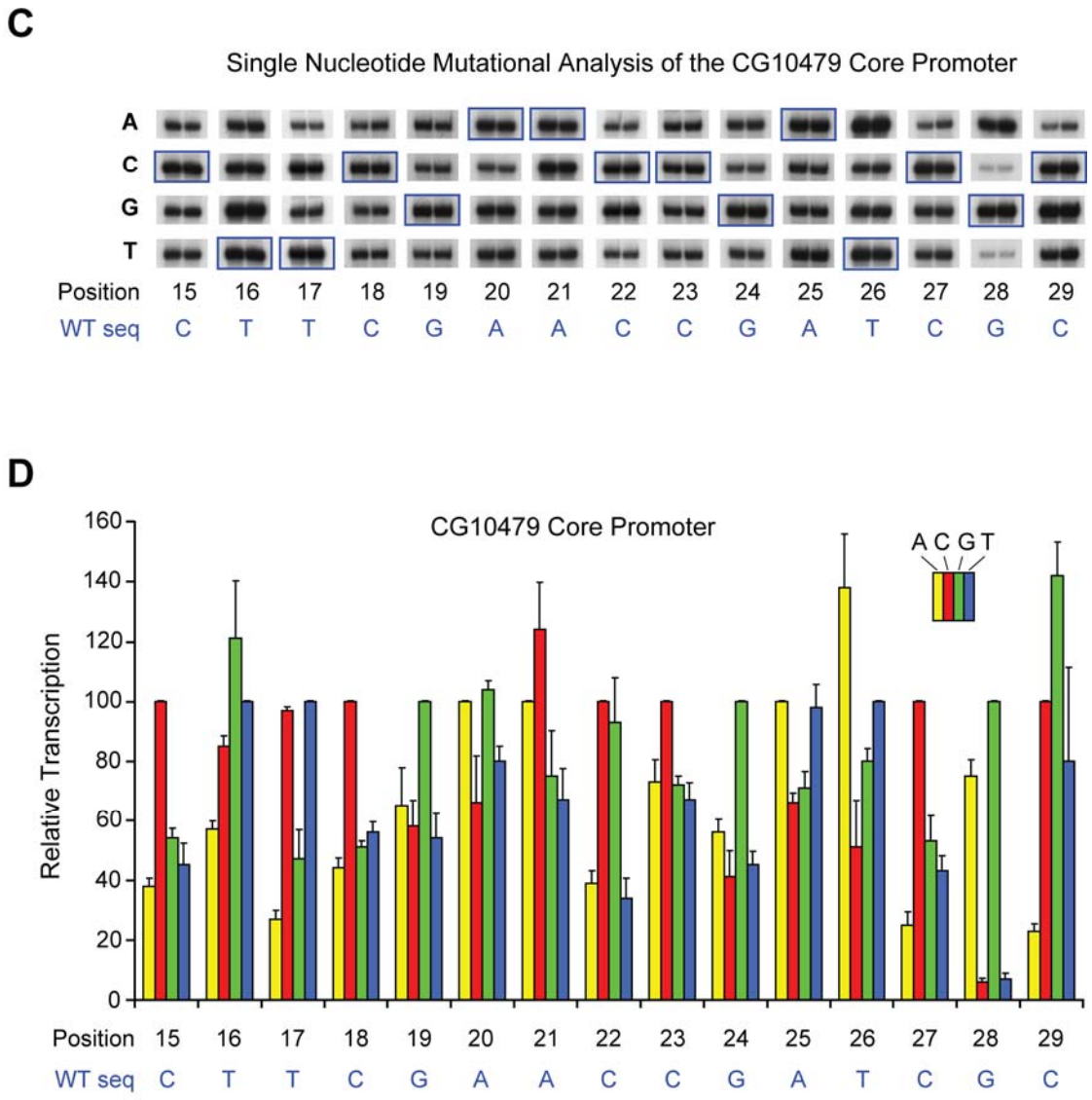


Figure 2.5 (cont.)

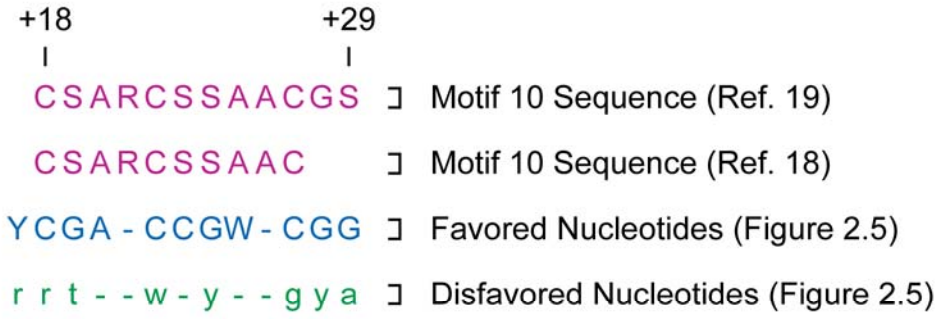
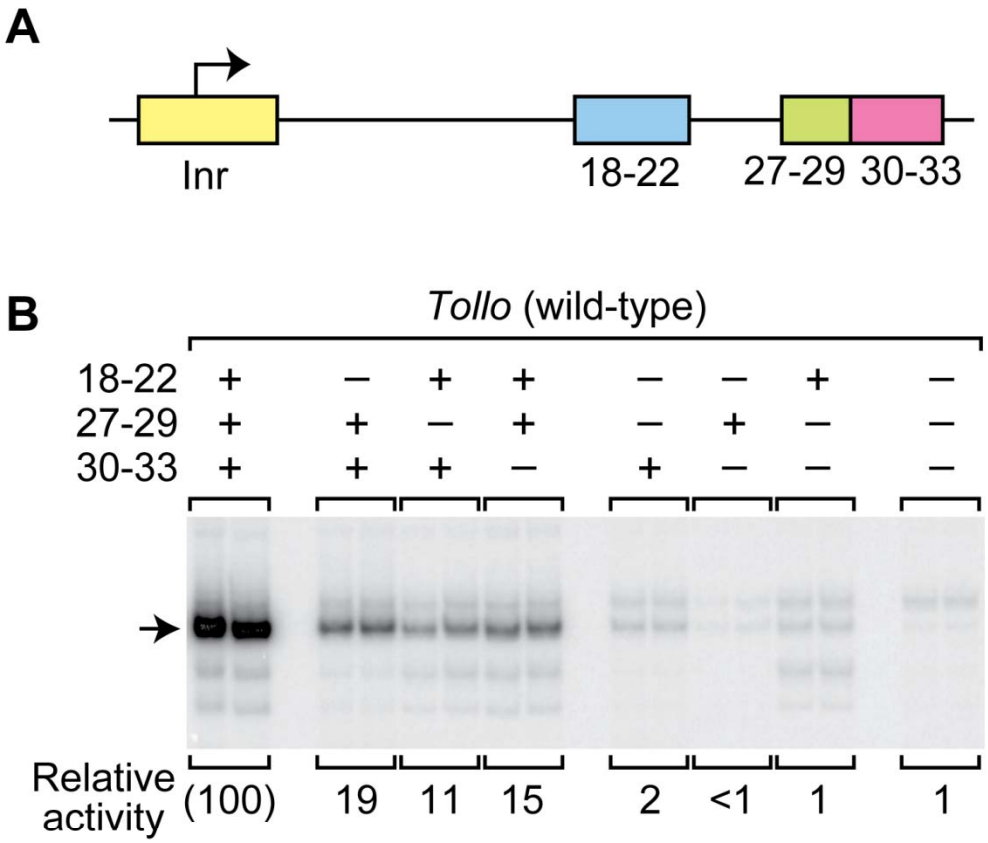


Figure 2.6: Identification of sequences that are important for MTE activity. The data in Figure 2.5 were analyzed as follows. For each position, nucleotides that result in >90% of the transcriptional activity of the wild-type nucleotide (defined to be 100%) for both the *Tollo* and CG10479 promoters were designated as 'Favored Nucleotides', whereas nucleotides that result in <60% of the transcriptional activity of the wild-type nucleotide for both promoters were designated as 'Disfavored Nucleotides'. The computationally derived motif 10 sequence (40) and the MTE sequence based on the initial characterization of the element (35) are also shown.

Figure 2.7: Analysis of three downstream subregions that are important for MTE and DPE activity. (A) Diagram of the Inr and downstream core promoter subregions. The relative locations of the elements are drawn approximately to scale. (B) Systematic analysis of three downstream subregions of the wild-type *Tollo* core promoter. A set of promoters with all possible combinations of the m18-22, m27-29, and m30-33 mutations in the *Tollo* core promoter was constructed. For each promoter, the wild-type (+) and mutant (−) versions of each sequence are indicated. The numbers below each promoter construct are the mean of three or four independent experiments relative to the wild-type *Tollo* promoter. (C,D) Analysis of three downstream subregions in *E74B*- and *Doc*-based core promoters. The relative contributions of the 18-22, 27-29, and 30-33 sequences were tested by using the hybrid *E74B*+MTE and *Doc*+MTE core promoters, which consist of the natural E74B and Doc promoters containing the *Tollo* MTE sequences from +18 to +27 relative to the A+1 in the Inr. Thus, the hybrid promoters contain Inr, MTE, and DPE motifs. Beginning with each hybrid promoter, a set of promoters that comprises all possible combinations of the m18-22, m27-29, and m30-33 mutations was constructed. For each promoter, the wild-type (+) and mutant (−) versions of each sequence are indicated. The numbers below each promoter construct indicate the mean of three or four independent experiments relative to the hybrid core promoter containing the optimal sequences in the Inr, MTE, and DPE motifs.


A
Inr
18-22
27-29
30-33
B
Tollo (wild-type)
18-22 + – + + – – + –
27-29 + + – + – + – –
30-33 + + + – + – – –
Relative activity
(100)
19
11
15
2
<1
1
1

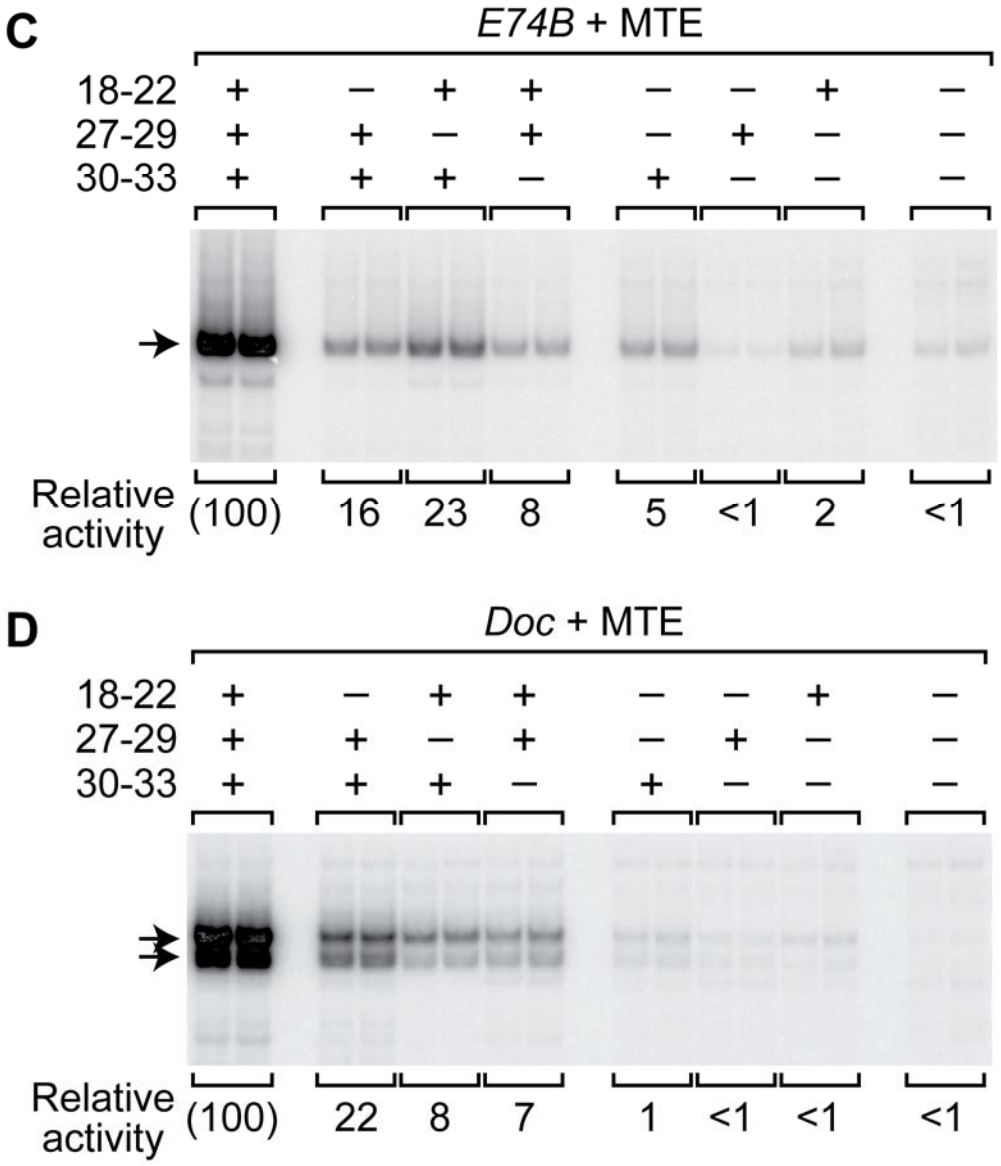


Figure 2.7 (cont.)

Figure 2.8: Natural core promoters that are driven predominantly by the MTE or Bridge core promoter motifs. (A,B) Natural TATA-less, MTE-containing core promoters that lack a strong DPE motif. Wild-type and mutant versions of the CG5397 (A) and CG6980 (B) core promoters were subjected to in vitro transcription and primer extension analyses. In addition to the standard m18-22 and m30-33 mutations, we analyzed promoters in which a consensus DPE sequence was introduced from +30 to +33 (Consensus 30-33). The relative activity values are the mean of three or four independent experiments normalized to the cognate wild-type promoter. (C) Natural TATA-less, Bridge (18-22 + 30-33)-containing core promoter that lacks a strong +27-29 sequence. Wild-type and mutant versions of the CG15253 core promoter were subjected to in vitro transcription and primer extension analyses. In addition to the m18-22, m27-29, and m30-33 mutations, we tested a mutant promoter containing an optimal 27-29 sequence (Optimal 27-29; see Figure 2.6). The relative activity values are the mean of three or four independent experiments normalized to the cognate wild-type promoter.

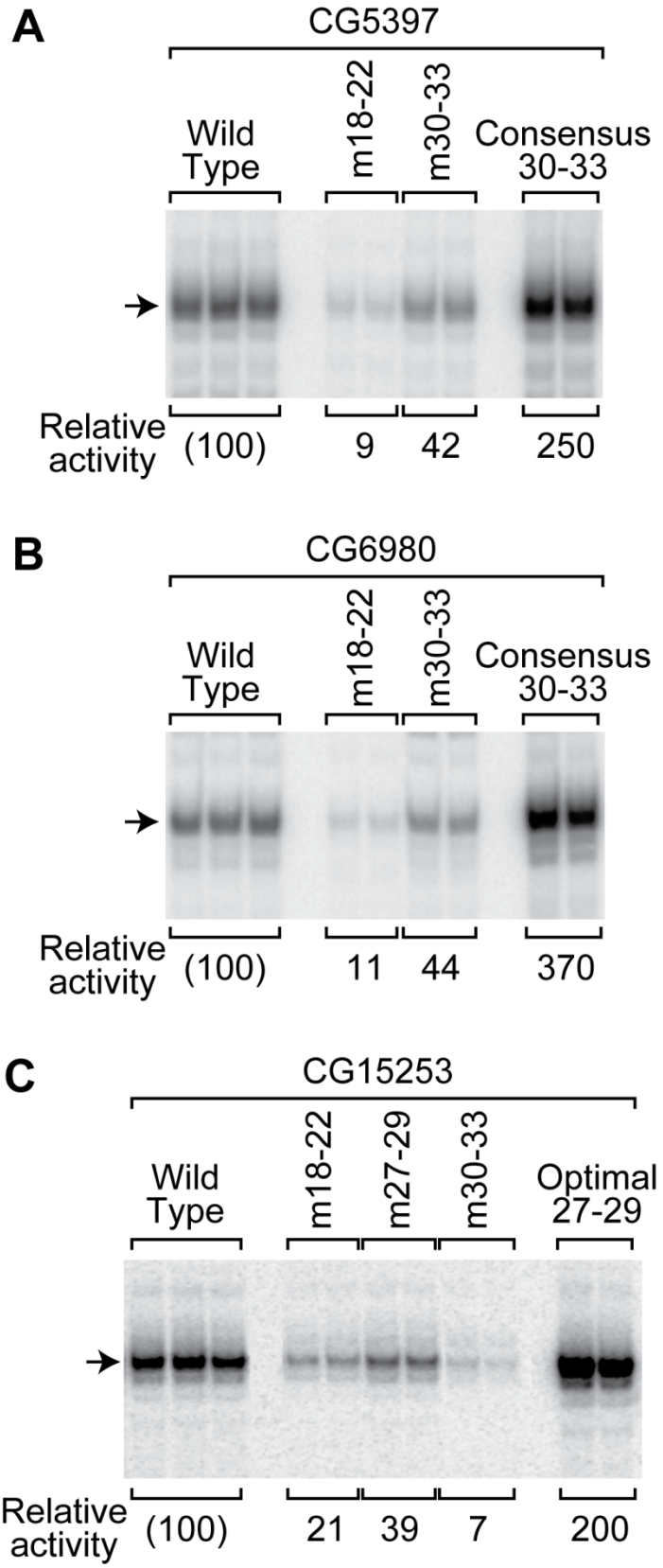

A
CG5397
Wild Type
m18-22
m30-33
Consensus 30-33
Relative activity
(100)
9
42
250
B
CG6980
Wild Type
m18-22
m30-33
Consensus 30-33
Relative activity
(100)
11
44
370
C
CG15253
Wild Type
m18-22
m27-29
m30-33
Optimal 27-29
Relative activity
(100)
21
39
7
200

Table 2.2: Co-occurrence of downstream promoter elements with the initiator (Inr) in *Drosophila melanogaster*. [a] The presence of an Inr motif was based on at least a 5/6 match with the Inr consensus of TCA+1KTY while not allowing for deviation from CA+1. [b] Core promoter sequences were based on the transcription start site data from the *Drosophila* MachiBase database (1). [c] Promoters that lack 18-22, 27-29, and 30-33 downstream core promoter subregions. The presence of the 18-22 subregion was based on at least a 3/4 match to the favored nucleotides in Fig. 6 while not allowing for any of the disfavored nucleotides in this region. The presence of the 27-29 subregion was based on a 3/3 match to the favored nucleotides in Fig. 6. The presence of the 30-33 subregion was based on at least a 3/4 match to the favored nucleotides WYGT while not allowing for any disfavored nucleotides (SRW from +30 to +32). [d] Promoters that contain the 18-22 and 27-29 subregions, but lack the 30-33 subregion. [e] Promoters that contain the 27-29 and 30-33 subregions, but lack the18-22 subregion. [f] Promoters that contain the 18-22 and 30-33 subregions, but lack the 27-29 subregion.

Core Promoter Class	Fraction with Inr[a]	Percent with Inr[a]
All core promoters[b]	3122/11638	27
No downstream subregions[c]	1880/8930	21
MTE (18-22 & 27-29, but not 30-33)[d]	49/73	67
DPE (27-29 & 30-33, but not 18-22)[e]	141/217	65
Bridge (18-22 & 30-33, but not 27-29)[f]	99/157	63

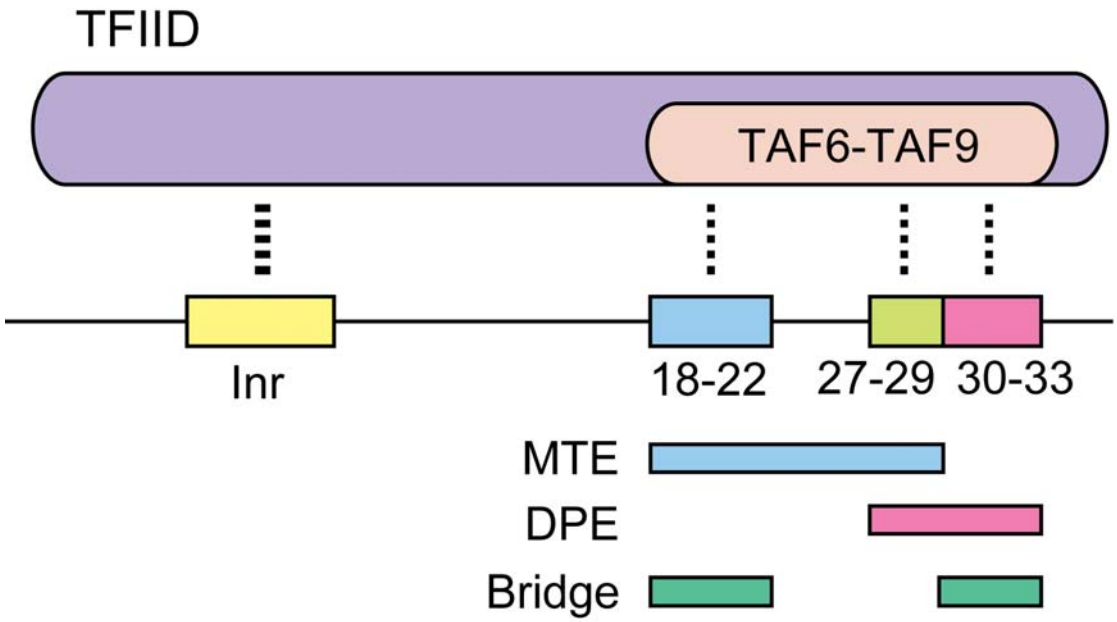


Figure 2.9: Model of a tripartite organization of key interaction points of TFIID with downstream core promoter sequences. In this model, the 18-22, 27-29, and 30-33 sequences are three key downstream points of interaction of TFIID with the core promoter. Based on the photocrosslinking studies (Figure 2.4), the TAF6 and TAF9 subunits of TFIID are shown in close proximity to the downstream core promoter region. The Inr is included because it has been previously shown that the MTE as well as the DPE act in a cooperative manner with the Inr (13, 55). The MTE comprises the 18-22 and 27-29 sequences, whereas the DPE contains the 27-29 and 30-33 sequences. In addition, the 18-22 and 30-33 sequences can function synergistically in the absence of an optimal 27-29 sequence and form the Bridge element.

ACKNOWLEDGMENTS

We are very grateful to Uwe Ohler, Tammy Juven-Gershon, Timur Yusufzai, and Sharon Torigoe for critical reading of this manuscript. We thank Drs. Chris Benner and Chris Glass (University of California, San Diego; La Jolla, CA) for sharing the *Drosophila* transcription start site data. We also thank Drs. George Kassavetis and E. Peter Geiduschek (University of California, San Diego; La Jolla, CA) for the gift of the AB-dUTP photocrosslinking reagent. This work was supported by a grant from the National Institutes of Health (GM041249) to J.T.K.

This chapter , in full, has been accepted for publication as: Theisen, J. W. M., C. Y. Lim, and J. T. Kadonaga. 2010. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Molecular and Cellular Biology*, in press..

Chapter 3

Repression of Transcription by MeCP2 Using a Histone Deacetylase-independent Mechanism

ABSTRACT

MeCP2 is a methyl-CpG-binding protein and transcriptional repressor that is required for normal neurologic function. MeCP2 dysfunction causes the neurologic disorder Rett syndrome. To gain a better understanding of MeCP2 function, we examined the mechanism of transcriptional repression by MeCP2 and the role that the number and location of methyl-CpG dinucleotides plays in directing repression. These studies identify a novel, histone deacetylase-independent mechanism of repression. This mechanism involves inhibition of pre-initiation complex formation and is not dependent on a previously identified transcription repression domain. Analysis of transcription templates with specifically placed CpG dinucleotides reveals a general methylation code, whereby constructs with more methylation sites or with sites closer to the transcription start site are subject to stronger repression by MeCP2. These findings show that MeCP2 has a second mechanism of transcriptional repression and highlight the importance of methylation patterns in directing MeCP2 function.

INTRODUCTION

Methyl-CpG-binding protein 2 (MeCP2) is the founding member of the methyl-CpG-binding domain (MBD) family of proteins (58). Mutations in MeCP2 cause Rett syndrome (3), an X-linked dominant neurodevelopmental disorder characterized by apparently normal development until 6-18 months of age, at which point the disorder manifests (24). Rett syndrome patients exhibit developmental regression, seizures and autistic behaviors. Approximately 80% of Rett syndrome

cases are caused by mutations in MeCP2 (11). MeCP2 is a 53 kDa, monomeric protein. It binds to DNA containing methy-CpG dinucleotides and can repress transcription. These activities are localized to an N-terminal MBD and a centrally located transcriptional repression domain (TRD) (60, 61), respectively. While these domains have been well studied, we still do not fully understand how MeCP2 selects target genes and represses transcription.

Most CpG dinucleotides are methylated in mammalian genomes (9), and MeCP2 can bind CpG-methylated DNA in any sequence context (52). These findings led to the prediction that MeCP2 might act as a global repressor of transcription. Yet, numerous lines of investigation have shown that MeCP2 acts selectively on a subset of genes (15, 37, 84). The determinants of this selectivity have not been established.

The mechanism of transcriptional repression by MeCP2 is also not entirely clear. MeCP2 has been shown to co-purify with RNA-, DNA- and chromatin-modifying activities (25, 30, 36, 62, 85). The best characterized of these interactions is the binding of MeCP2 to Sin3A-histone deacetylase (HDAC) complexes (30, 62). Consistent with this observation, MeCP2-mediated repression of transcription in cells is partially relieved by treatment with HDAC inhibitors. However, a significant component of this repression is HDAC-independent, and the majority of MeCP2 is not associated with HDAC-containing complexes in vivo (38). To date, the mechanism of HDAC-independent repression is completely unknown.

In order to study both the determinants of target selection by MeCP2 and the mechanisms of transcription repression, we have established an in vitro transcription

system in which repression of transcription by MeCP2 is dependent on CpG methylation. Using this system, we observe potent, HDAC-independent repression of transcription by MeCP2. We find that this repression occurs via inhibition of transcription pre-initiation complex (PIC) formation. We also show that this repression does not depend on binding of MeCP2 to the core promoter region. Finally, strong repression of transcription by MeCP2 requires the presence of multiple methyl-CpG dinucleotides.

MATERIALS AND METHODS

Purification of recombinant MeCP2. Recombinant wild-type and truncated MeCP2 proteins were expressed and purified as previously described (86), with minor modifications. Briefly, chitin binding domain (CBD) fusion proteins were expressed in bacteria overnight at 16 °C. Bacteria were lysed by sonication and the extracts were bound to chitin beads (New England Biolabs). The fusion proteins were cleaved from the beads by overnight incubation in buffer containing 50 mM DTT. Column eluate was dialyzed twice for 1 hour against KSB [25 mM HEPES (pH 7.6), 400 mM potassium acetate, 6 mM $MgCl_2$, 20% (v/v) glycerol, 1 mM DTT, 1 mM benzamidine-HCl, 0.2 mM PMSF]. The purity and concentration of the proteins was estimated by SDS-PAGE and comparison to bovine serum albumin standards.

Preparation of HeLa nuclear extract. The HeLa nuclear extract used in transcription reactions was prepared as follows. All operations were performed at 4 °C. HeLa-S3 cell pellets from 12 L of culture (approximately 50 mL of cells; from

National Cell Culture Center) were washed with 150 mL PBS + 5 mM $MgCl_2$. Cells were pelleted by centrifugation in a GSA rotor at 3000 rpm for 10 min. The cell pellet was resuspended in 150 mL PBS + 5 mM $MgCl_2$ and pelleted a second time by centrifugation in a GSA rotor at 3000 rpm for 10 min. The final pellet was resuspended in 60 mL Buffer H [10 mM Tris-HCl (pH 7.9), 10 mM KCl, 750 μM spermidine, 150 μM spermine, 0.1 mM EDTA, 0.1 mM EGTA]. Cells were incubated on ice for 20 min and then lysed by 20 strokes with a 40 mL Dounce using the B (loose) pestle. Nuclei were pelleted by centrifugation in an SS-34 rotor at 8500 rpm for 10 min. Pelleted nuclei were resuspended in 60 mL Buffer H and pelleted a second time by centrifugation in an SS-34 rotor at 8500 rpm for 10 min. Pelleted nuclei were resuspended in 40 mL Buffer AB [15 mM HEPES (pH 7.6), 110 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 1 mM benzamidine-HCl, 0.2 mM PMSF, 1 mM sodium bisulfite]. Clumps of nuclei were dispersed by 3 strokes with a 40 mL Dounce using the B (loose) pestle. Another 60 mL of Buffer AB was added to the dispersed nuclei, bringing the total volume to approximately 110 mL. Nuclei were then lysed by addition of $1/10^{th}$ volume (approximately 11 mL) of 4 M ammonium sulfate and incubation for 20 min on a rotating wheel. Nuclear debris was pelleted by ultracentrifugation in a Ti45 rotor at 35,000 rpm for 60 min. The supernatant was transferred to a 250 mL beaker and 0.3 g of pulverized ammonium sulfate was added per 1 mL of supernatant. The solution was stirred for 15 min and then precipitated proteins were pelleted by centrifugation in an SS-34 rotor at 15,000 rpm for 20 min. Protein pellets were resuspended in 5 mL HEG Buffer [25 mM HEPES (pH 7.6),

0.1 mM EDTA, 10% (v/v) glycerol] containing 0.1 M KCl, 1 mM DTT, 1 mM benzamidine-HCl, 0.2 mM PMSF and 1 mM sodium bisulfite. Proteins were dispersed by 10 strokes with a 15 mL Dounce using the A (tight) pestle. Insoluble debris was pelleted by centrifugation in an SS-34 rotor at 10,000 rpm for 10 min. The supernatant was then dialyzed three times for 50 min against 2 L HEG containing 0.1 M KCl, 1 mM DTT, 1 mM benzamidine-HCl, 0.2 mM PMSF and 1 mM sodium bisulfite. Insoluble debris was pelleted by centrifugation in an SS-34 rotor at 10,000 rpm for 10 min. The supernatant was then aliquoted, flash frozen and stored at −80 °C.

Generation of the pCpG-Txn plasmid. The pCpG-Txn plasmid was derived from the pCpG-Luc plasmid (Invivogen, San Diego, CA) as follows. pCpG-Luc was digested with *Pst* I to remove the mCMV enhancer and hEF1 promoter. The resulting plasmid was then digested with *Pci* I, blunted with Klenow and religated to destroy the *Pci* I restriction site. The resulting plasmid was then modified by the insertion of double-stranded oligonucleotides into the *Pst* I and *Xba* I sites. This destroyed the existing multiple cloning site and generated a new proximal multiple cloning site (MCSp). The resulting plasmid was then modified by the insertion of double-stranded oligonucleotides into the *Sph* I site. This created a distal multiple cloning site (MCSd), and completed the construction of pCpG-Txn. The sequences of the MCSp and MCSd inserts are detailed in Table 3.1.

Core promoter constructs. Plasmids containing variations of the Super Core Promoter were constructed by the insertion of double-stranded oligonucleotides into the *Xba* I and *Pst* I sites of the multiple cloning region of pUC119 or the MCSp of

pCpG-Txn. The pCpG-SCPX-0CpG-12CpG-US was constructed by insertion of double-stranded oligonucleotides into the *Pci* I and *Bgl* II sites of the MCSp of pCpG-SCPX-0CpG. The sequences of all of the SCPX variations are in Table 3.2. The sequence of the 12CpG-US insert is CATGTCGAACGAGATCTCGAACGTACGA-ACGTTCGAACGATCGTACGATCGAACGTACGT.

In vitro transcription. In vitro transcription reactions were carried out as follows. For reactions involving pUC119-based templates, 250 ng of template and 750 ng of unmethylated pUC119 competitor were included in each reaction. For reactions involving pCpG-Txn-based templates, 250 ng of template, 750 ng of unmethylated pCpG-Txn competitor and 250 ng of methylated pUC119 competitor were included in each reaction. MeCP2 was incubated at 30 °C with template and competitor DNA for 20 min in a total volume of 36 uL in buffer containing 20 mM HEPES (pH 8.0), 69 mM potassium acetate, 8 mM $MgCl_2$, 3.5% (w/v) polyvinyl alcohol, 4.2 mM ATP and 3.5% (v/v) glycerol. After this incubation, 10 uL of HeLa nuclear extract diluted in HEG containing 0.1 M KCl was added and the reaction was incubated at 30 °C for 90 min in a total volume of 46 uL in buffer containing 21 mM HEPES (pH 8.0), 22 mM KCl, 54 mM potassium acetate, 6.5 mM $MgCl_2$, 2.7% (w/v) polyvinyl alcohol, 3.3 mM ATP and 5% (v/v) glycerol. After this incubation, 4 uL of 5 mM rNTP solution was added and the reaction was incubated at 30 °C for 20 min in a final volume of 50 uL in buffer containing a final concentration of 20 mM HEPES (pH 8.0), 20 mM KCl, 50 mM potassium acetate, 6 mM $MgCl_2$, 2.5% (w/v) polyvinyl alcohol, 3.4 mM ATP, 0.4 mM CTP, GTP and TTP, and 4.5% (v/v) glycerol. In

reactions that included Sarkosyl, initial buffer volumes were adjusted to allow the addition of 2 uL of 5% (v/v) Sarkosyl for a final concentration of 0.2% (v/v) Sarkosyl. In reactions that included trichostatin A (TSA), initial buffer volumes were adjusted to allow the addition of 2 uL of TSA solution at 25 times the desired final concentration. Reactions were stopped by the addition of 100 uL of Stop Buffer [20 mM EDTA, 200 mM sodium chloride, 1% (w/v) sodium dodecyl sulfate, 0.3 mg/mL glycogen] and 5 uL of 2.5 mg/mL proteinase K and incubated for 20 min at 37 °C. Transcripts were subjected to primer extension analysis as previously described (34) using 5'-^{32}P-labeled M13 reverse sequencing primer (AGCGGATAACAATTTCACACAGGA) for pUC119-based constructs or 5'-^{32}P-labeled pCpG-PE1 primer (GGAAAGAGAAGA-AGGTTAGTACAATTGT) for pCpG-Txn based constructs. Quantitation of reverse transcription products was carried out with a Typhoon Trio (GE Health Sciences). All experiments were carried out a minimum of three independent times to ensure reproducibility of the data.

RESULTS

Recombinant MeCP2 represses in vitro transcription in a CpG-methylation-dependent manner. To investigate the mechanism of transcriptional repression by human MeCP2, we performed in vitro transcription reactions with HeLa nuclear extract in the presence of full-length, wild-type MeCP2. Wild-type MeCP2 and Rett syndrome-associated truncations were expressed as chitin binding domain (CBD) fusion proteins in *E. coli*. The fusion proteins were purified to near

homogeneity using chitin resin, and the CBD was removed by intein mediated cleavage (Figure 3.1). Addition of wild-type MeCP2 to reactions with unmethylated template had little or no effect on transcription (Figure 3.2). However, addition of MeCP2 to reactions with CpG-methylated template resulted in nearly complete repression of transcription. Previous studies have examined repression of transcription by MeCP2 in vitro. However, in our system this repression is specific to CpG-methylated constructs, and methylated constructs are repressed only in the presence of MeCP2

Repression by MeCP2 in vitro does not rely on class I or class II histone deacetylase activity. Past work on repression of transcription by MeCP2 using GAL4-fusion proteins identified histone deacetylase- (HDAC-) dependent mechanism of repression (30, 62). These studies also suggested the existence of an additional, HDAC-independent mechanism of repression. To establish whether the repression seen in our transcription system is dependent on HDAC activity, we added trichostatin A (TSA) to the transcription reaction. TSA is an HDAC inhibitor with broad specificity for class I and class II HDACs (80). In the absence of MeCP2, TSA had no effect on basal transcription from CpG-methylated constructs (Figure 3.3). Addition of TSA also did not relieve MeCP2-mediated repression of transcription. These findings indicate that the mechanism of transcriptional repression utilized by MeCP2 in vitro does not depend on class I or class II HDAC activity. This mechanism may be related to the HDAC-independent component of repression seen in cells (30, 62).

MeCP2 has a second transcription repression domain. MeCP2 has a

transcription repression domain (TRD) that was delineated based on transfection experiments with GAL4-fusion proteins (61). This TRD partially overlaps the co-repressor interacting domain necessary for interaction with HDACs (30, 62), suggesting that the TRD represses transcription by recruiting HDACs. However, published reports and our results (Figure 3.3) indicate that MeCP2 can also repress transcription in an HDAC-independent manner. In order to elucidate whether the HDAC-independent mechanism of transcriptional repression by MeCP2 requires the previously described TRD, we generated truncated MeCP2 proteins that lack this domain (Figure 3.1). We then added these truncated proteins to HeLa transcription reactions with methylated templates (Figure 3.4). The R294X and E235X proteins, which are capable of binding methylated DNA but lack sequences necessary for HDAC-mediated repression in cells (61), were both able to repress transcription in vitro. However, the R168X construct, which also retains the ability to bind methylated DNA (60), was unable to repress transcription. These results clearly demonstrate the presence of a second MeCP2 transcription repression domain between amino acid 167 and amino acid 234. This second transcription repression domain is distinct from, though possibly partially overlapping with, the previously defined TRD.

MeCP2 inhibits pre-initiation complex formation. Given our evidence that MeCP2 can repress transcription in a HDAC-independent manner via a second transcription repression domain, we next attempted to determine the mechanism of this repression. The process of basal transcription can be divided into three main stages: pre-initiation complex (PIC) formation, initiation, and elongation (Figure 3.5).

These stages can be separated experimentally through ordered addition of factors (Figure 3.5A). Inclusion of Sarkosyl limits the transcription reaction to a single round, preventing reinitiation events, which would complicate this analysis.

Addition of MeCP2 to the transcription reaction at later time points (Figure 3.5, time points 3 and 4) resulted in little inhibition of transcription. This indicates that MeCP2 has little effect on the initiation and elongation stages of transcription. In contrast, incubation of MeCP2 with the transcription template before addition of the HeLa nuclear extract (Figure 3.5, time point 1) allowed MeCP2 to influence all stages of the transcription reaction, resulting in near total repression of transcription. Since the data from time points 3 and 4 indicate that MeCP2 does not significantly inhibit initiation or elongation, the repression seen with addition of MeCP2 at time point 1 must be primarily due to inhibition of PIC formation. Interestingly, adding MeCP2 to the reaction at the same time as the HeLa nuclear extract (Figure 3.5, time point 2) resulted in an intermediate level of repression. This suggests that pre-binding of MeCP2 to DNA can prevent PIC formation but MeCP2 cannot disrupt PICs that are already assembled.

Repression by MeCP2 does not require binding in the core promoter. Our finding that MeCP2 represses transcription by inhibiting PIC formation suggested that this repression may be due to steric inhibition of general transcription factor binding to the core promoter. Assembly of the PIC requires that TFIID, RNA polymerase II and other basal transcription factors bind directly to the region of DNA 40 bp upstream and downstream of the transcription start site (TSS) (23). Binding of MeCP2 to

methyl-CpG dinucleotides in the core promoter could physically block interaction of the basal transcription machinery with promoter DNA, thus inhibiting transcription.

To test this model, we developed a core promoter that has no CpG dinucleotides within 50 bp of the TSS. While the region of DNA occupied by the binding of full-length MeCP2 is not known, binding of the isolated MBD of MeCP2 results in a DNaseI footprint that only extends six to eight base pairs in either direction from the methyl-CpG dinucleotide (60). Based on these findings, we expect that a 100 bp CpG-free window should allow simultaneous binding of MeCP2 and the basal transcription factors. If steric interference is the sole mechanism of repression, transcription from this construct should not be repressed by MeCP2. However, MeCP2 was able to repress transcription from this construct by 15-fold (Figure 3.6), despite the lack of methyl-CpG dinucleotides near the TSS. This result indicates that transcriptional repression by MeCP2 does not depend on binding to the core promoter.

MeCP2 cannot repress transcription in the absence of CpG dinucleotides. Our analysis of MeCP2 repression on a construct with a CpG-free core promoter led us to ask how the pattern of CpG methylation might affect MeCP2 function. In mammalian genomes, 60% to 90% of CpG dinucleotides are methylated (9). MeCP2 is capable of binding methyl-CpG moieties in almost any sequence context (39, 52), yet MeCP2 is able to selectively repress a subset of genes and may even activate transcription from others (15, 84). These findings highlight the question of how patterns of CpG-methylation direct MeCP2 activity.

In order to evaluate this question, we used a CpG-free plasmid, which allowed

us to systematically place CpG dinucleotides at specific positions in the transcription construct. Using this system, we generated constructs with either zero or six CpG dinucleotides in the core promoter and no CpG dinucleotides in the plasmid. We then treated these constructs with M.SssI methyltransferase and used them in HeLa transcription reactions in the presence or absence of MeCP2 (Figure 3.8). Addition of MeCP2 to transcription reactions with the template with six methyl-CpG dinucleotides (6mCpG) resulted in approximately 6-fold reduction in transcription. This indicates that MeCP2 can repress transcription in the context of the CpG-free plasmid if CpG dinucleotides are present in the core promoter. Importantly, MeCP2 was unable to repress transcription from the construct with no CpG dinucleotides in the plasmid or the core promoter (0mCpG), even after treatment with methyltransferase. This is an important control, as it demonstrates that the repressive effect of MeCP2 is specific to CpG-methylation and is not due to some artifact of the methylation reaction.

Strong repression of transcription by MeCP2 requires more than one methyl-CpG dinucleotide. Biochemical studies of the relationship between the density of CpG-methylation transcriptional repression indicate that low levels of methylation can cause significant reductions in transcription, but full methylation is required for maximal repression (28, 61). In addition, a recent genome-wide analysis found that MeCP2 binding correlates with CpG methylation (74). However, a previous genome-wide study found little correlation between MeCP2 binding, density of DNA methylation, and transcriptional repression (84). Another study reported that MeCP2 may activate transcription from genes with low levels of promoter methylation (15). In

addition, analysis of the putative MeCP2 target gene BDNF has raised the possibility that a single methyl-CpG dinucleotide might be sufficient for repression by MeCP2 (16).

To test whether the number of methylation sites affects repression by MeCP2, we methylated constructs with one, two or three CpG dinucleotides in the core promoter (Figure 3.9). In this analysis, constructs with a single methyl-CpG dinucleotide (1mCpG) more than 40 bp away from the TSS were not repressed by MeCP2. While 1mCpG constructs with the methyl-CpG dinucleotide closer to the TSS were partially repressed by MeCP2, none of the 1mCpG templates were repressed to the same extent as the 6mCpG templates (Figure 3.8). Transcription templates with two or three methyl-CpG dinucleotides in the core promoter show progressively greater susceptibility to repression by MeCP2 (data not shown). These findings indicate that repression by MeCP2 is not all-or-nothing, but instead show that more methylation results in more repression.

DISCUSSION

In this study, we have explored the mechanism of transcriptional repression by MeCP2 and the role that methyl-CpG dinucleotide number and position plays in directing MeCP2 activity. This work has demonstrated that MeCP2 can repress transcription in a HDAC-independent manner by inhibiting PIC formation. We have also shown that this repression is mediated by a second, previously unidentified transcription repression domain. Finally, we have begun to elucidate how the number

and location of methyl-CpG dinucleotides can affect repression by MeCP2. These findings have contributed to a greater understanding of the function of MeCP2.

Mechanisms of transcriptional repression by MeCP2. In our analysis of in vitro transcriptional repression by wild-type MeCP2, we showed that this repression is dependent on CpG methylation (Figure 3.2) but does not require the activity of TSA-sensitive HDACs (Figure 3.3). These experiments confirmed the existence of the HDAC-independent mechanism that had been hypothesized (62). They also validated our system as a means to study HDAC-independent repression, which had been obscured by the HDAC-dependent mechanism in past studies.

In order to further understand HDAC-independent repression of transcription by MeCP2, we next asked whether the region of MeCP2 that is responsible for this repression is the same as the region that mediates HDAC-dependent repression. Using C-terminal truncations, we found that MeCP2 mutants that are unable to sustain HDAC-mediated repression of transcription in cells are able to fully repress transcription in our system (Figure 3.4). These findings confirm and extend previous work done with MeCP2 in cells (42), which also showed repression of transcription by MeCP2 truncations that lacked the previously defined TRD. However, complications due to overlap of the TRD and the MBD with two nuclear localization signals caused the author to concluded that this repression was the result of binding of MeCP2 to the template DNA, in the absence of a specific transcription repression domain. Our new results clearly demonstrate the existence of a second transcription repression domain, which provides an alternative interpretation of the previous observations.

Patterns of CpG methylation affect MeCP2 function. We observe that both the number and position of methyl-CpG dinucleotides influences transcriptional repression by MeCP2. Previous studies have found that strong repression of transcription by MeCP2 requires CpG methylation that exceeds some threshold (28, 61). However, these studies were performed using various methyltransferases with dozens or hundreds of recognition sites throughout the plasmids and the promoter regions. As a result, the effect of average methyl-CpG density could not be separated from the effect of specific methyl-CpG positions. We have solved this problem by using a CpG-free plasmid as the starting point for our analysis. Using these templates, we have shown that both the number and position of methyl-CpG dinucleotides affect repression by MeCP2 (Figure 3.7 and 2.8). In general, we find that more methylation results in more repression, and methylation closer to the TSS induces more repression than methylation farther from the TSS.

The results presented here confirm the existence of a second mechanism of transcriptional repression by MeCP2 and enhance our understanding of the methylation patterns required for MeCP2 activity. It is possible that the HDAC-independent mechanism of repression demonstrated here may be dysregulated in certain cases of Rett syndrome. Given the separate but overlapping nature of the two transcription repression domains, the specific mutation that a patient has may dictate whether proposed therapies that modulate the HDAC-dependent activity should be used alone or in conjunction with potential therapies that modulate the HDAC-independent activity. Similarly, our new understanding of how methylation patterns

influence MeCP2 activity will aid in validation of putative MeCP2 target genes. Genome-wide studies have identified potential MeCP2 targets, but these studies often cannot separate direct effects from indirect effects. With an understanding of the required methylation patterns necessary for MeCP2 function, researchers will be able to exclude many misidentified MeCP2 targets. Together, these findings will be important for advancing our understanding of normal MeCP2 function and its dysfunction in Rett syndrome and for developing therapies for this disease.

Table 3.1: Oligonucleotides used to create multiple cloning sites in the pCpG-Txn plasmid.

Oligo Name	Sequence
MCSp_For	TACATGTACTTGTAGATCTTCTAGATGGGAACTGCAGGGATCCATGACATCATGAA
MCSp_Rev	CTAGTTCATGATGTCATGGATCCCTGCAGTTCCCATCTAGAAGATCTACAAGTACATGTATGCA
MCSd_For	TAAGCTTTGACTATGTACAACTAGTTGACTAATGCATGGGCCCTGACTACTTAAGACATG
MCSd _Rev	TCTTAAGTAGTCAGGGCCCATGCATTAGTCAACTAGTTGTACATAGTCAAAGCTTACATG

Table 3.2: Oligonucleotides used to prepare core promoter constructs. Core promoter sequences are represented by uppercase lettering, whereas the sequences used for subcloning of the oligonucleotides into the *Xba* I and *Pst* I restriction sites of pUC119 or pCpG-Txn are denoted by lowercase lettering. The A+1 position in the Inr is highlighted in teal. CpG dinucleotides are highlighted in red. Mutations to remove CpG dinucleotides are highlighted in green.

Construct	Sequence
SCPX	ctagaGCTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGTCGAGCCGAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-0CpG	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCTCAGATTGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG-49	ctagaGCGTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCTCAGATTGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG-14	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGCGTAGTGAACCTCAGATTGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG-4	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATTGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG+5	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCTCAGATCGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG+10	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCTCAGATTGCCTCGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG+23	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCTCAGATTGCCTGGAGACCTCAAGCCGAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG+49	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCTCAGATTGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACCGctgca
SCPX-CpG-14-4	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGCGTAGTGAACCGTCAGATTGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG-14+5	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGCGTAGTGAACCCTCAGATCGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG-4+5	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca
SCPX-CpG-14-4+5	ctagaGCTTGTACTATGGGAGGTCTATATAAGCAGAGCTGGCGTAGTGAACCGTCAGATCGCCTGGAGACCTCAAGCCAAGTGGTTGTGCCTCCATAGAAGACACctgca

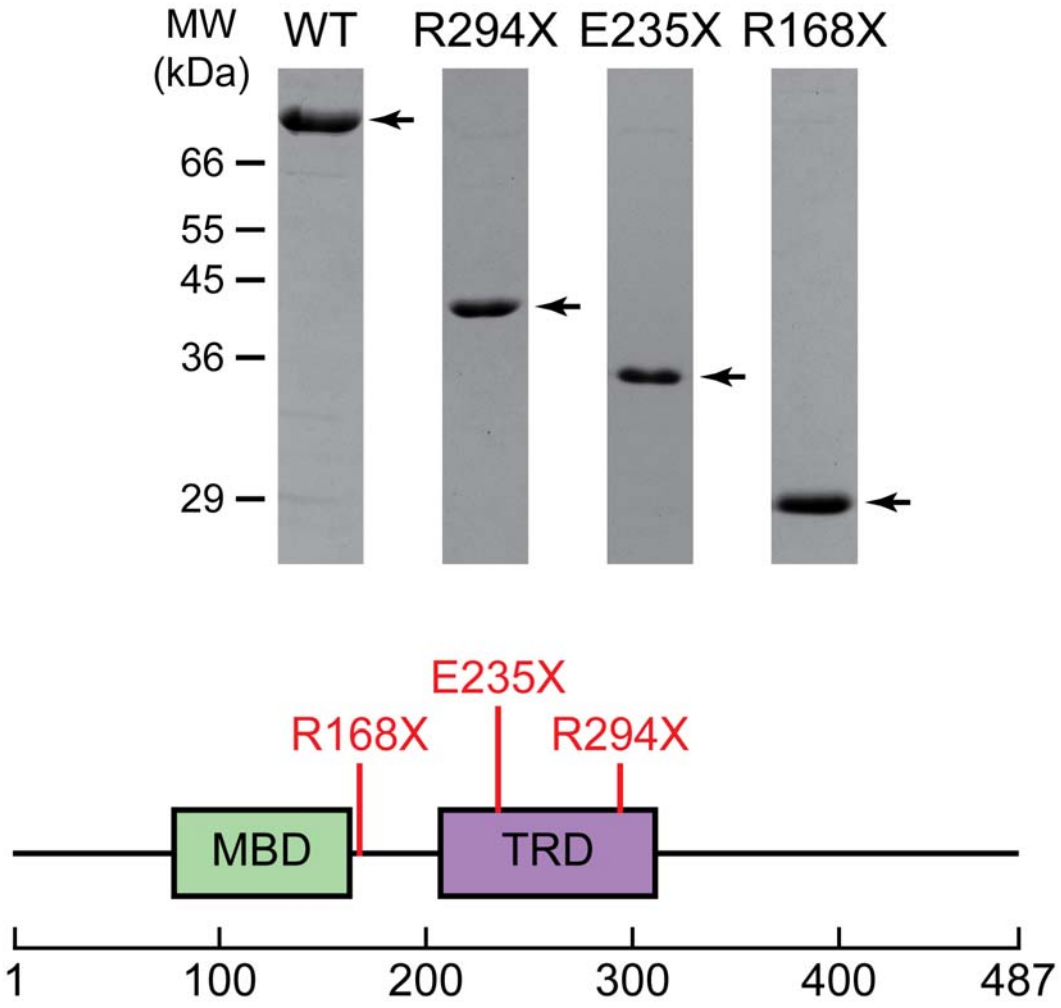


Figure 3.1: Expression and purification of full-length, wild-type MeCP2 and Rett syndrome-associated truncations. Coomassie blue stained SDS-PAGE of recombinant MeCP2-chitin binding domain (CBD) fusion proteins purified using a chitin resin. The CBD was removed from full-length or trucated MeCP2 proteins by intein-mediated clevage. Below is a schematic of the domain structure of MeCP2, with an N-terminal methyl-CpG binding domain (MBD), a centrally located transcription repression domain (TRD) and an uncharacterized C-terminal domain. The relative positions of the MeCP2 truncations are labeled in red.

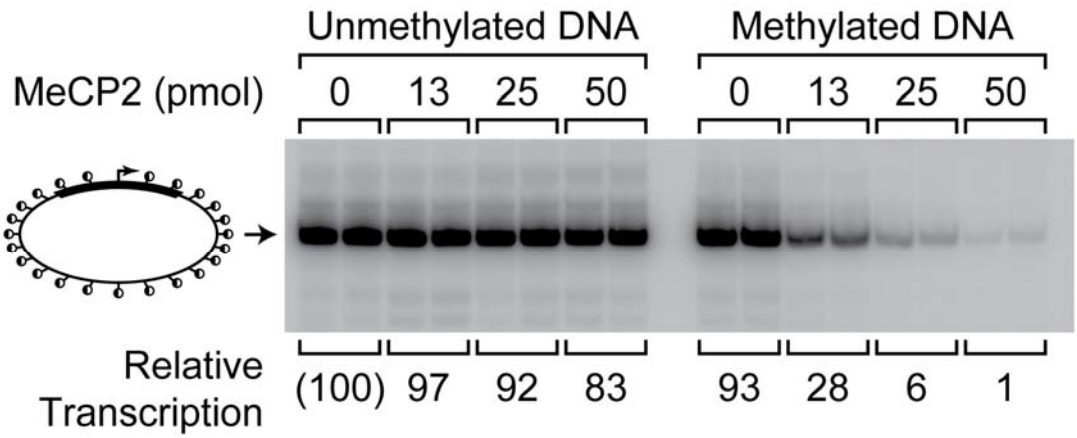


Figure 3.2: Repression of transcription by MeCP2 requires CpG methylation. Addition of MeCP2 to the in vitro HeLa transcription reaction results in dose-dependent repression of transcription from pUC-SCP that is fully methylated with M.SssI methyltransferase but not from unmethylated pUC-SCPX. pUC-SCPX has CpG dinucleotides in the pUC vector and the SCPX core promoter.

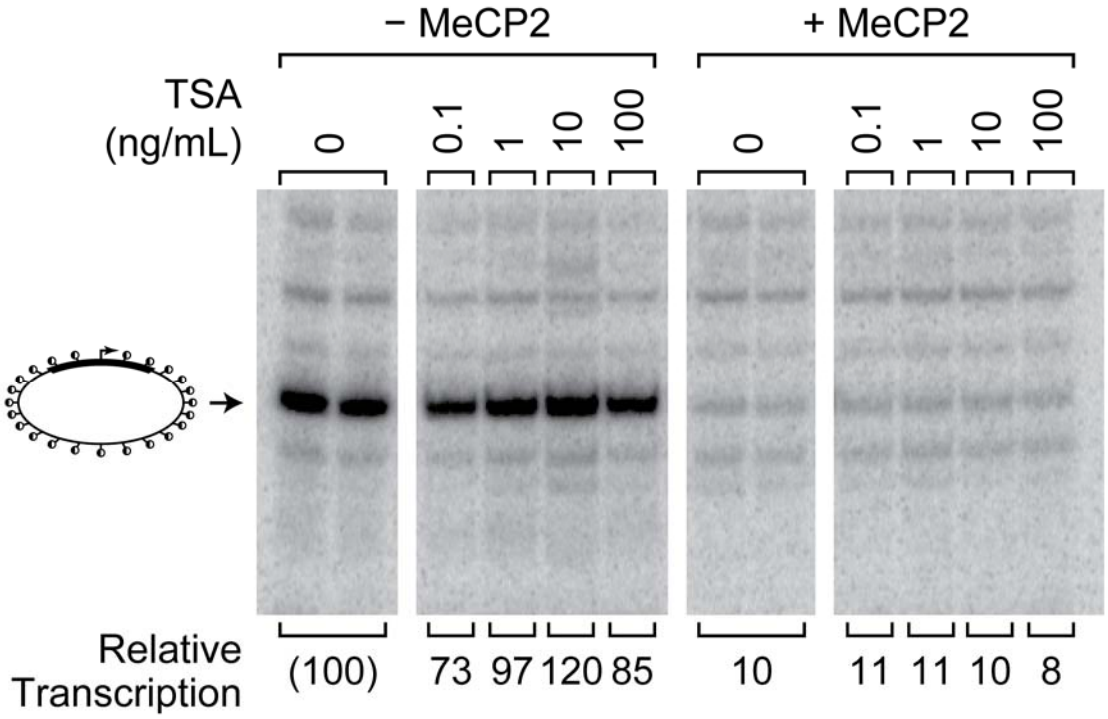
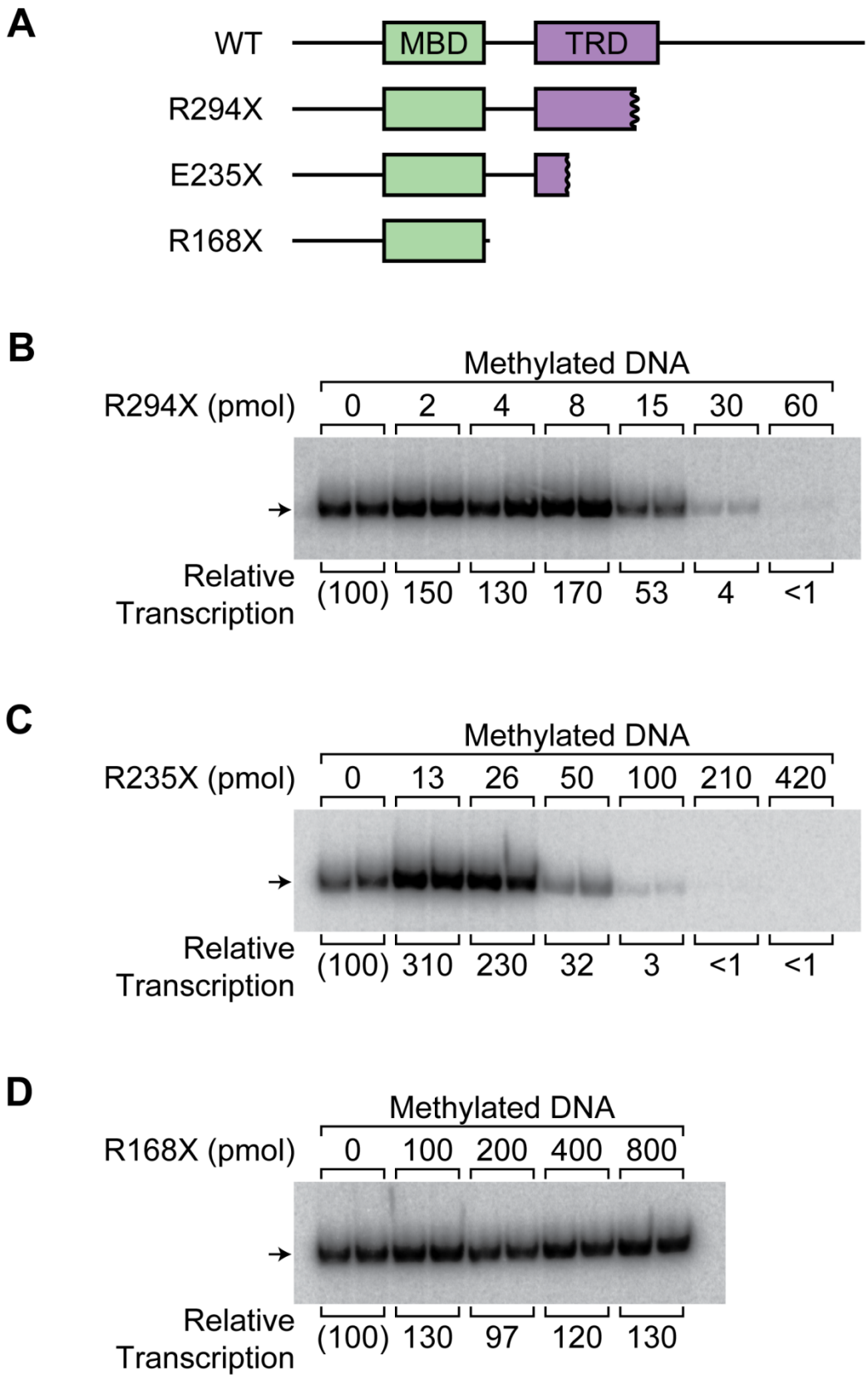


Figure 3.3: Repression of in vitro transcription by MeCP2 does not appear to depend on histone deacetylase activity. Fully-methylated template was transcribed with or without MeCP2 in the presence of increasing amounts of the histone deacetylase inhibitor trichostatin A (TSA).

Figure 3.4: Repression of transcription by MeCP2 in vitro does not require a previously identified transcription repression domain. (A) Schematic of the domain structure of full-length MeCP2 (WT) and the truncated proteins used in this analysis. (B and C) The R294X and E235X truncations are able to repress transcription to the same extent as and at similar concentrations as the full-length protein (see Figure 3.2). (D) The R168X truncation is unable to repress transcription even at concentrations far in excess of those needed for repression by wild-type MeCP2 or the other truncations.


A
WT
MBD
TRD
R294X
E235X
R168X
B
Methylated DNA
R294X (pmol)
0
2
4
8
15
30
60
Relative Transcription
(100)
150
130
170
53
4
<1
C
Methylated DNA
R235X (pmol)
0
13
26
50
100
210
420
Relative Transcription
(100)
310
230
32
3
<1
<1
D
Methylated DNA
R168X (pmol)
0
100
200
400
800
Relative Transcription
(100)
130
97
120
130

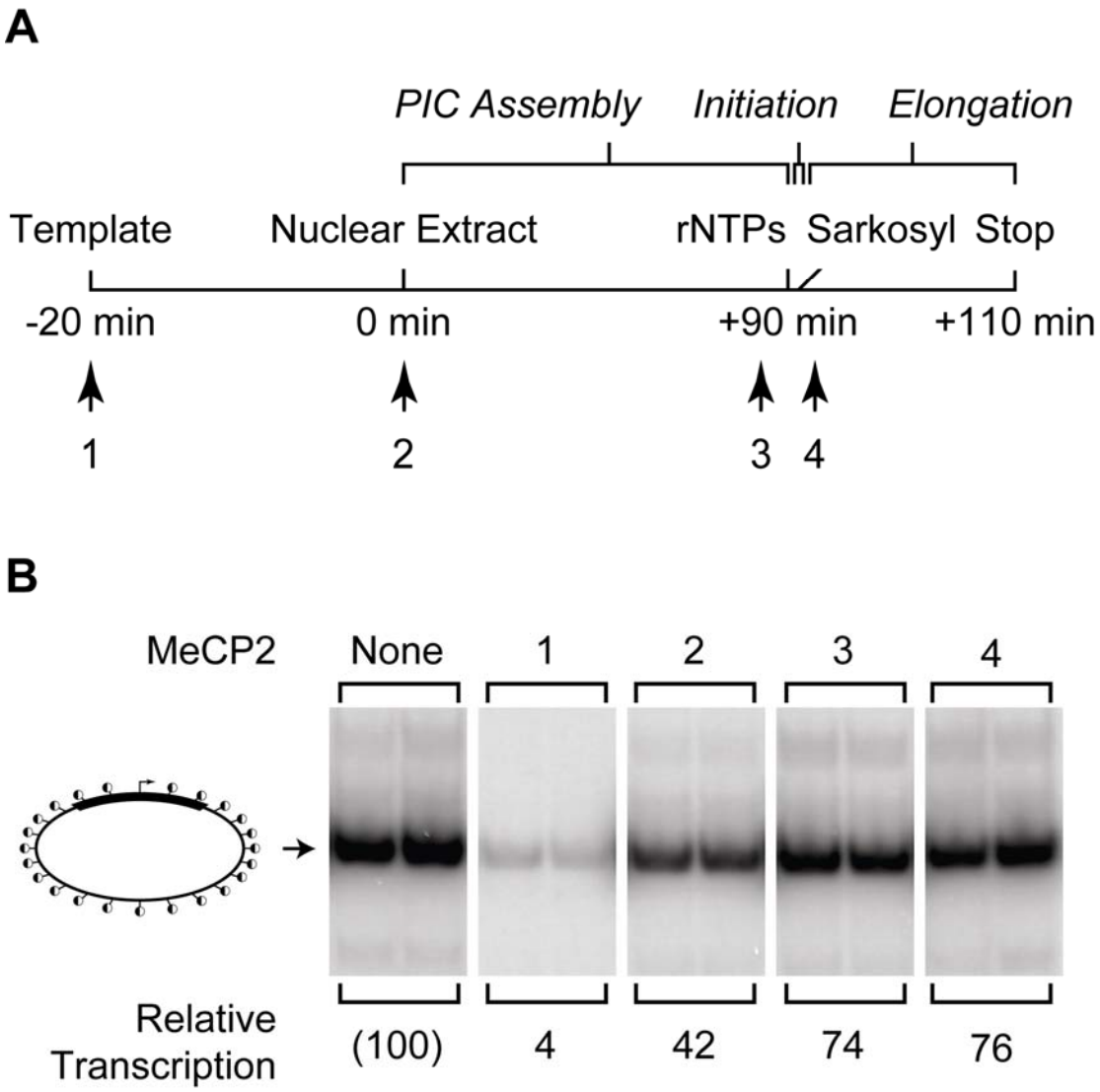


Figure 3.5: MeCP2 inhibits pre-initiation complex formation. (A) Schematic of the stages of the transcription reaction and how they relate to the steps in the in vitro HeLa transcription protocol. (B) Addition of MeCP2 at various time points during transcription of fully-methylated pUC-SCPX. Complete repression of transcription is seen only when MeCP2 is added before PIC formation has started.

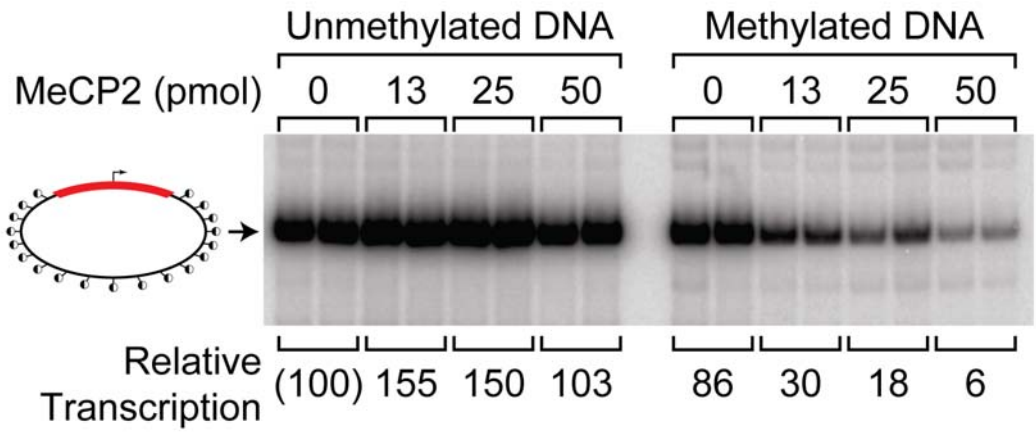


Figure 3.6: Repression of transcription by MeCP2 does not require methylation in the core promoter. Addition of MeCP2 to the transcription reaction represses transcription from pUC-SCPX-0CpG that is fully methylated with M.SssI methyltransferase but not from unmethylated pUC-SCPX-0CpG. pUC-SCPX-0CpG has CpG dinucleotides in the pUC vector but lacks CpG dinucleotides in the core promoter, from −50 to +50 relative to the transcription start site.

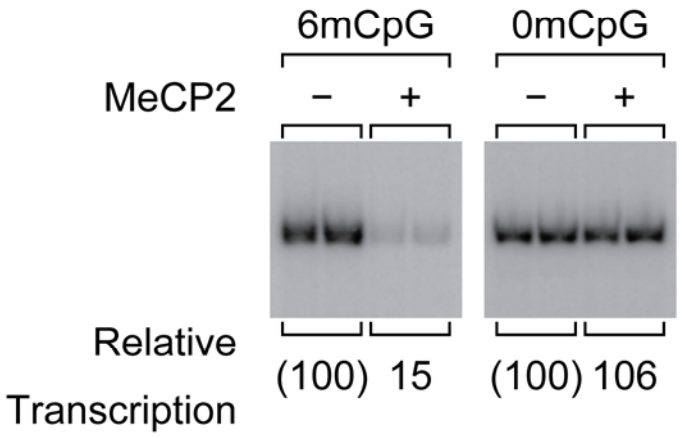


Figure 3.7: MeCP2 does not repress transcription from a methyl-transferase-treated, CpG-free construct. Constructs with no CpG dinucleotides in the vector and zero or six CpG dinucleotides in the core promoter were methylated with M.SssI methyltransferase and subjected to in vitro transcription with HeLa nuclear extract in the presence or absence of MeCP2. MeCP2 represses transcription from the construct with only six methylated CpG dinucleotides in the core promoter (6mCpG) but is unable to repress transcription from a CpG-free construct (0mCpG).

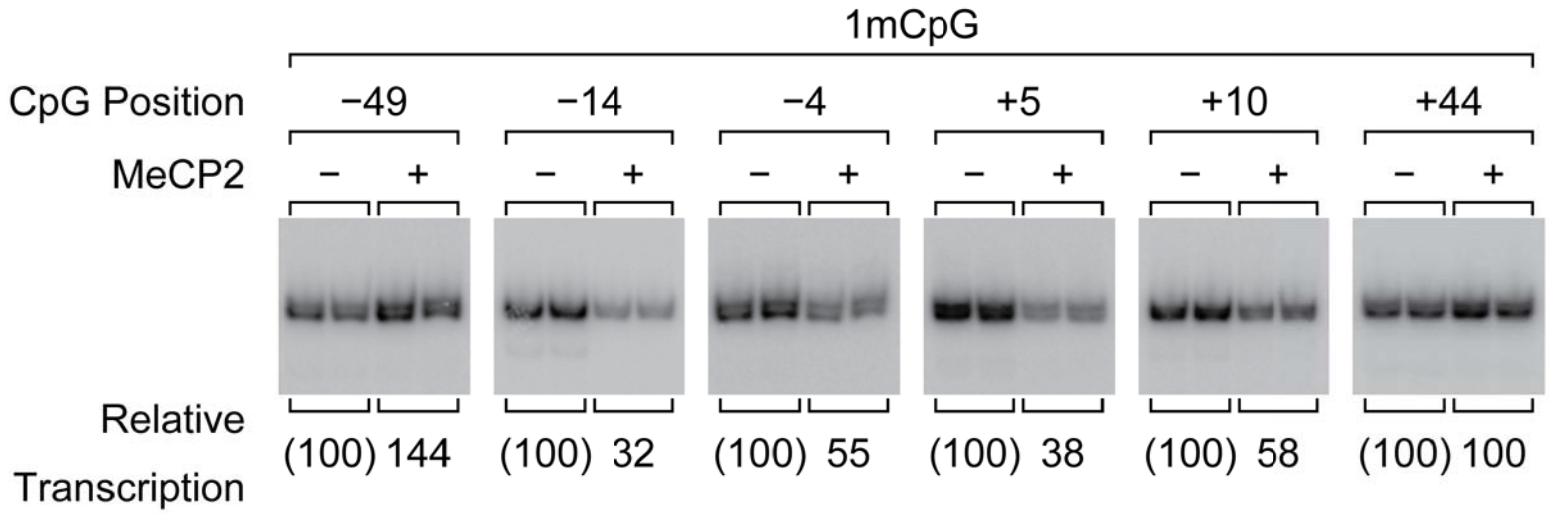


Figure 3.8: A single MeCP2 binding site is not sufficient for strong repression of transcription. Constructs with a single CpG dinucleotide in the core promoter were methylated with M.SssI methyltransferase and transcribed in the presence or absence of MeCP2. None of these 1mCpG constructs are repressed to the same extent as the 6mCpG construct (see Figure 3.8). Constructs with two or three methyl-CpG dinucleotides show an additive effect on transcriptional repression (data not shown).

Chapter 4

Conclusions

Regulation of transcription in eukaryotic cells is a highly complex task. The activity of tens of thousands of genes must be regulated simultaneously, but independently, in response to normal physiology or environmental stimuli. Many layers of signaling and control mechanisms achieve this daunting goal. This dissertation has explored two of these mechanisms.

My work on the nature of the downstream elements of the core promoter has improved our understanding of what the MTE is and how it functions with the DPE. I have shown that the downstream region of the core promoter contains three subregions that each contribute to binding of TFIID to the core promoter and to transcription. I have shown that the MTE can be defined as encompassing two of these subregions (18-22 and 27-29). The DPE also encompasses two of these subregions (27-29 and 30-33). In addition, I have identified a new class of core promoter element, the Bridge element, which encompasses part of the MTE sequence and part of the DPE sequence (18-22 and 30-33, respectively) without having the full consensus of either element. Given the role that combinations of core promoter sequences play in determining enhancer-promoter specificity, this improved understanding of core promoter architecture will be important for understanding the detailed regulatory mechanisms used by many important genes.

In addition to the clear-cut conclusions described above, my work on core promoter elements has illuminated a difficult, but important concept. Using our understanding of the preferred sequences for each of the downstream subregion, I was able to identify promoters that had optimal or suboptimal sequences as these positions.

When I mutated suboptimal regions to the best possible sequence, transcription from these promoters increased. Similarly, transcription from these promoters decreased when I mutated optimal regions to the worst possible sequence. However, when I mutated suboptimal regions to the worst possible sequence, transcription was still reduced. Despite the lack of consensus sequence in these regions, the natural sequence was still able to support transcription better than the worst possible sequence. This suggests that core promoter elements should not be judged in terms of presence or absence but rather in terms of degrees of optimality.

Another complication revealed by my work is the concept of context dependence. It has been shown for transcription factor binding sites that the nucleotide present at one position of the binding site can influence the preferred nucleotide at a different position. As a hypothetical example, in a binding site with the general consensus AWST, it might be that having and A at the second position is preferred when there is a G at the third position, but that a T at the second position is preferred when there is a C at the third position. In other words, the preferred nucleotide at the second position is dependent on its sequence context, i.e. the nucleotide at the third position. I have shown that this concept of context dependence also applies to the preferred nucleotides in core promoter elements. How this context dependence affects the function of core promoters in the more complex regulatory environment of a living cell remains to be seen.

My work on the mechanism of repression of transcription by MeCP2 is less surprising in that it confirms existing hypotheses, but it may have more immediate

application. I have shown that MeCP2 has a second transcription repression domain that functions through a completely different mechanism than the previously defined repression domain, a finding that may prove important for Rett syndrome therapy. Since the two repression activities of MeCP2 are contained in different portions of the protein, patients with less severe truncations or different point mutations might retain different relative amounts of each activity. This knowledge will be important when developing and testing therapies and when deciding which therapies are appropriate for which patients.

The biggest mystery facing research on the ontology of Rett syndrome is the need for improved understanding of how MeCP2 selects target genes. My investigations of the role that the position and number of methyl-CpG dinucleotides plays in repression in vitro are the first step towards this understanding. However, I think that the real contribution of this works will come through expanded use of the pGpG-Txn plasmid to study theses parameters in cell lines and, eventually, primary neurons.

References

1. **Ahsan, B., T. L. Saito, S. Hashimoto, K. Muramatsu, M. Tsuda, A. Sasaki, K. Matsushima, T. Aigaki, and S. Morishita.** 2009. MachiBase: a Drosophila melanogaster 5'-end mRNA transcription database. Nucleic Acids Res **37:**D49-53.

2. **Alland, L., R. Muhle, H. Hou, Jr., J. Potes, L. Chin, N. Schreiber-Agus, and R. A. DePinho.** 1997. Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. Nature **387:**49-55.

3. **Amir, R. E., I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, and H. Y. Zoghbi.** 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet **23:**185-8.

4. **Ballas, N., C. Grunseich, D. D. Lu, J. C. Speh, and G. Mandel.** 2005. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell **121:**645-657.

5. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA-sequences. Cell **27:**299-308.

6. **Bartholomew, B., G. A. Kassavetis, B. R. Braun, and E. P. Geiduschek.** 1990. The subunit structure of Saccharomyces cerevisiae transcription factor IIIC probed with a novel photocrosslinking reagent. EMBO J **9:**2197-205.

7. **Bartholomew, B., G. A. Kassavetis, and E. P. Geiduschek.** 1991. Two components of Saccharomyces cerevisiae transcription factor IIIB (TFIIIB) are stereospecifically located upstream of a tRNA gene and interact with the second-largest subunit of TFIIIC. Mol Cell Biol **11:**5181-9.

8. **Bhaumik, S. R., E. Smith, and A. Shilatifard.** 2007. Covalent modifications of histones during development and disease pathogenesis. Nature Structural & Molecular Biology **14:**1008-1016.

9. **Bird, A. P.** 1986. CpG-rich islands and the function of DNA methylation. Nature **321:**209-13.

10. **Bird, A. P., and A. P. Wolffe.** 1999. Methylation-induced repression - Belts, braces, and chromatin. Cell **99:**451-454.

11. **Bourdon, V., C. Philippe, O. Labrune, D. Amsallem, C. Arnould, and P. Jonveaux.** 2001. A detailed analysis of the MECP2 gene: prevalence of recurrent mutations and gross DNA rearrangements in Rett syndrome patients. Hum Genet **108:**43-50.

12. **Burke, T. W., and J. T. Kadonaga.** 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev **11:**3020-31.

13. **Burke, T. W., and J. T. Kadonaga.** 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev **10:**711-24.

14. **Burley, S. K., and R. G. Roeder.** 1996. Biochemistry and structural biology of transcription factor IID (TFIID). Annu Rev Biochem **65:**769-99.

15. **Chahrour, M., S. Y. Jung, C. Shaw, X. Zhou, S. T. Wong, J. Qin, and H. Y. Zoghbi.** 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science **320:**1224-9.

16. **Chen, W. G., Q. Chang, Y. X. Lin, A. Meissner, A. E. West, E. C. Griffith, R. Jaenisch, and M. E. Greenberg.** 2003. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science **302:**885-889.

17. **D'Alessio, J. A., K. J. Wright, and R. Tjian.** 2009. Shifting players and paradigms in cell-specific transcription. Mol Cell **36:**924-31.

18. **Deng, W., and S. G. Roberts.** 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev **19:**2418-23.

19. **Deng, W., and S. G. Roberts.** 2007. TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma **116:**417-29.

20. **Dynlacht, B. D., T. Hoey, and R. Tjian.** 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell **66:**563-76.

21. **Ernst, P., and S. T. Smale.** 1995. Combinatorial regulation of transcription .1. General-aspects of transcriptional control. Immunity **2:**311-319.

22. **Goldberg, M. L.** 1979. Ph.D. Thesis Title. Stanford University, Palo Alto, CA.

23. **Gralla, J. D.** 1996. Global steps during initiation by RNA polymerase II. Methods Enzymol **273:**99-110.

24. **Hagberg, B., J. Aicardi, K. Dias, and O. Ramos.** 1983. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. Ann Neurol **14:**471-9.

25. **Harikrishnan, K. N., M. Z. Chow, E. K. Baker, S. Pal, S. Bassal, D. Brasacchio, L. Wang, J. M. Craig, P. L. Jones, S. Sif, and A. El-Osta.** 2005. Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nature Genetics **37:**254-264.

26. **Heinzel, T., R. M. Lavinsky, T. M. Mullen, M. Soderstrom, C. D. Laherty, J. Torchia, W. M. Yang, G. Brard, S. D. Ngo, J. R. Davie, E. Seto, R. N. Eisenman, D. W. Rose, C. K. Glass, and M. G. Rosenfeld.** 1997. A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. Nature **387:**43-8.

27. **Hoey, T., R. O. Weinzierl, G. Gill, J. L. Chen, B. D. Dynlacht, and R. Tjian.** 1993. Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. Cell **72:**247-60.

28. **Hsieh, C. L.** 1994. Dependence of transcriptional repression on CpG methylation density. Mol Cell Biol **14:**5487-94.

29. **Jenuwein, T., and C. D. Allis.** 2001. Translating the histone code. Science **293:**1074-80.

30. **Jones, P. L., G. J. C. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis, and A. P. Wolffe.** 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genetics **19:**187-191.

31. **Juven-Gershon, T., S. Cheng, and J. T. Kadonaga.** 2006. Rational design of a super core promoter that enhances gene expression. Nat Methods **3:**917-22.

32. **Juven-Gershon, T., J. Y. Hsu, and J. T. Kadonaga.** 2008. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. Genes Dev **22:**2823-30.

33. **Juven-Gershon, T., and J. T. Kadonaga.** 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol **339:**225-9.

34. **Kadonaga, J. T.** 1990. Assembly and disassembly of the Drosophila RNA polymerase II complex during transcription. J Biol Chem **265:**2624-31.

35. **Kadonaga, J. T.** 1991. Purification of sequence-specific binding proteins by DNA affinity chromatography. Methods Enzymol **208:**10-23.

36. **Kimura, H., and K. Shiota.** 2003. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. Journal of Biological Chemistry **278:**4806-4812.

37. **Klose, R., and A. Bird.** 2003. Molecular biology. MeCP2 repression goes nonglobal. Science **302:**793-5.

38. **Klose, R. J., and A. P. Bird.** 2004. MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J Biol Chem **279:**46490-6.

39. **Klose, R. J., S. A. Sarraf, L. Schmiedeberg, S. M. McDermott, I. Stancheva, and A. P. Bird.** 2005. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell **19:**667-78.

40. **Kokubo, T., D. W. Gong, J. C. Wootton, M. Horikoshi, R. G. Roeder, and Y. Nakatani.** 1994. Molecular cloning of Drosophila TFIID subunits. Nature **367:**484-7.

41. **Kokura, K., S. C. Kaul, R. Wadhwa, T. Nomura, M. M. Khan, T. Shinagawa, T. Yasukawa, C. Colmenares, and S. Ishii.** 2001. The ski protein family is required for MeCP2-mediated transcriptional repression. Journal of Biological Chemistry **276:**34115-34121.

42. **Kudo, S.** 1998. Methyl-CpG-binding protein MeCP2 represses Sp1-activated transcription of the human leukosialin gene when the promoter is methylated. Mol Cell Biol **18:**5492-9.

43. **Kutach, A. K., and J. T. Kadonaga.** 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol **20:**4754-64.

44. **Laemmli, U. K.** 1978. Levels of organization of the DNA in eucaryotic chromosomes. Pharmacol Rev **30:**469-76.

45. **Lagrange, T., A. N. Kapanidis, H. Tang, D. Reinberg, and R. H. Ebright.** 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev **12:**34-44.

46. **Laherty, C. D., W. M. Yang, J. M. Sun, J. R. Davie, E. Seto, and R. N. Eisenman.** 1997. Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell **89:**349-56.

47. **Lee, D. H., N. Gershenzon, M. Gupta, I. P. Ioshikhes, D. Reinberg, and B. A. Lewis.** 2005. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol **25:**9674-86.

48. **Lee, J. W., Y. C. Lee, S. Y. Na, D. J. Jung, and S. K. Lee.** 2001. Transcriptional coregulators of the nuclear receptor superfamily: coactivators and corepressors. Cellular and Molecular Life Sciences **58:**289-297.

49. **Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li.** 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development **122:**3195-3205.

50. **Lemon, B., and R. Tjian.** 2000. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev **14:**2551-69.

51. **Lewis, B. A., T. K. Kim, and S. H. Orkin.** 2000. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. Proc Natl Acad Sci U S A **97:**7172-7.

52. **Lewis, J. D., R. R. Meehan, W. J. Henzel, I. Maurerfogy, P. Jeppesen, F. Klein, and A. Bird.** 1992. Purification, sequence, and cellular-localization of a novel chromosomal protein that binds to methylated DNA. Cell **69:**905-914.

53. **Li, B., M. Carey, and J. L. Workman.** 2007. The role of chromatin during transcription. Cell **128:**707-19.

54. **Li, E., T. H. Bestor, and R. Jaenisch.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:**915-926.

55. **Lim, C. Y., B. Santoso, T. Boulay, E. Dong, U. Ohler, and J. T. Kadonaga.** 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev **18:**1606-17.

56. **Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond.** 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389:**251-60.

57. **Lunyak, V. V., R. Burgess, G. G. Prefontaine, C. Nelson, S. H. Sze, J. Chenoweth, P. Schwartz, P. A. Pevzner, C. Glass, G. Mandel, and M. G. Rosenfeld.** 2002. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science **298:**1747-1752.

58. **Meehan, R., J. Lewis, S. Cross, X. Nan, P. Jeppesen, and A. Bird.** 1992. Transcriptional repression by methylation of CpG. J Cell Sci Suppl **16:**9-14.

59. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371-8.

60. **Nan, X., R. R. Meehan, and A. Bird.** 1993. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res **21:**4886-92.

61. **Nan, X. S., F. J. Campoy, and A. Bird.** 1997. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell **88:**471-481.

62. **Nan, X. S., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird.** 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393:**386-389.

63. **Ohler, U., G. C. Liao, H. Niemann, and G. M. Rubin.** 2002. Computational analysis of core promoters in the Drosophila genome. Genome Biol **3:**RESEARCH0087.

64. **Okano, M., D. W. Bell, D. A. Haber, and E. Li.** 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99:**247-257.

65. **Paule, M. R., and R. J. White.** 2000. Transcription by RNA polymerases I and III. Nucleic Acids Research **28:**1283-1298.

66. **Reese, J. C.** 2003. Basal transcription factors. Curr Opin Genet Dev **13:**114-8.

67. **Rice, J. C., and C. D. Allis.** 2001. Histone methylation versus histone acetylation: new insights into epigenetic regulation. Current Opinion in Cell Biology **13:**263-273.

68. **Roeder, R. G.** 1996. Nuclear RNA polymerases: role of general initiation factors and cofactors in eukaryotic transcription. Methods Enzymol **273:**165-71.

69. **Roth, S. Y., J. M. Denu, and C. D. Allis.** 2001. Histone acetyltransferases. Annual Review of Biochemistry **70:**81-120.

70. **Shahbazian, M. D., and M. Grunstein.** 2007. Functions of site-specific histone acerylation and deacetylation. Annual Review of Biochemistry **76:**75-100.

71. **Shao, H., M. Revach, S. Moshonov, Y. Tzuman, K. Gazit, S. Albeck, T. Unger, and R. Dikstein.** 2005. Core promoter binding by histone-like TAF complexes. Mol Cell Biol **25:**206-19.

72. **Shiio, Y., and R. N. Eisenman.** 2003. Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A **100:**13225-30.

73. **Sims, R. J., 3rd, S. S. Mandal, and D. Reinberg.** 2004. Recent highlights of RNA-polymerase-II-mediated transcription. Curr Opin Cell Biol **16:**263-71.

74. **Skene, P. J., R. S. Illingworth, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, R. Andrews, and A. P. Bird.** 2010. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell **37:**457-68.

75. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57:**103-13.

76. **Smale, S. T., and J. T. Kadonaga.** 2003. The RNA polymerase II core promoter. Annu Rev Biochem **72:**449-79.

77. **Soeller, W. C., S. J. Poole, and T. Kornberg.** 1988. In vitro transcription of the Drosophila engrailed gene. Genes Dev **2:**68-81.

78. **Thomas, M. C., and C. M. Chiang.** 2006. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol **41:**105-78.

79. **Tokusumi, Y., Y. Ma, X. Song, R. H. Jacobson, and S. Takada.** 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:**1844-58.

80. **Vanhaecke, T., P. Papeleu, G. Elaut, and V. Rogiers.** 2004. Trichostatin A-like hydroxamate histone deacetylase inhibitors as therapeutic agents: toxicological point of view. Curr Med Chem **11:**1629-43.

81. **Wampler, S. L., C. M. Tyree, and J. T. Kadonaga.** 1990. Fractionation of the general RNA polymerase II transcription factors from Drosophila embryos. J Biol Chem **265:**21223-31.

82. **Weake, V. M., and J. L. Workman.** 2008. Histone ubiquitination: Triggering gene activity. Molecular Cell **29:**653-663.

83. **Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, and L. A. Romano.** 2003. The evolution of transcriptional regulation in eukaryotes. Molecular Biology and Evolution **20:**1377-1419.

84. **Yasui, D. H., S. Peddada, M. C. Bieda, R. O. Vallero, A. Hogart, R. P. Nagarajan, K. N. Thatcher, P. J. Farnham, and J. M. Lasalle.** 2007. Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A **104:**19416-21.

85. **Young, J. I., E. P. Hong, J. C. Castle, J. Crespo-Barreto, A. B. Bowman, M. F. Rose, D. Kang, R. Richman, J. M. Johnson, S. Berget, and H. Y. Zoghbi.** 2005. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proceedings of the National Academy of Sciences of the United States of America **102:**17551-17558.

86. **Yusufzai, T. M., and A. P. Wolffe.** 2000. Functional consequences of Rett syndrome mutations on human MeCP2. Nucleic Acids Res **28:**4172-9.