

UNIVERSITY OF CALIFORNIA, SAN DIEGO

The TCT motif, a core promoter element for the translational machinery

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in
Biology

by
Trevor John Parry

Committee in charge:

Professor James Kadonaga, Chair
Professor Terence Hwa
Professor William McGinnis

2010

UMI Number: 1474762

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 1474762

Copyright 2010 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



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

Copyright

Trevor John Parry, 2010

All rights reserved.

The thesis of Trevor John Parry is approved, and it is acceptable
in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

I dedicate this thesis to Stephanie, Tom, Julie, and Skyler,
for their love and support.

Table of Contents

Signature Page	iii
Dedication	iv
Table of Contents	v
List of Figures	vi
Acknowledgments	vii
Abstract	viii
I. Introduction	1
II. Results	12
III. Discussion	30
IV. Materials and Methods	36
References	39

List of Figures

Figure 1.	A shared sequence motif at the transcription start sites of ribosomal protein genes	20
Figure 2.	The critical nucleotides for transcription from the TCT	24
Figure 3.	The TCT does not function with a TATA box or DPE	27
Figure 4.	TFIID does not bind the <i>RpLPI</i> core promoter	29

Acknowledgments

I would first like to express my appreciation to Dr. Kadonaga for giving me the opportunity to join his lab. The time I have spent in the lab has not only taught me the skills needed for basic research, but ignited my passion to pursue science as a life-long goal. Dr. Kadonaga has always treated me kindly, and offered support and guidance whenever needed.

I would also like to thank Joshua Theisen for training me in the techniques needed to succeed in the lab, as well as for being a mentor to me. Josh was always willing to help me with experimental design, interpreting results, or troubleshooting, and for his guidance I am very grateful.

Next, I would like to thank all of the members of the Kadonaga lab for their support. I would especially like to thank Arthur Hsu for helping me develop my research skills, and for contributing to a portion of figure 4 in this paper.

I would like to express my gratitude to Dr. Chris Benner, Dr. Chris Glass and Dr. Uwe Ohler for the data that they provided to me; it was paramount to the success of this research.

Lastly, I would like to thank all of my friends and family who have encouraged and supported me while I worked towards this graduate degree.

ABSTRACT OF THE THESIS

The TCT motif, a core promoter element for the translational machinery

by

Trevor John Parry

Master of Science in Biology

University of California, San Diego, 2010

Professor James Kadonaga, Chair

The RNA polymerase II core promoter is a diverse and complex region of DNA that contains the transcription start site. Core promoter elements are conserved DNA sequences within the core promoter that direct transcription to the start site and function in the regulation of gene expression. Here I describe the TCT motif, a newly discovered core promoter element that encompasses the transcription start site and is conserved among the ribosomal protein genes. Mutational analysis of three ribosomal protein gene promoters, *RpLP1*, *RpS12*, and *RpS15*, demonstrates the importance of this DNA element. A severe down-regulation of transcription is seen when the wild-

type sequence is mutated, with positions 3-5 (+3 to +5 relative to the +1 transcription start site) being the most important. I also show that the TCT does not work synergistically with other previously characterized core promoter elements, and the *RpLPI* promoter is not bound by Transcription Factor IID (TFIID). The TCT is a necessary core promoter element for the transcription of the ribosomal protein genes. Its inability to work synergistically with other known promoter elements, along with its inability to bind TFIID, may suggest a parallel mechanism of RNA polymerase II transcription for the ribosomal protein genes that is independent of the mechanism used for all previously studied genes.

I:
Introduction

The core promoter

The core promoter is regarded as the region of DNA that is necessary to drive transcription of RNA polymerase II, typically spanning from positions -40 to +40 relative to the transcription start site (review in Juven-Gershon and Kadonaga, 2009). Two main types of core promoters have been described: focused and dispersed. A focused promoter is one in which transcription starts from a distinct nucleotide, or within a small group of nucleotides and are associated with regulated genes. Dispersed promoters, on the other hand, are ones in which transcription initiates weakly from many sites and are associated most often with constitutively expressed genes. The rest of this paper will only discuss focused core promoters for two main reasons; the majority of research performed on the core promoter has been on focused promoters, and the major mode of transcription in *Drosophila melanogaster*, the organism whose genome was used for this study, appears to be from focused promoters.

RNA polymerase II alone is not sufficient to recognize the core promoter and drive transcription, and a number of basal transcription factors are needed to recruit RNA polymerase II. Transcription Factor IIA (TFIIA), TFIIB, TFIID, TFIIE, TFIIIF, and TFIIF are the essential factors for RNA polymerase II-dependent transcription (Thomas and Chiang, 2006). TFIID is especially important as it is a key basal transcription factor used for the recognition and accurate transcription of focused promoters.

The TATA box, the first known core promoter element

A core promoter element is a conserved sequence in the core promoter that is recognized by the basal transcription factors, and which serves as a nucleating site for the formation of the pre-initiation complex (PIC). The TATA box, an A/T-rich sequence approximately 30 nucleotides upstream of the transcription start site (TSS) was the first core promoter element discovered (Goldberg, 1979). Through mutational analysis, the consensus sequence for the TATA box was originally identified as TATAAA (Chen and Struhl, 1988), though it has been shown more recently that a variety of A/T-rich sequences divergent from the consensus sequence can activate transcription (Singer et al. 1990). The TATA box directs transcription by recruiting TFIID through the binding of a TFIID subunit, the TBP (TATA box Binding Protein). The association of TFIID and the TATA box functions to recruit the rest of the proteins necessary for the PIC and directs RNA polymerase to the TSS. The TATA box is present, however, in only approximately 25% of focused promoters; thus, there must be additional core promoter elements directing the initiation of transcription of focused promoters.

The Initiator

The Initiator was the second core promoter element to be discovered. The transcription of the *deoxynucleotidyltransferase* gene was shown to be dependent upon the DNA sequence encompassing the transcription start site which was termed the Initiator (Smale and Baltimore, 1989). The Initiator has since been found in both *Drosophila* and mammalian promoters. Both computational and mutational analyses

have found the consensus sequence for the *Drosophila* Initiator to be TCA₊₁KTY (IUPAC nucleotide code) where transcription initiates at the A residue designated A₊₁ (Purnell et al. 1994; Ohler et al. 2002). The consensus sequence for mammalian Initiators is more degenerate, found to be YYA₊₁NWYY (Javahery et al. 1994).

The mechanism through which the Initiator functions is not fully understood. It has been shown that transcriptional activity from this core promoter element correlates best with TFIID binding, though it can be utilized by other factors (Verrijzer et al. 1995; Weis and Reinberg 1997). Computational analysis has suggested that the Initiator is the most prevalent core promoter element in the *Drosophila* genome (Ohler et al. 2002), and the Initiator is present in both TATA-containing and TATA-less promoters (Arkhipova 1995). The Initiator has been found to work in tandem with the TATA box to promote transcription, and two core promoter elements downstream of the Initiator have been found whose function is dependent upon the presence of, and distance from, the Initiator.

The Downstream Promoter Element (DPE) and Motif Ten Element (MTE)

The DPE was the next core promoter element to be discovered, found through footprinting analysis using purified TFIID. It was observed that TFIID protected an area of TATA-less promoters approximately 35 nucleotides downstream of the transcription start site (Burke and Kadonaga, 1996). The consensus sequence was found to be RGWCGT located at positions +28 to +33 downstream of the TSS. It was also shown that the spacing between the DPE and Initiator was critical for

transcription. Moving the DPE by a single nucleotide caused a four-fold decrease in transcription (Kutach and Kadonaga, 2000).

The DPE appears to be equally as prevalent in *Drosophila* core promoters as the TATA box. It was observed that a high salt nuclear extract was able to support transcription from DPE-dependent promoters while lower salt extracts could not. Biochemical fractionation of this high salt extract led to the isolation of a factor that was able to activate transcription of DPE-dependent promoters and repress transcription of TATA-dependent promoters (Willy et al. 2000). NC2 (also known as DR1-Drap1) was shown to be a bifunctional transcription factor that functioned in transcription differently depending on the context of the promoter and the core promoter elements present. Later, it was shown that NC2 works with another protein, Mot1, to remove TBP from DNA (Hsu et al. 2008). TBP has been shown to activate transcription from promoters containing a TATA box, and repress transcription of promoters containing a DPE. NC2 functions with Mot1 to block this effect, thus activating transcription from DPE-dependent promoters and repressing transcription from TATA-dependent promoters. This regulatory circuit for transcription is one example of the complexity of the core promoter's influence on transcription.

Using bioinformatics, a second downstream core promoter element was identified. An overrepresented DNA sequence in the *Drosophila* core promoter, labeled Motif Ten, was found with a consensus sequence of CSARCSSAACGS from positions +18 to +29 relative to the TSS (Ohler et al. 2002). Subsequent functional investigations have confirmed the function of this motif as a core promoter element,

the MTE. The MTE is dependent upon the presence of an Initiator, is as position sensitive as the DPE, and is found independently of the TATA box and DPE (Chin et al. 2004; Theisen et al. in submission). Moreover, the MTE, like the TATA box and DPE, is a recognition site for TFIID (Theisen et al. in submission).

The TFIIB Recognition Element (BRE)

The only known core promoter motif that is not a binding site for TFIID is the TFIIB Recognition Element (BRE). The BRE was first discovered as a DNA sequence directly upstream of the TATA box that is bound by TFIIB (Lagrange et al. 1998). It was originally thought that TFIIB binding the BRE helped facilitate the binding of TBP to the TATA box. A second TFIIB binding sequence was found downstream of the TATA box (Deng and Roberts, 2005). The discovery of this second BRE element prompted the upstream BRE to be renamed the BREu, while the downstream BRE is named the BREd. The BRE acts in conjunction with the TATA box in basal transcription, though conflicting data has shown the BRE to both increase and decrease transcription from TATA containing promoters. A more recent study has implicated the BREu in playing a role on enhancer activity on known core promoter elements (Juven-Gershon et al. 2008). Specifically, the presence of a BREu suppresses the ability of the enhancer Caudal to activate transcription from TATA box-dependent promoters.

Regulation of gene expression through the core promoter

All of the mechanisms involved in the expression and regulation of genes ultimately act through the core promoter, where the transcription start site is located. It is easy to imagine, then, that the composition of core promoter elements present in the promoter region can act to regulate specific classes of genes. It was found that nearly all of the Homeotic (Hox) genes in *Drosophila melanogaster* have DPE-dependent promoters. It was then shown that Caudal, a DNA sequence-specific transcription factor involved in regulation of the Hox gene network, preferentially upregulated expression from DPE-dependent promoters (Juven-Gershon et al. 2008). The Hox genes code for transcription factors necessary for the development of many anatomical structures, and the expression and regulation of these genes is dependent on the core promoter elements present in their promoters. This example underscores how core promoter elements can be used to regulate the expression of an important class of proteins. There may be undiscovered core promoter elements that are involved in the expression of key gene networks.

The Ribosome and ribosomal proteins

The ribosome is one of the most important complexes in the cell, responsible for the accurate production of almost all of the proteins in a cell. The eukaryotic ribosome consists of 4 rRNAs (5S, 5.8S, 18S, 28S) and approximately 80 ribosomal proteins (this number varies slightly by species) (Wool, 1979). The rRNA and ribosomal proteins arrange in to two subunits, the small and large subunits. The 40S small ribosomal subunit consists of the 18S rRNA and 33 proteins (termed the RpS

proteins for Ribosomal protein Small subunit). The 60S large subunit consists of the 5S, 5.8S, and 28S rRNAs and 49 proteins (termed the RpL proteins for Ribosomal protein Large subunit). These two subunits associate to form the 80S ribosome. The four rRNAs form the catalytic site responsible for translating mRNA sequences into polypeptide chains. The ribosomal proteins act as scaffolding for the rRNAs. In addition to providing the structural support for the rRNAs, the ribosomal proteins have been implicated in a wide array of non-ribosomal processes including apoptosis, DNA repair, and transcription (review in Lindstrom 2008). The roles of ribosomal proteins in these pathways are still debated.

In addition to the ribosomal proteins, other proteins are involved in the assembly and function of the ribosome. The Eukaryotic Initiation Factors are five multi-subunit proteins that help the large and small subunits associate, stabilize the interaction between the ribosome and the mRNA, and help recruit the tRNAs to the ribosome (Thomas et. al 1981). Two other factors, known as the Eukaryotic Elongation Factors, help in the process of translating the mRNA and making the polypeptide chain (Riis et al. 1990).

A significant amount of energy is spent on the production of ribosomes. For example, in yeast, each cell contains approximately 200,000 ribosomes, and must make about 2000 ribosomes per minute. Moreover, in rapidly growing yeast, 50% of the RNA polymerase II-dependent transcription initiation events occur at the ribosomal protein genes (Warner, 1999). The amount of energy the cell spends on

ribosome biogenesis is significant, which emphasizes how important these complexes are for cell growth and proliferation.

Ribosomal protein stoichiometry is strictly controlled. One of each ribosomal protein is present in every ribosome, except RpLP1 and RpLP2 which have two copies in each ribosome (review in Taylor et al. 2007). Controlling the expression of the ribosomal protein genes must be exquisitely coordinated to maintain equimolar amounts of all 80 proteins. Ribosome biogenesis must be tightly regulated and coordinated at the levels of transcription and translation.

Regulation of the ribosomal proteins and the 5' TOP

Ribosome synthesis occurs at high levels in rapidly growing cells and at low to undetectable levels in quiescent cells, suggesting that there exists a tight regulation of the expression of the genes that make up the ribosome. The ribosomal proteins are a class of constitutively expressed genes, and it is thought that the cell maintains the equimolar amounts of these proteins by controlling their translation. In rat liver cells, the decrease in ribosome synthesis during development of actively growing fetal liver cells in to non-growing adult liver cells is due to decreased association of ribosomal protein mRNAs (rpmRNAs) with polysomes (Aloni et al. 1992). The polysome is a group of ribosomes concurrently translating the same mRNA, and a loss of association of mRNA and the polysome is linked to a loss of translation. There must then be some feature of the ribosomal protein mRNA that allows the cells to recognize and control the translation of these transcripts. Two key characteristics of the eukaryotic rpmRNAs are that they all start with a C residue and they all have a stretch of 4 to 13

pyrimidines known as the 5' terminal oligopyrimidine tract (5'TOP) (Meyuhas et al. 1996). It is uncommon to find eukaryotic transcripts that start with a C residue. Only approximately 17% of mammalian transcripts have a C residue at the cap site (Schibler et al. 1977), and when looking at predicted start sites of 10,981 *Drosophila melanogaster* genes using CAGE analysis, about 1.5% have a C at the transcription start site (C. Benner and C. Glass, pers. comm.). The 5' TOP is present in the 5' untranslated region of the ribosomal protein genes, and plays a role in the association of these mRNAs with the polysome in translational control. A number of different factors have been shown to interact with the 5' TOP of the ribosomal proteins, but it still remains unclear which of these factors actually contributes to the translational control of these genes (review in Hamilton et al. 2006).

The TOP sequence is found in the mRNAs in both ribosomal proteins and other proteins involved in translation, including the initiation and elongation factors. This sequence may play an important role in the translation of all of the components necessary for translation (Meyuhas, 2000). A line of evidence suggests that the ribosomal protein genes are not solely translationally controlled, as a 2-6 fold decrease in transcription has been noted for ribosomal protein genes following mouse myoblast differentiation (Agrawal and Bowman, 1987). The mechanism of the down-regulation of transcription from these genes after differentiation is still not understood, but this result demonstrates that ribosome biogenesis is also controlled at the level of transcription. A conserved 10 nucleotide pyrimidine sequence encompassing the TSS has been described in ribosomal protein promoters using bioinformatics (Perry 2005),

however, the biochemical function of this sequence has not been examined. This element may provide a mechanism for coordinated regulation of ribosomal protein gene transcription.

The synthesis of ribosomes is an energy intensive process, and the cell has adapted mechanisms for controlling the levels of ribosomal proteins. The 5' TOP contributes to this regulation by coordinating translation of the ribosomal protein mRNA. The evidence cited above suggests a synchronized control of the transcription of the ribosomal protein genes, but further work is needed to understand this method of regulation of ribosome biogenesis.

II:
Results

Discovery of a conserved pyrimidine sequence in the promoters of the ribosomal proteins

I began this project by analyzing a database of predicted transcription start sites of 10,981 *Drosophila* genes (Ashan et al. 2009; C. Benner and C. Glass, pers. comm.), looking for promoters that lacked all known core promoter elements but that still had focused start sites as predicted by the Cap Analysis of Gene Expression (CAGE) dataset. This dataset predicts transcription start sites by aligning the sequenced 5' end of mRNAs to the genome, and picks the nucleotide that has the highest frequency of mRNAs starting at that position for each gene. A number of these promoters lacking known elements were cloned from -50 to +50 relative to the TSS, and were transcribed *in vitro*. Briefly, these constructs were transcribed using *Drosophila* nuclear extract, and the resultant transcripts were subjected to reverse transcription/primer extension using a radio-labeled primer, run on a gel and quantified. The transcription of these promoters lacking known elements were also done in the presence of α -amanitin, a chemical inhibitor of RNA polymerase (α -amanitin inhibits RNA polymerase II when at a final concentration of 4 μ g/mL). Using this approach, a number of functional promoters lacking known promoter elements were identified. Of particular interest, the promoters of two ribosomal protein genes, *Ribosomal protein LPI (RpLPI)* and *Ribosomal protein S15 (RpS15)*, were found to transcribe *in vitro* in the absence of α -amanitin, but lacked transcription in its presence (figure 1a). These RNA polymerase II-dependent genes lack any known core promoter elements, yet both have focused transcription start sites.

Next, sequence analysis of these two promoters revealed that both of these genes have a matching pyrimidine tract spanning the TSS, and when the *in vitro* start site of *RpLPI* was mapped, this promoter started at the predicted C₊₁ residue in the pyrimidine sequence (figure 1b). It was important to see if this consensus sequence was a unique feature of these two promoters, or conserved amongst all of the ribosomal protein genes. When the predicted start sites of the 52 ribosomal protein genes for which I have reliable TSS data were aligned, a highly conserved pyrimidine sequence spanning the TSS was apparent (figure 1c). I compiled a table of these 52 ribosomal protein promoters from positions -10 to +10 (figure 1d). 100% of these 52 ribosomal protein genes are predicted to have a cytosine in the +1 position, with positions -1, +2, +4, and +5 also 100% conserved among these promoters. The consensus sequence from -2 to +6 determined from this list of promoters for this pyrimidine sequence is YYC₊₁TTTTYY. We named this ribosomal protein promoter sequence the TCT motif, due to the signature pyrimidine sequence at the TSS. The consensus sequence identified for the TCT motif looks strikingly similar to the pyrimidine stretch identified in the mouse and human ribosomal protein promoters, YYC₊₁TYTTYYY (Perry 2005).

The TCT is necessary for accurate transcription from the ribosomal protein promoters

I have identified a consensus sequence in the *Drosophila* ribosomal protein promoters, but the presence of this sequence does not prove any functionality. To examine the function of the TCT, I set out to make a set of mutations that will test the

importance of 3 nucleotide sections of this sequence, as well as determine what the 5' and 3' borders of the element are. Beginning with the position -4 relative to the TSS, a series of mutations was made that replaced three nucleotide groups of the wild type sequence with 3 guanines. The reason that a GGG mutation was chosen was because we wanted to choose nucleotides that were not represented in the consensus sequence (either A or G), but the addition of adenine in the TCT may accidentally introduce an Initiator (TCA₊₁KTY) into this region of the promoter. The simplest way to mutate out sections of the TCT without adding in an Initiator was to use guanines. These three nucleotide mutations were made across the TCT at positions -4 to -2, -1 to +2, +3 to +5, +6 to +8, and +9 to +11. A second mutation, this time using thymines instead of guanines for positions +6 to +8, was necessary for the promoters *RpLPI* and *RpS15*, because these promoters have two and one guanine in this three nucleotide set, respectively. These scanning mutations were introduced into the promoters for *RpLPI* (figure 2a) and *RpS15* (figure 2b), as well as another ribosomal protein promoter that gave a strong signal *in vitro*, *RpS12* (figure 2c). Along with the transcriptions, a schematic for each promoter showing the wild-type sequence and the locations of the mutations is included. Mutations to -4 to -2, +6 to +8, and +9 to +11 caused modest or inconsistent affects on transcription among the three promoters. When mutating the nucleotides at positions -1 to +2, which span the TSS, a 3-10 fold decrease in transcription can be observed for the three promoters. Mutations to positions +3 to +5 have an even more detrimental effect on transcription than mutating the TSS. A triple G mutation to positions +3-+5 gives a 6 fold decrease in signal for *RpS12*, a 25 fold decrease in signal for *RpLPI*, and no signal can be detected above background for

RpS15. The results from the mutational analysis show that the TCT motif is in fact a functional core promoter element necessary for transcription from the ribosomal protein genes.

The TCT cannot replace an Initiator in TATA or DPE-dependent promoters

When our lab discovers a core promoter element, we often test whether the new element has the ability to work synergistically with either the TATA box or DPE. *Hunchback P2* (*hbP2*), a TATA dependent promoter, and *E74B*, a DPE-dependent promoter, were chosen to test whether the TCT can synergize with these well studied core promoter elements. These promoters were chosen because they are also Initiator-dependent. *In vitro* transcriptions were performed to test the relative ability of the TCT to replace the Initiator in these two promoters, and the signals from the primer extension were quantified relative to their respective wild-type promoters. When the Initiator is mutated out of *hbP2*, a 10 fold decrease in transcription can be observed versus the wild type, while replacing the Initiator with the TCT causes approximately a 33 fold decrease in transcription (figure 3a). Similarly, mutational analysis of the DPE-dependent promoter *E74B*, reveals that a mutation to the Initiator causes a complete loss of signal, and no recovery of transcription is seen when the TCT is added in place of the Initiator (figure 3b). These results demonstrate that the TCT cannot compensate for the loss of the Initiator in either *HbP2* or *E74b*.

TFIID does not bind a TCT-dependent promoter

TFIID recognizes and binds both the TATA box and the DPE. Because no synergy between the TCT and other known core promoter elements was observed, I wanted to test whether TFIID binds the promoter of *RpLP1*, our prototypical TCT-dependent promoter. A TFIID footprint was performed to test how well this transcription factor binds the wild-type *RpLP1* as well as *RpLP1* with a GGG mutation at positions +3 to +5 (figure 4a). There is minimal hypersensitivity or protection seen in the footprint for both the wild-type and mutant promoters, suggesting that TFIID does not interact with or bind either of these promoters. Next, to test whether TFIID truly was not binding *RpLP1*, or if there was some problem with this footprinting analysis or the TFIID itself, footprints were performed on both *hbP2* and *E74B*, promoters containing elements known to interact with this transcription factor. Areas of protection from, and hypersensitivity to, DNaseI can be seen around the TATA box and Initiator of *hbP2* (figure 4a). Likewise, areas of hypersensitivity can be seen at the Initiator, and areas of protection can be seen around the Initiator and DPE of *E74B* (figure 4a). TFIID binding is often a good predictor of transcription strength, so it would be expected that the *RpLP1* promoter would have a much weaker strength of signal than *hbP2* or *E74B* based on the footprint analysis. When the signal strengths of the four promoters were quantified relative to wild-type *RpLP1*, the mutation to positions +3 to +5 again shows a 25 fold decrease in signal, while the combined signal from the two TSS in *hbP2* is slightly stronger than the signal from *RpLP1*. *E74B* transcribes at about half the signal strength of *RpLP1* (figure 4b). Wild-type *RpLP1*

transcribes at relatively the same level as *hbP2* and *E74B* even though TFIID does not interact with the TCT-dependent promoter. The discrepancy between the footprint and the transcription strength shows that the TCT is a core promoter element that is not bound by TFIID, and suggests the transcription from these ribosomal protein promoters may be TFIID-independent.

Figure 1. A shared sequence motif at the transcription start site of ribosomal protein genes

- (A) *In vitro* transcriptions of two wild type ribosomal proteins (*RpLp1* and *RpS15*) in the absence and presence of α -amanitin (at a final concentration of 4 $\mu\text{g}/\text{mL}$)
- (B) *In vitro* start site mapping of *RpLp1*. The sequence on the left is determined from the ladder, and the transcription start site is indicated with the arrow on the C.
- (C) Sequence logo for positions -10 to +11 of 52 ribosomal protein promoters aligned at their transcription start sites.
- (D) A list of the 52 ribosomal protein promoters from -10 to +10 used to generate the sequence logo. The nucleotides in each promoter that match the consensus sequence below the table are highlighted in blue. The transcription start site is annotated in red.

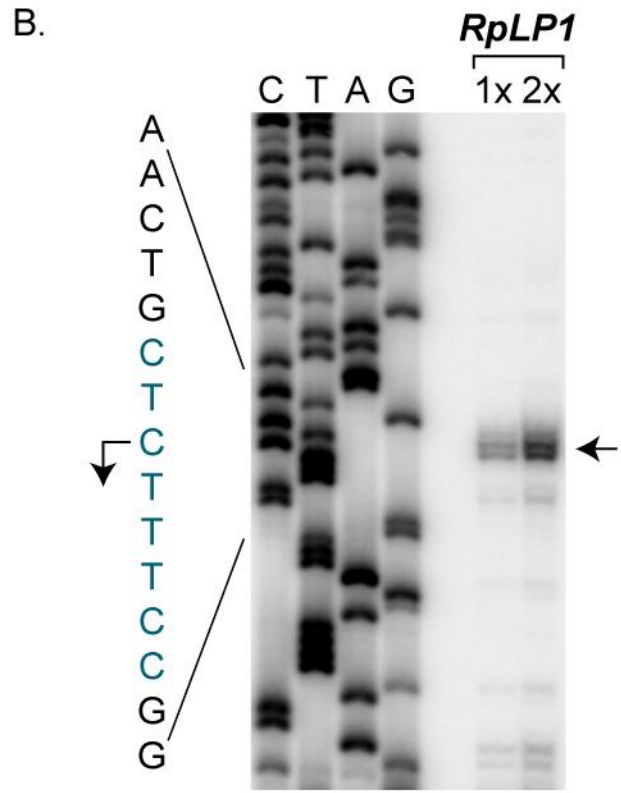
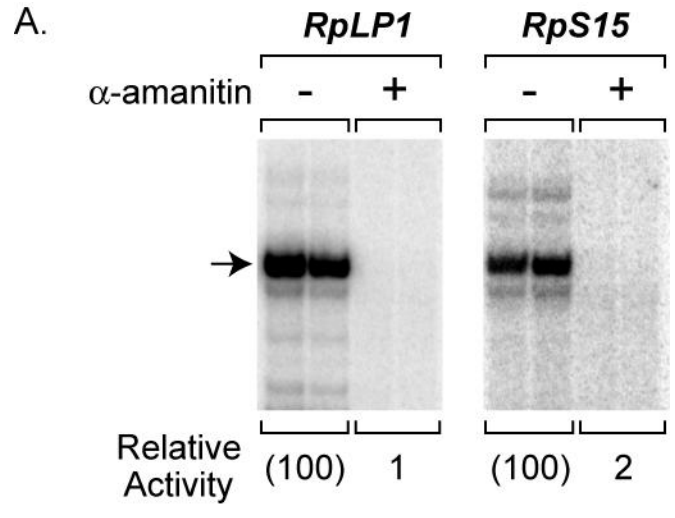
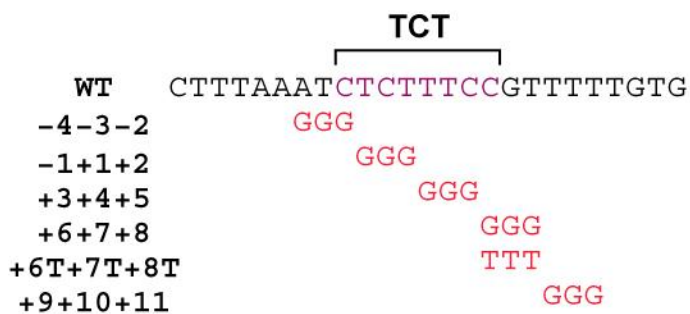
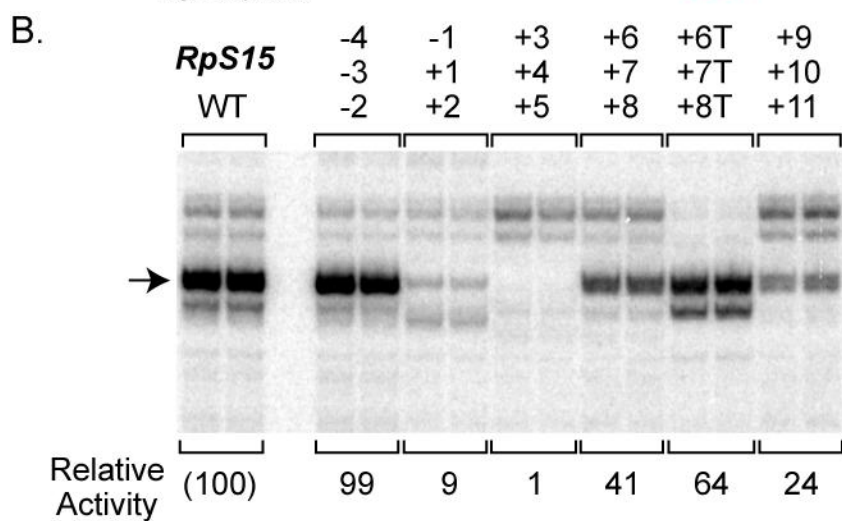
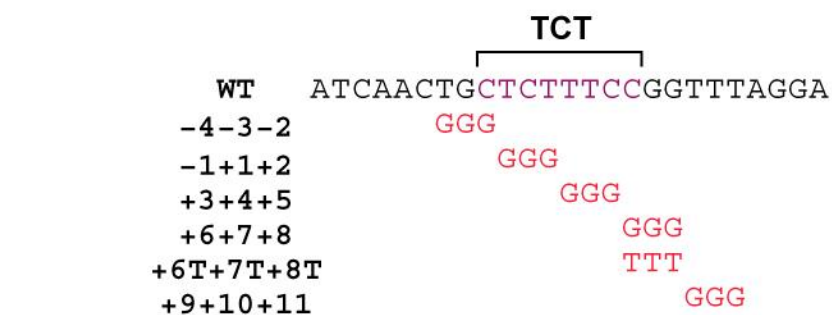
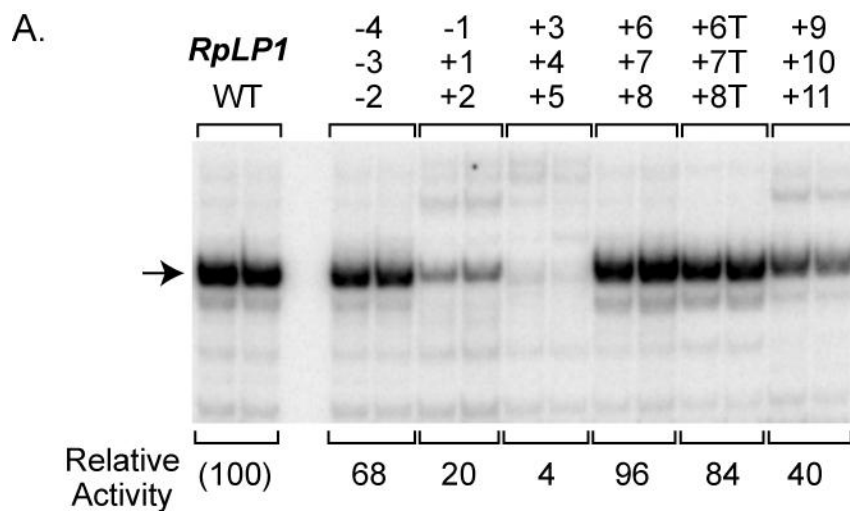


Figure 2. The critical nucleotides for transcription from the TCT

In vitro transcription analysis of a set of mutations made to the *RpLPI* (A), *RpS15* (B), and *RpS12* (C) promoters. Transcripts are quantified relative to the wild-type (WT). The mutations are indicated above the lanes on the gel, and a schematic to the right shows the locations of the mutations (in red) in the promoter region of each gene. The TCT is highlighted in purple.



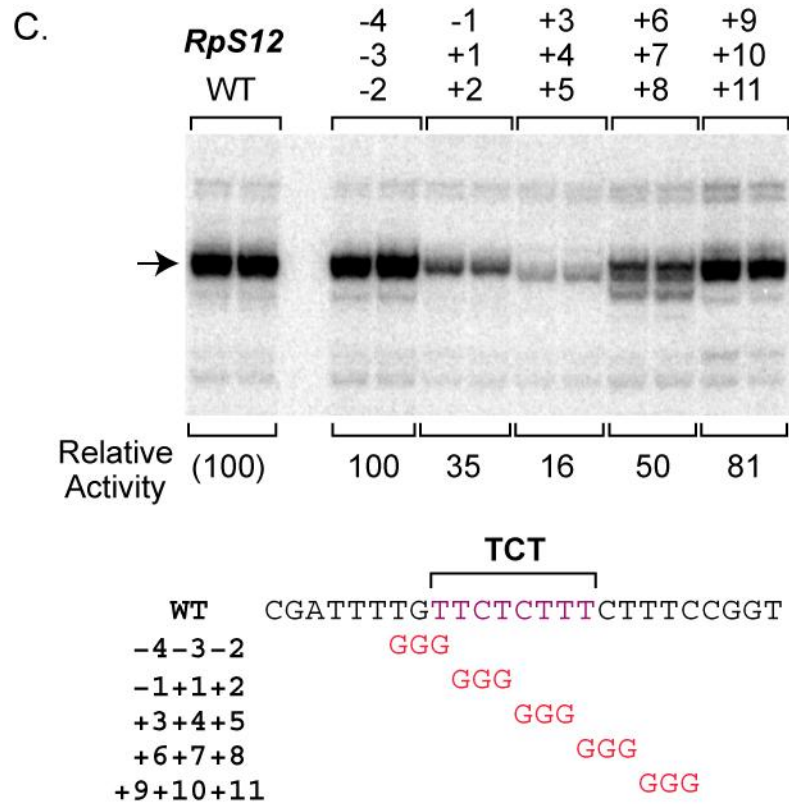


Figure 2. continued

Figure 3. The TCT does not function with a TATA box or DPE

- A) *In vitro* transcription of wild-type *HunchbackP2*, as well as a promoter with a mutant Initiator (mInr), and one which has the TCT added in place of the Initiator (TCT). The schematic on the right describes each promoter. The transcriptions are quantified relative to the wild-type.
- B) *In vitro* transcription of wild-type *E74B*, as well as a promoter with a mutant Initiator (mInr), and one which has the TCT added in place of the Initiator (TCT). The schematic on the right describes each promoter. The transcriptions are quantified relative to the wild-type.

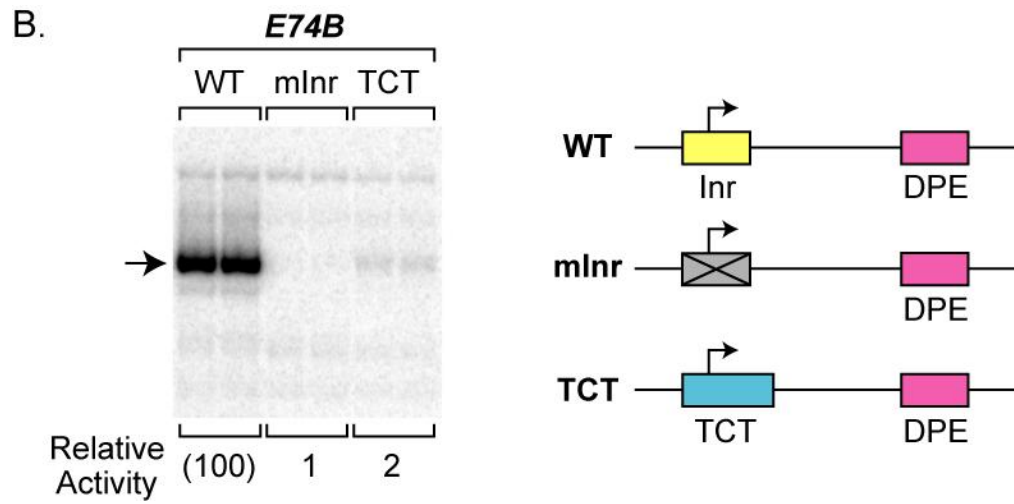
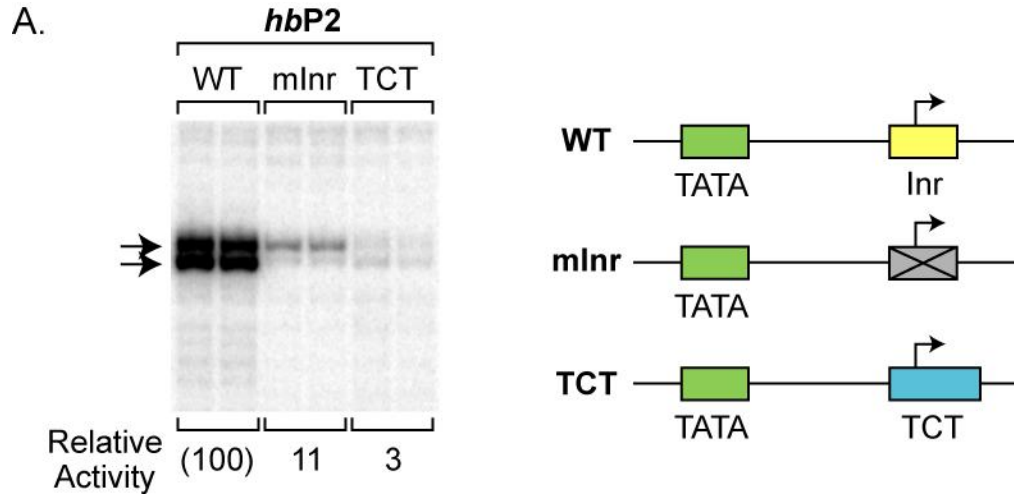
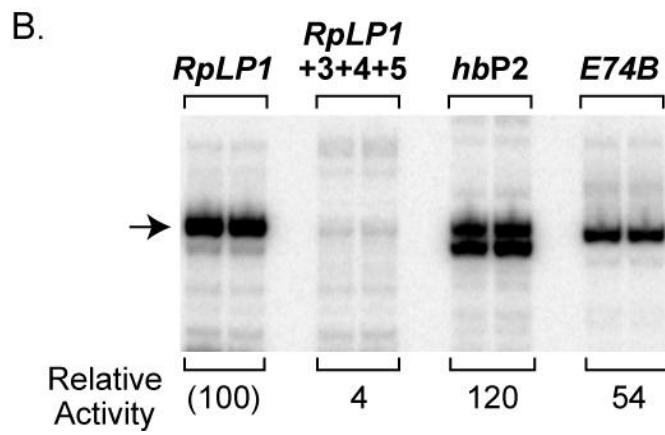
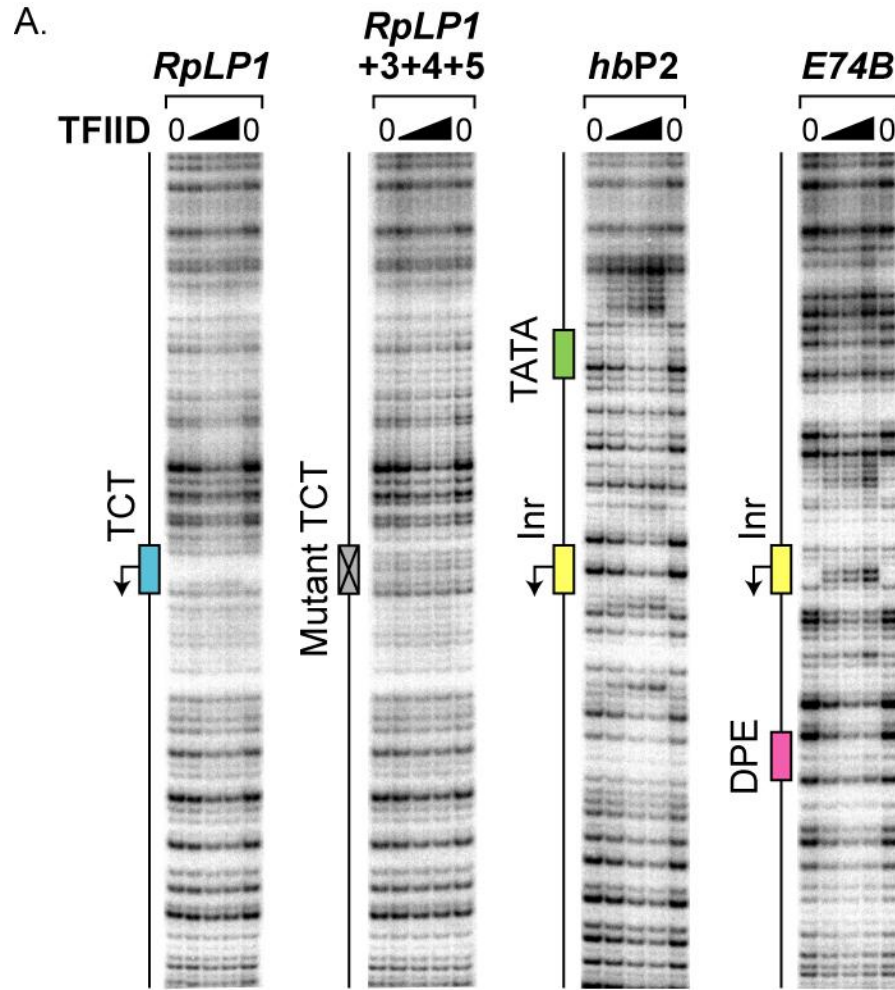


Figure 4. TFIIID does not bind the RpLP1 core promoter

- (A) DNaseI footprint analysis with increasing concentrations purified TFIIID on wild-type *RpLP1*, mutant *RpLP1* (a mutation at positions +3 to +5), *hbP2*, and *E74B* promoters. The schematic to the left of each footprint describes what core promoter elements are present in each promoter.
- (B) *In vitro* transcriptions for the four promoters footprinted in (A). The transcriptions were quantified relative to wild-type *RpLP1*.



III:

Discussion

In this work, I have analyzed the function of a newly discovered core promoter element, the TCT, which is conserved among the ribosomal proteins in *Drosophila melanogaster*. Through mutational analysis, I have shown that the TCT is necessary for transcription of these promoters. I have also shown that the TCT cannot replace the Initiator in two natural promoters and does not function synergistically with either the TATA box or the DPE. Finally, my work shows that TFIID does not bind to the promoter region of TCT-dependent promoters. These results are significant in three ways.

The TCT is a newly discovered core promoter element

In the investigation of promoters lacking any known core promoter elements, two ribosomal protein promoters were identified that had strong *in vitro* transcription signals, and were RNA polymerase II dependent (figure 1). Sequence analysis of these two genes led to the discovery of a strongly conserved pyrimidine sequence that encompassed the TSS from positions -2 to +6 of all of the ribosomal protein genes with reliable predicted TSSs, giving the consensus sequence $YYC_{+1}TTTTY$ (figure 1). Through biochemical analysis of three ribosomal protein gene promoters, I showed that transcription of these promoters is dependent on this new promoter element (figure 2). The core of this sequence appears to be positions -1 to +5, as mutations to this region caused the largest reduction in transcription from three different ribosomal protein promoters. More importantly, mutational analysis revealed that the three most important nucleotides for this sequence are at positions +3 to +5, and the mutation of the wild-type nucleotides at these positions could cause a more severe loss of

transcription from these promoters than a mutation at any other location in the conserved sequence. Surprisingly, the TCT element was more dependent on these three nucleotides than on those spanning the transcription start site, which would seem most critical because this is where RNA polymerase II begins transcription. The functional analysis correlates well with the bioinformatic analysis, showing that the most conserved positions are also the most critical for transcription. The nucleotides at positions +9 to +11, though not as highly conserved as the core of the TCT, seem to play a part in transcription from *RpLPI* and *RpS15*. This may be due to the fact that this sequence falls about 1 turn of the DNA helix (10 base pairs) away from the start site, causing this sequence to be on the same face of the DNA surface as the TCT. Further work is necessary to elucidate the function of this more distal portion of the TCT.

The identification of a new core promoter element conserved in a small set of genes (about 1% of the *Drosophila* genome) reveals how conserved DNA sequences can go unnoticed when looking at large cross sections of promoters. Better tools need to be developed so that promoters will not only be studied in a genome wide scale, but classes of promoters will be grouped together and studied independently to identify class specific DNA motifs. The sequence of the TCT is almost identical to the pyrimidine rich sequence found in murine and human ribosomal proteins (Perry 2005). This suggests that the TCT may be a core promoter element conserved from *Drosophila* to humans, and is essential for the transcription of metazoan ribosomal protein genes.

A new mechanism of RNA polymerase II transcription

I set out to try to understand what other promoter elements the TCT works with, to better understand how this element might be functioning in cells. Upon the addition of the TCT to either TATA box- or DPE-containing promoters, I saw no synergism between the TCT and the previously described elements (figure 3). In fact, the addition of the TCT to the TATA box-containing promoter actually caused a more severe reduction in transcription than mutating out the Initiator. This could be due to the fact that TFIID, the transcription factor that recognizes and binds the TATA box, Initiator, and DPE, cannot recognize or bind the TCT, which had replaced the Initiator in the hybrid promoters.

Footprints of TFIID on the ribosomal protein promoter indicates little or no association between TFIID and *RpLP1* (figure 4). This is important because the lack of synergism between the TCT and the TATA box or DPE, coupled with the lack of interaction between the TCT and TFIID seen in the footprint suggests that there may exist a mode of transcription for the ribosomal protein genes that is independent of TFIID. *RpLP1* transcribes at relatively the same strength as *hbP2* and *E74B*, yet it does not bind the major transcription factor used for focused promoters. Recently it has been shown that after myoblast differentiation, myotubes shift from a TFIID-dependent transcription system to a TRF3/TAF3-dependent system (Deato and Tjian 2007). The data from this study underlines the emerging idea that there exist multiple RNA polymerase II transcription systems that function in parallel to one another. This

work characterizing the TCT may be the first step in discovering another parallel mechanism of RNA polymerase II transcription.

The TCT, a regulatory element for the translational machinery

Through this work, it has become apparent that the TCT is required for transcription from ribosomal protein promoters. The fact that the TCT is so highly conserved among this important group of genes may give insight into how the levels of ribosomal protein gene transcripts, and ultimately the number of ribosomes, is controlled. Work has shown that the 5' TOP affects levels of ribosomal proteins by controlling their translation. Previous work in our lab has shown how core promoters can play a key role in the expression of certain classes of genes (Juven-Gershon et al. 2008). The TCT may allow specific, coordinated regulation of the ribosomal protein genes at the level of transcription. If this is true, the cell can control both the amount of mRNA being transcribed from TCT containing genes as well as the amount of protein that is expressed from the mRNA through the 5' TOP. The TCT may be the master regulator of transcription for ribosome biogenesis.

Looking forward

The work on the TCT may be expanded into genes other than the ribosomal protein genes. The 5' TOP has been shown to exist in many genes encoding proteins involved in translation, and the TCT may control the transcription of many of these genes. In addition, a search of promoter databases for the TCT sequence may identify genes with previously unidentified function in translation.

The number of ribosomes present in a cell is an important part of cell proliferation, as a cell will generate a large amount of ribosomes before it divides, and quiescent cells limit the amount of ribosomal proteins being translated. The TCT may prove to be a *cis*-element that can be exploited to promote cell growth and proliferation through increased ribosome biogenesis, or may be a new target of cancer therapies that act to slow the proliferation of tumor cells.

IV:

Materials and Methods

DNA constructs

All of the promoters used in these experiments, both wild type and mutant, were cloned from -50 to +50 relative to the predicted transcription start site, acquired from Chris Benner's CAGE dataset. The double stranded oligonucleotides for these promoters were cloned into the Xba1 and Pst1 sites in the polylinker of the plasmid pUC119. The scanning mutations were made by replacing the wild type sequences at the noted locations (figure 2) with 3 guanine residues, or in the case of the m+6 to +8 TTT mutation, three thymine residues. The initiator mutation (mInr) for *hbP2* and *E74B* replaced the wild type sequence from positions -2 to +4 with GTG₊₁ACA. The addition of the TCT to *hbP2* and *E74B* was made by replacing the wild type sequence from positions -2 to +8 with CTC₊₁TTTCCGG.

In vitro transcription analysis

All in vitro transcription reactions were performed as previously described (Wampler et al. 1990) using 250 ng of supercoiled DNA constructs with *Drosophila* SK nuclear extract (Soeller et al. 1988). The transcripts then underwent reverse transcription/primer extension using the RS2 reverse sequencing primer (AGCGGATAACAATTTTACACAGGA). Transcription gels were fixed in 10% acetic acid before they were dried. Resultant reverse transcription products were then quantified using a PhosphorImager (GE Health Sciences). All quantifications are products of three independent experiments to ensure accuracy.

DNase 1 footprinting analysis

The TFIID footprint was performed by Dr. Jer-Yuan (Arthur) Hsu, a former post-doc in Dr. Kadonaga's lab. Footprinting probes were first prepared by PCR amplification using the unlabeled upstream universal M13 primer and a 5' ³²P-labeled downstream M13 primer, which flank the promoter region. The PCR products were gel purified on a 5% native acrylamide gel, and a DNaseI digestion/TFIID footprint was performed as previously described (Burke and Kadonaga 1996). The TFIID was purified as previously described (Theisen et al. in submission). Briefly, FLAG-tagged TBP was purified from *Drosophila* S2 cell nuclear extract using DNA affinity chromatography, and then further purified using immunoaffinity chromatography.

Sequence logo

The sequence logo was generated as previously described (Schneider and Stephens, 1990; Crooks et al. 2004) by aligning the predicted start sites of 52 ribosomal proteins taken from our CAGE dataset.

References

- Agrawal M. G. and Bowman L.H. 1987. Transcriptional and Translational Regulation of Ribosomal Protein Formation during Mouse Myoblast Differentiation. *The Journal of Biological Chemistry* **262**: 4868-4875
- Aloni R., Peleg D. and Meyuhas O. 1992. Selective translational control and nonspecific posttranscriptional regulation of ribosomal protein gene expression during development and regeneration of rat liver. *Molecular and Cell Biology* **12**: 2203-2212
- Arkhipova I. R. 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* **139**: 1359-1369
- Ashan B., Saito T. L., Hashimoto S., Muramatsu K., Tsuda M., Sasaki A., Matsushima K., Aigaki T. and Morishita S. 2009. Machibase: a *Drosophila melanogaster* 5'-end mRNA transcription database. *Nucleic Acids Research* **37**: D49-53
- Burke T. W. and Kadonaga J. T. 1996. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes and Development* **10**: 711-724
- Chen W. and Struhl K. 1988. Saturation Mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein. *PNAS* **85**: 2691-2695
- Crooks G. E., Hon G., Chandonia J. M. and Brenner S. E. 2004. A sequence logo generator. *Genome Research* **14**: 1188-1190
- Deato M. D. E. and Tjian R. 2007. Switching of the core transcription machinery during myogenesis. Cold Spring Harbor Laboratory Press **21**: 2137-2149
- Deng W. and Roberts S. G. 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes and Development* **19**: 2418-2423
- Goldberg M. L. 1979. Ph. D. Thesis, Stanford University, Stanford, CA, USA
- Hamilton T. L., Stoneley M., Spriggs K. A. and Bushell M. 2006. TOPs and their regulation. *Biochemical Society Transactions* **34**: 12-16
- Hsu J., Juven-Gershon T., Marr II M. T., Wright K. J., Tjian R. and Kadonaga J. T. 2008. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes and Development* **22**: 2353-2358
- Jahavery R., Khachi A., Lo K., Zenzie-Gregory B. and Smale S.T. 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell Bio* **14**: 116-127

- Juven-Gershon T. and Kadonaga J. T. 2009. Regulation of gene expression via the core promoter and the basal transcription machinery. *Developmental Biology* **339**: 225-229
- Juven-Gershon T., Hsu J. and Kadonaga J. T. 2008. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes and Development* **22**: 2823-2830
- Kutach A. K. and Kadonaga J. T. 2000. The Downstream Promoter Element DPE Appears To Be as Widely Used as the TATA Box in *Drosophila* Core Promoters. *Molecular and Cellular Biology* **20**: 4754-4764
- Lagrange T., Kapanidis A., Tang H., Reinberg D. and Ebright R. 1998. New Core promoter element in RNA polymerase II dependent transcription: sequence specific DNA binding by transcription factor IIB. *Genes and Development* **12**: 33-44
- Lim C. Y., Santoso B., Boulay T., Dong E., Ohler U., and Kadonaga J.T. 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes and Development* **18**: 1606-1617
- Lindstrom M. S. 2008. Emerging functions of ribosomal proteins in gene-specific transcription and translation. *Biochem. and Biophys. Research Communications* **379**: 167-170
- Meyuhas O. 2000. Synthesis of the translational apparatus is regulated at the translational level. *Eur. J. Biochem.* **267**: 6321-6330
- Meyuhas O. Avni D. and Shama S. 1996. Translational Control of Ribosomal Protein mRNAs in Eukaryotes. *Cold Spring Harbor Laboratory Press*: 363-388
- Ohler U., Liao G. C., Niemann H. and Rubin G. M. 2002. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* **3(12)**: 0087.1-0087.12
- Perry R. P. 2005. The architecture of mammalian ribosomal protein promoters. *BMC Evolutionary Biology* **5**: 15
- Purnell B. A., Emanuel P. A. and Gilmour D. S. 1994. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes and Development* **8**: 830-842
- Riis B., Rattan S. I., Clark B. F. and Merrick W. C. 1990. Eukaryotic protein elongation factors. *Trends in Biochemical Science* **15**: 420-424
- Schibler U., Kelley D. E. and Perry R. 1977. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *Journal of Molecular Biology* **115**: 695-714
- Schneider T. D. and Stephens R. M. 1990. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* **18**: 6097-6100

- Singer V. L., Wobbe C. R. and Sturhl K. 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcription activation. *Genes and Development* **4**: 636-645
- Smale S. T. and Baltimore D. 1989. The “initiator” as a transcriptional control element. *Cell* **57**: 103-113
- Soeller W. C., Poole S. J. and Kornberg T. 1988. In vitro transcription of the Drosophila engrailed gene. *Genes and Development* **2**: 68-81
- Taylor D. J., Frank J. and Kinzy T. G. 2007. Translational Control in Biology and Medicine. *Cold Springs Harbor Press*. 59-79
- Theisen J. W. M., Lim C. Y. and Kadonaga J. T. In submission. Three Key Subregions Contribute to the Function of the MTE and DPE Core Promoter Motifs.
- Thomas A. A. M., Benne R. and Voorma H. O. 1981. Initiation of eukaryotic protein synthesis. *FEBS letters* **128**: 177-185
- Thomas M. C. and Chiang C. M. 2006. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Bio.* **41**: 105-178
- Verrijzer C. P., Chen J. L., Yokomori K. and Tjian R. 1994. Binding of TAFs to core promoter elements directs promoter selectivity by RNA polymerase II. *Cell* **81**: 1115-1125.
- Wampler S. L., Tyree C. M. and Kadonaga J. T. 1990. Fractionation of the general RNA polymerase II transcription factors from Drosophila. *J. Biol. Chem.* **265**: 21223-21231
- Warner J. R. 1999. The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Science* (**24**): 437-440
- Weis L. and Reinberg D. 1997. Accurate Positioning of RNA Polymerase II on a Natural TATA-less Promoter is Independent of TATA-Binding-Protein-Associated Factors and Initiator-Binding proteins. *Molecular and Cellular Biology* **17**: 2973-2984
- Willy P. J., Kobayashi R. and Kadonaga J. T. 2000. A basal transcription factor that activates or represses transcription. *Science* **290**: 982-984
- Wool I. G. 1979. The structure and function of eukaryotic ribosomes. *Ann. Rev. Biochem.* **48**: 719-754