6-1-2014

Y- Box Protein 2 (YBX2) is a Major mRNA Specific Regulator of Translation in Spermatogenesis and the Translational Regulation of the Sperm Mitochondria Associated Cysteine Rich Protein (Smcp) mRNA

Tamjid A. Chowdhury
University of Massachusetts Boston

Y-BOX PROTEIN 2 (YBX2) IS A MAJOR mRNA SPECIFIC REGULATOR OF TRANSLATION IN SPERMATOGENESIS AND THE TRANSLATIONAL REGULATION OF THE SPERM MITOCHONDRIA ASSOCIATED CYSTEINE RICH PROTEIN (SMCP) mRNA

A Dissertation Presented

by

TAMJID A. CHOWDHURY

Submitted to the Office of Graduate Studies,
University of Massachusetts Boston,
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

June 2014

Molecular, Cellular and Organismal Biology

Y-BOX PROTEIN 2 (YBX2) IS A MAJOR mRNA SPECIFIC REGULATOR OF TRANSLATION IN SPERMATOGENESIS AND THE TRANSLATIONAL REGULATION OF THE SPERM MITOCHONDRIA ASSOCIATED CYSTEINE RICH PROTEIN (SMCP) mRNA

A Dissertation Presented

by

TAMJID A. CHOWDHURY

Approved as to style and content by:

_____
Kenneth Kleene, Professor Emeritus
Chairperson of Committee

_____
Rick Kesseli, Professor
Member

_____
Linda, Huang, Associate Professor
Member

_____
Alexey Veraksa, Associate Professor
Member

_____
Rahul Kulkarni, Associate Professor
Member

_____
Linda Huang, Program Director
Biology Graduate Program

_____
Richard Kesseli, Chairperson
Biology Department

ABSTRACT

Y-BOX PROTEIN 2 (YBX2) IS A MAJOR mRNA SPECIFIC REGULATOR OF TRANSLATION IN SPERMATOGENESIS AND THE TRANSLATIONAL REGULATION OF THE SPERM MITOCHONDRIA ASSOCIATED CYSTEINE RICH PROTEIN (SMCP) mRNA

June 2014

Tamjid A. Chowdhury, B.S., University of Massachusetts Amherst
Ph.D., University of Massachusetts Boston

Directed by Professor Kenneth C. Kleene

Spermatogenesis is the process by which diploid stem cells differentiate into haploid male gametes, spermatozoa. As haploid cells differentiate into spermatozoa, they undergo profound changes in morphology and physiology, including total reorganization of chromatin in the nucleus which results in the inability to synthesize new mRNAs about one week before the differentiation of sperm is complete. Consequently, many mRNAs, such as the protamine 1 (Prm1) mRNA and sperm mitochondria-associated cysteine rich-protein (Smcp) mRNA, are transcribed in early haploid spermatogenic cells and stored as translationally inactive messenger ribonucleoprotein particles (free-mRNPs) for several days to a week before the mRNA is translated to make protein in late haploid

spermatogenic cells. The initial translational repression is critical for normal sperm development, since premature activation of translation leads to deformed spermatozoa and male subfertility or infertility. The mechanisms that regulate mRNA translation in haploid spermatogenic cells are poorly understood. My research explores the mechanisms of translational regulation of Prm1 and Smcp mRNAs. The goal of my research is to identify factors and elements that repress the translation of Prm1 and Smcp mRNAs in round spermatids and activate their translation in late spermatids.

Prm1 and Smcp mRNA are the only two mRNAs extensively studied through mutations in transgenic mice. Mutation in transgenic mice is the best method of identifying cis-elements in spermatogenic cells. Previous studies of mutations in transgenic mice have identified cis-elements which are necessary to repress the Prm1 mRNA and Smcp mRNAs in transgenic mice. Using UV-crosslinking RNA binding assays, RNA affinity chromatography and mass spectrometry sequencing, my research demonstrates that Y-box protein 2 (YBX2) binds both elements suggesting that YBX2 is an important mRNA specific translational repressor. UV-crosslinking assays also reveal that YBX2 is less specific in its binding specificity than was previously known, implying that YBX2 represses many mRNAs. My research also demonstrates that proper repression of the Smcp mRNA in early haploid cells requires interactions between multiple elements in the 5'UTR and the 3'UTR.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Figures Page

Supplemental Figures

LIST OF TABLES

CHAPTER 1

MECHANISMS OF TRANSLATIONAL CONTROL IN SPERMATOGENESIS

1.1 Introduction

Translation of mRNA into proteins is a complex process, especially in eukaryotes. The process of translation can be subdivided into 3 major phases: initiation, elongation and termination (Merrick et al., 2000). Of the 3 phases, initiation is the most complex; it involves at least 11 initiation factors and over 30 polypeptides (Sonnenberg and Hinnebusch, 2009). Regulation of translation, mostly during translational initiation, is an important and widely used mechanism, at least in eukaryotes, for the regulation of gene expression (Gebauer and Hentze, 2004). Translational control is also critical in mammalian spermatogenesis where transcription ceases in post-meiotic spermatids due to chromatin remodeling (Meistrich et al., 2003). This chapter briefly introduces the molecular mechanisms of translational control, spermatogenesis, translational control in spermatogenesis, the role of Y-box proteins in translational repression, the translational control of the *Sperm Mitochondria associated Cysteine rich Protein (Smcp)* mRNA and the objectives of this dissertation.

1.2 Molecular mechanisms of translational control

Translation begins with "initiation" which can be defined as the process of assembly of elongation competent 80S ribosome onto a mature mRNA (mRNA devoid of introns and possesses a 5'cap and a poly(A) tail) (Jackson et al., 2010). In fact, the 80S ribosome, a large complex protein made up of the 40S and 60S ribosome, is assembled on the mRNA to be translated in several steps. First, the heteromultimeric eukaryotic initiation factor eIF4F (made up of cap-binding protein eIF4E, the RNA-dependent helicase eIF4A, the scaffolding factor eIF4G, and the factors eIF4B and eIF4H) binds the 7-methyl guanosine cap at the 5' terminus. Then, the 43S pre-initiation complex (consisting of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF2-GTP, eIF5) charged with the initiator tRNA (Met-tRNAi) binds eIF4F through eIF3 to form the 48S pre-initiation complex which scans the 5'UTR for the initiation codon. However, alternate methods to scanning such as internal ribosomal entry (IRES) or ribosome shunting are also used in certain cases (Sonnenberg and Hinnebusch, 2009). Once an initiation codon in a strong context is recognized, as defined by Kozak (1991) (a purine in -3 position and a G at +4 with the A of the first AUG being +1), eIF1 and eIF1A causes a conformational change in the 48S complex that allows eIF5 to hydrolyze the GTP of eIF2. This causes the expulsion of eIF2-GDP from the complex which is followed by the expulsion of eIF5, eIF1, eIF1A and eIF3, building a platform for the binding of the 60S subunit and commitment to translation (Jackson et al., 2010).

Another thing that happens during the initiation of translation is circularization of the mRNA, also known as the formation of the "closed loop" (Sonnenberg and

Hinnebusch, 2009). Circularization is achieved by the binding of the poly(A) binding protein, PABP, to the poly(A) tail in the 3'UTR and the scaffolding protein eIF4G (Mazumder *et al*, 2003). Circularization or closed loop formation increases the efficiency of initiation and reinitiation by post-termination ribosomes (Sonnenberg and Hinnebusch, 2009). Translation initiation is the rate limiting step of translation and it can be regulated in several different ways, both at a wide global level as well as an mRNA specific level.

Global control of translation is generally achieved by preventing interactions between initiation factors or their regulators through changes in their phosphorylation state (Gebauer and Hentze, 2004). For example, in order for a functional 43S ternary complex to form, eIF2 must be charged with GTP. eIF2 is made up of 3 subunits – α, β and γ. The exchange of GDP for GTP on eIF2 is blocked by the phosphorylation of the α subunit of eIF2 (Gebauer and Hentze, 2004). Thus, by modulating the phosphorylation state of eIF2, translation of all the mRNAs in a cell can be controlled.

mRNA specific translational control is achieved, most typically, through the interaction of a *cis* element in the 5' or 3'UTR with its *trans* factor such as an RNA-binding protein (RBP) or small non-coding RNA (Kleene, 2013). For example, the iron regulatory proteins, IRP1 and IRP2, bind the iron-responsive (*cis*) element (IRE) in the 5'UTR of ferritin heavy and light – chain mRNAs and prevent the 43S complex from binding the 5' end of the mRNA (Gebauer and Hentze, 2004). *Cis* elements can also be found in the 3'UTR where they bind specific *trans* factors that interact with eIF4F and disrupt "closed loop" formation (Gebauer and Hentze, 2004).

mRNA specific translational control can also be achieved at post-initiation step through upstream open reading frames (uORFs) in the 5'UTR. uORFs lie between the 5'terminus and the principle coding region, and may harbor AUG in strong context to initiate translation. However, uORFs usually sediment with monosomes or small polysomes which means that they only code for very short peptides (Kleene, 2003). It is estimated that 30%-50% of human mRNAs contain uORFs which have varied effect on translation, from fine tuning to repression (Bartel, 2009; Fabian & Sonenberg, 2012).

1.3 Spermatogenesis

The process of male germ cell differentiation from diploid stem cells into haploid gametes, spermatozoa, is known as spermatogenesis. During spermatogenesis, developing sperm cells progressively move from the basal lamina towards the lumen of the seminiferous tubules as they mature. This is a highly regulated process which has been divided into three main phases based on morphology and physiology of the developing cells: the proliferative phase in which diploid spermatogonia replicate rapidly in the mitotic cycle, the meiotic phase in which spermatocytes undergo segregation, crossing over and reduction in chromosome number; and the differentiation phase in which haploid spermatids develop the highly specialized organelles of the mature spermatozoon (reviewed in Russell et al., 1990; Tanaka et al., 2007). The process of spermatogenesis takes about a month in mice and two months in humans (reviewed in Tanaka et al., 2007).

Spermatogonia are the stem cells of the male germ line. The diploid spermatogonia undergo multiple mitotic divisions, and then withdraw from the cell cycle to become primary spermatocytes, which undergo meiosis, and divide once to become secondary spermatocytes with diploid chromosome number (reviewed in Tanaka *et al.*, 2007). Secondary spermatocytes divide again to become spermatids with haploid chromosome number. Spermatids undergo a remarkable process of differentiation known as spermiogenesis which involves major changes in the structure of the nucleus and flagellum and the synthesis of many proteins that are not synthesized in any other cell type in the mammalian body (Eddy and O'Brien, 1994).

Based on the morphology of the acrosome (a spermatogenic-cell specific secretory vesicle), tail and nucleus, spermatids can be categorized into different developmental stages which are referred to as "steps" and are denoted by Arabic numerals. Morphologically, haploid spermatogenic cells can be classified into 16 developmental stages in mice (Russell et al., 1990). The developmental stages of spermatids are referred to as "steps" to distinguish them from developmental stages of seminiferous tubules (Russell et al., 1990).

Spermatogenesis occurs in overlapping waves with the germ cells developing in synchrony in localized regions of seminiferous tubules (Russell et al., 1990; Eddy, 2002). A cross-section of the seminiferous tubule reveals that spermatogonia are located at the periphery of the tubules and increasingly mature cells are located toward the center (Russell et al., 1990; Eddy, 2002). Due to synchronized development of the germ cells in short segments of seminiferous tubules, particular cell types are associated with specific

regions of the tubule as diagrammed in Figure 1.1. Cell associations, also referred to as "stages of the seminiferous cycle", are designated by Roman numerals. There are 14 stages in rat and 12 in mice (Russell et al., 1990). In mice, the sequential progression of the stages of sperm development is complete ~34 days after birth. The mitotic phase lasts about 11 days, the meiotic phase lasts about 10 days and post-meiotic phase lasts about 13 days (reviewed in Russell et al., 1990; Eddy, 2002).

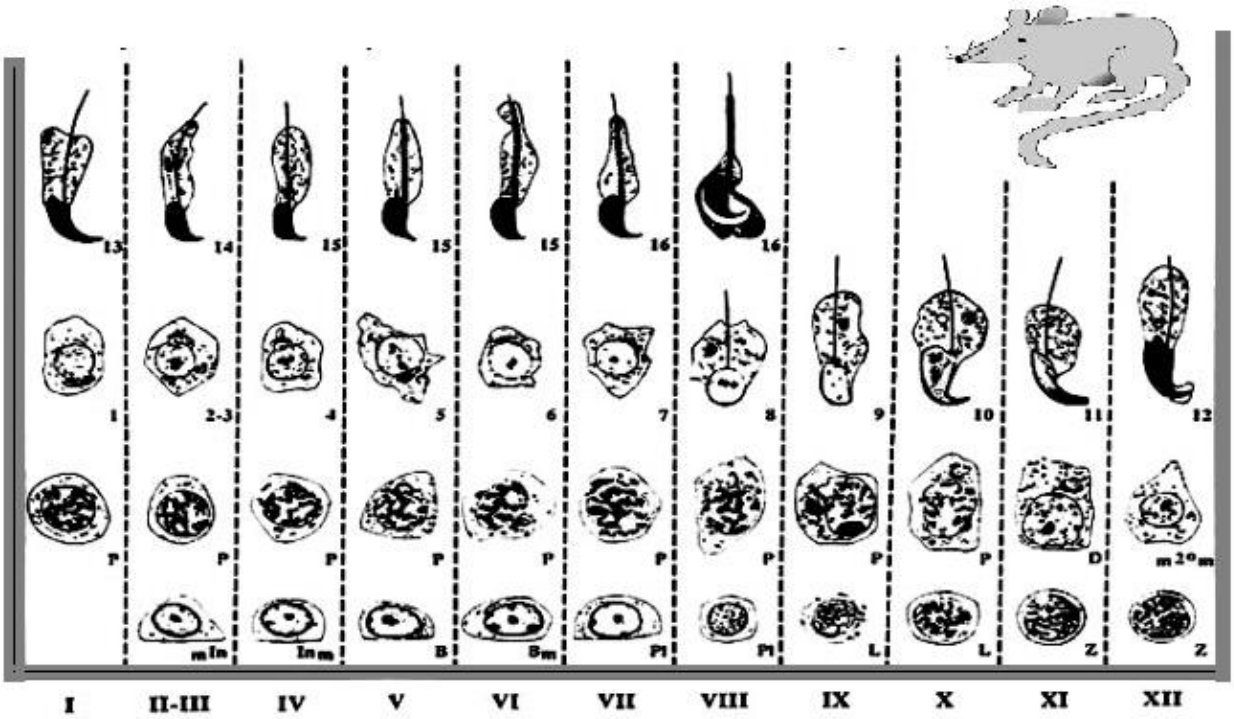


Figure 1.1 Spermatogenic cell-types and cell associations in the stages of the seminiferous cycle of the mouse. The stages of the seminferous cycle are represented by Roman numerals, the steps of spermatids are represented by Aarabic numerals, and the stages of meiosis are represented by letters (PL, preleptotene; L, leptotene; Z, zygotene; P, pachytene; D, diplotene). The vertical dashed lines show cell-associations present at each stage. The cells at the bottom of the figure are at the periphery of the tubule and the cells at the top are adjacent to the lumen. This figure is adapted from Russell *et al.* (1990)

1.4 Translational control in mammalian spermatogenesis

In mice, transcription ceases at step 11 of spermatogenesis (Kierszenbaum and Tres, 1975; reviewed in Kleene, 1996). In general, early haploid cells in steps 1-8 (round spermatids) are transcriptionally active but late haploid cells, elongated spermatids in step 12-16, are not (Kierszenbaum and Tres, 1973; reviewed in Kleene, 1996). The inactivation of transcription happens because the histones (the basic chromosomal proteins involved in the packaging of double stranded DNA in eukaryotic cells and spermatogenic cells through round spermatids) are replaced first by transition proteins (TNP1 and TNP2) and finally by protamines (PRM1 and PRM2). Protamines are small basic proteins that fold DNA in a zigzag fashion and enable the strands of DNA to be closer together than histones, packaging DNA into a volume about one-twentieth of that in somatic cells (reviewed in Eddy 2002; Balhorn, 1982). The lack of transcription in the elongated spermatids means that gene expression has to be controlled by other mechanisms, translational control being the major mechanism though other mechanisms such as protein inactivation are possible. Since transcription in haploid spermatogenic cells stops at step 11, mRNAs that are translated only in elongated spermatids are transcribed in early haploid cells and stored initially in a repressed condition as translationally inactive free messenger ribonucleoprotein particles (free mRNPs) that are not associated with ribosomes. Translation of these mRNAs is activated at specific developmental steps, usually days after the mRNA has been transcribed (Kleene, 1996).

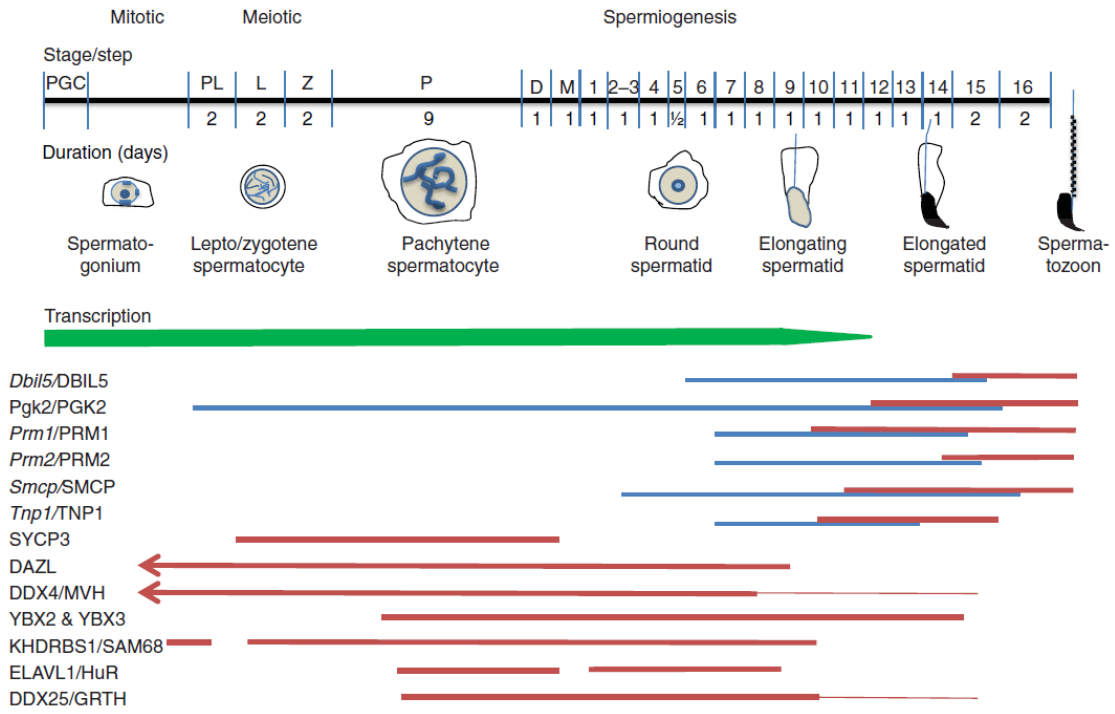


Figure 1.2 Developmental expression of mRNAs and proteins during spermatogenesis in prepubertal and adult mice. The stages of meiosis are denoted PL, preleptotene; L, leptotene; Z, zygotene; P, pachytene; D, diplotene; and M, meiotic divisions 1 and 2. Spermiogenesis is subdivided into three main phases in mice: round spermatids, steps 1–8; elongating spermatids, steps 9–early 12; and elongated spermatids, steps 12–16. The diagrams of spermatogenic cells illustrate the gross morphology, and the black nuclei symbolize the highly condensed chromatin of elongated spermatids and spermatozoa (adapted from Russell et al. (1990)). The decrease in transcriptional activity in elongating spermatids is based on the study of Kierszenbaum & Tres (1975). The duration of each phase of meiosis and steps of spermiogenesis have been rounded to the nearest number of days (Russell et al. 1990). The stages of expression of mRNAs and proteins are depicted respectively by blue and brown lines. The thickness of the lines symbolizes the levels of expression, and the arrows on the left ends of the lines depicting the expression of DAZL and DDX4/MVH indicate that both proteins are expressed in utero. The sources of information concerning the stages of expression of mRNAs and proteins are as follows: Dbil5 and Pgk2 mRNAs and proteins, cited in the text; Prm1, Prm2, and Tnp1 mRNAs and proteins (Mali et al. 1989, Meistrich et al. 2003); Smcp mRNA and protein (Shih & Kleene 1992, Cataldo et al. 1996); DDX4/MVH (Onohara et al. 2010); GRTH/DDX25 (Onohara & Yokota 2012); DAZL (Brook et al. 2009); SYCP3 (Cohen et al. 2006); YBX2/MSY2 and YBX3/MSY4 (Oko et al. 1996, Davies et al. 2000, Giorgini et al. 2001); ELAVL1/HuR (Nguyen Chi et al. 2009); and KHDRBS1/SAM68 (Paronetto et al. 2009).

Translational regulation in spermatids is demonstrated by two general experimental approaches. First, a variety of approaches including analyses of Northern and Western blots of staged prepubertal mice and purified spermatogenic cells, and *in situ* hybridization and immunocytochemistry demonstrate that certain mRNAs first appear in early haploid cells and the protein first appears much later in elongating or elongated spermatids (Kleene, 1996). Second, the translational activity of particular mRNA species can be quantified through sucrose and Nycodenz gradient analysis. Sucrose gradients separate free mRNPs and polysomes by differences in sedimentation velocity while Nycodenz gradients separate free mRNPs and polysomal mRNA by differences in buoyant density. The proportion of total mRNA associated with polysomes is known as polysomal loading and is considered usually to be a measure of translational efficiency. At present, about 14-20 mRNAs have been rigorously identified that show developmental lags between the appearance of the mRNA and protein (Kleene, 1996; Chowdhury and Kleene, 2012), and five of these mRNAs have been shown to be stored in free-mRNPs in early haploid cells and actively translated on polysomes in late haploid cells, but the number is thought to be much greater. In addition, it is known that while most mRNAs in pre-meiotic and testicular somatic cells show high polysomal loading (85-90%), all mRNAs in meiotic and haploid spermatogenic cells exhibit lower levels of polysomal loading, from essentially none to no more than 65% (reviewed in Kleene, 1996). Thus, mRNA translation is also globally repressed in spermatogenic cells (Kleene, 1996; Schmidt et al., 1999).

1.5 Identification of *cis*- element that regulate mRNA translation in spermatids by analysis of mutations in transgenic mice

It is important to identify the regulatory sequences involved in translational regulation in order to completely understand the mechanisms of translational control during spermatogenesis. Unfortunately, rapid methods for investigating translational regulation in cell culture or cell-free translation systems have not been established, and the atypical patterns of gene expression in spermatogenic cells invalidate systems based on somatic cells. Thus, the only rigorous method for studying translational control elements in spermatogenesis is the use of mutations in transgenic mice. Mutations in elements can show exactly how an mRNA is translationally regulated. In addition, it can provide decisive evidence whether a specific factor has a large or small effect on translation (reviewed in Kleene, 2013).

Currently, the *Prm1* mRNA is the only mRNA that has been studied extensively in an attempt to identify the *cis*-elements that control translation in spermatids. *Prm1* mRNA is transcribed in step 7 spermatids but the protein is not seen until step 10 (reviewed in Kleene, 2013). Working with transgenic mice, Braun *et al.* (1989) showed that the *Prm1* 3'UTR is solely responsible for the translational control of *Prm1* mRNA. Subsequent work in transgenic mice showed that the translational repression of the *Prm1* mRNA in spermatids is mediated by two copies of a *cis*-element known as the translational control element (TCE), GAAAAAUGCCACCGUC and GAACAAUGCCACCUGUC, located just upstream of the poly(A) signal (Zhong et al. 2001). Although TCE is both necessary and sufficient for the translational repression of

Prm1 mRNA, *Prm1* 3'UTR also contains another cis element: a 7 nt element, 5'UCCAUCA3', that binds Y-box proteins - YBX2 (MSY2) and YBX3L (MSY3L/MSY4) (Fajardo et al., 1994; Davies et al., 2000; Giorgini et al., 2001). This sequence, known as the Y-box recognition sequence (YRS), does not cause any delay in protein expression in its original position 110 nt upstream of the poly(A) signal; but when put 16 nt upstream of the poly(A) signal by deleting the center of the *Prm1* 3'UTR, it causes a partial delay in hGH expression (Fajardo et al. 1997; Zhong et al., 2001).

1.6 Y-box proteins

Y-box proteins are multifunctional ssDNA and ssRNA binding proteins. Eukaryotic Y-box proteins contain an alanine and proline rich N-terminus, a highly conserved central cold-shock domain (CSD), and a variable C-terminal domain containing alternating clusters of acidic/hydroxyl and basic/aromatic residues (reviewed in Matsumoto and Wolffe 1998; Eliseeva et al, 2011). Y-box proteins are known to function in every aspect of gene expression including the activation and repression of transcription, alternative splicing, translational activation and repression, and stability (Eliseeva et al. 2011). However, their role as mRNA specific translational repressor is particularly well studied. Y-box proteins have been found to be the major mRNA binding proteins in the cytoplasm (reviewed in Eliseeva et al., 2011).

```
YBX3S   MSEAGEATTGGTTLPQAAADAPAA---------APPDPAPKSPAASGAPQAPAPAALLA 50
YBX3L   MSEAGEATTGGTTLPQAAADAPAA---------APPDPAPKSPAASGAPQAPAPAALLA 50
YBX1    MSSEAETQQ------PPAAPAAAL---------SAADTKPGS-TGSGAG---------S 34
YBX2    MSEAEASVVATAAPAATVPATAAGVVAVVVPVPAGEPQKAGGGAGGGGGAASGPAAGTP 59
                  . .*:         .:. *.*           :     . . :..*.
 .
```

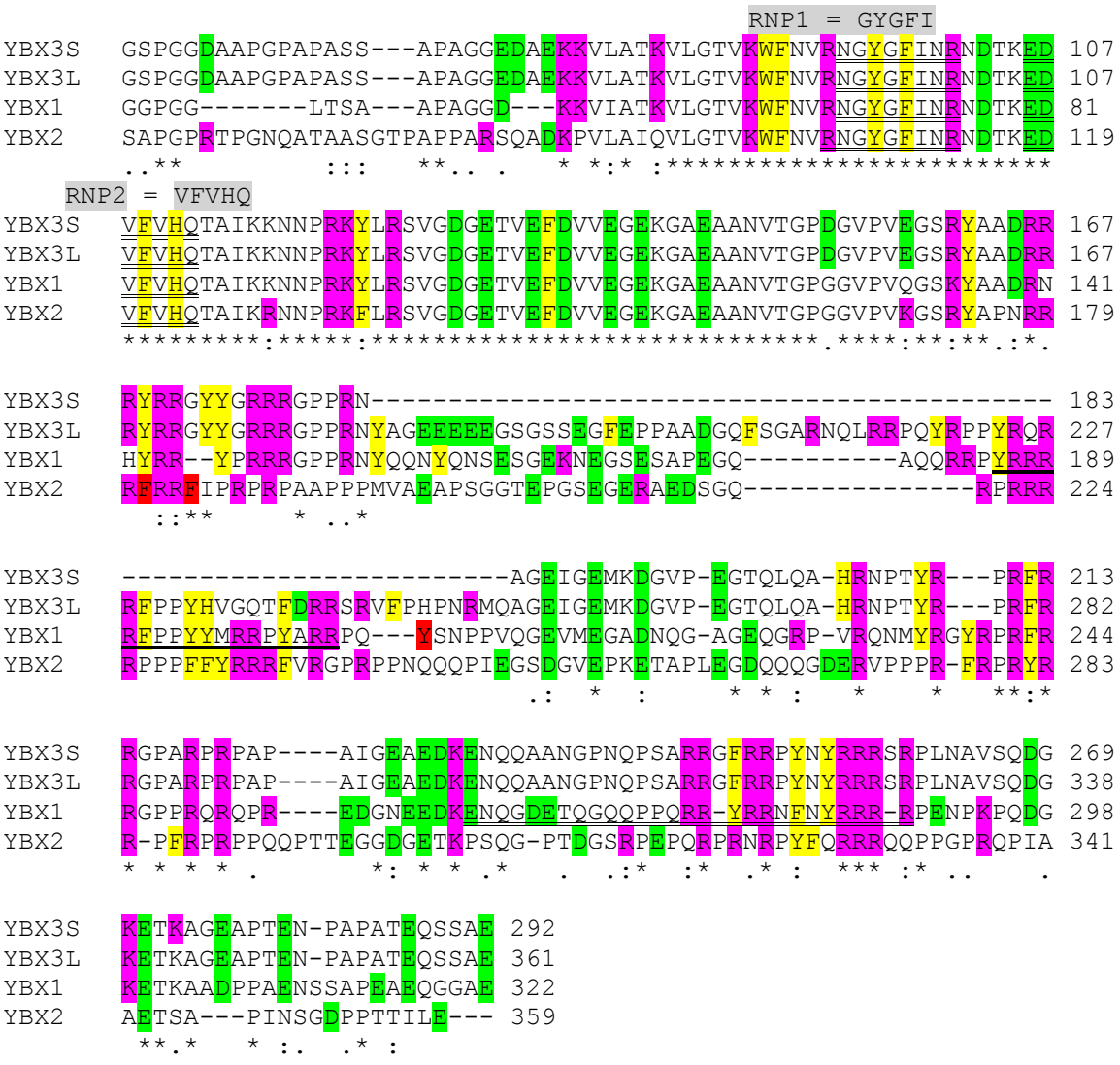


Figure 1.3 Clustal omega alignment of the conserved functional organizations of *Mus musculus* Y-box proteins YBX1, YBX2, YBX3S and YBX3L. Sites exhibiting identical amino acids, high similarity and low similarity are denoted by *,: and ‘ respectively. Acidic amino acids (D and E), aromatic amino acids (F, H, Y, and W), and basic amino acids (K and R) are highlighted green, red and magenta, respectively. Note the high conservation of the cold shock domain and alternating basic/aromatic and acidic islands in the C-terminal domain.

In mice, there are 3 Y-box protein genes – *Ybx1 (Msy1, YB-1, Csdb, DbpB, Nsep1), Ybx2 (Msy2, DbpC, contrin),* and *Ybx3 (Msy3, Msy4, Csda, Dbpa).* The meiotic

and haploid spematogenic cells in mice express 3 isoforms of Y-box proteins encoded by alternatively spliced transcripts of *Ybx2* and *Ybx3*, YBX2 (MSY2) and YBX3S (MSY3S) and YBX3L (MSY3L/MSY4) (Mastrangelo and Kleene, 2000), while somatic cells only express YBX1. Figure 1.2 depicts a CLUSTAL Omega alignment of YBX1, YBX2, YBX3S and YBX3L. Mouse testis also expresses high levels of *Ybx1* mRNA, but the levels of YBX1 protein are very much lower because virtually all of the *Ybx1* mRNA in testis is in free-mRNPs (Tafuri et al., 1993; Cataldo et al., 1999; Mastrangelo and Kleene, 2000). The multiplicity of Y-box protein isoforms is one of the many differences in the pattern of gene expression between somatic and spermatogenic cells, and presumably reflects the unusual complexity of post-transcriptional control in spermatids.

YBX2 and YBX3 are first detected in early pachytene spermatocytes (stage V); their levels increase in late pachytene and reach a maximum plateau in round spermatids, and then decrease progressively to an undetectable level in elongated spermatids (Oko et al., 1996; Davies et al., 2000; Giorgini et al., 2001). The patterns of YBX2 and YBX3 expression suggest that they are involved in the translational repression of many mRNAs in round spermatids. Studies of gene knock-out shows that YBX2 may play a more prominent role in translational repression as compared to YBX3. YBX2 knock out in male mice causes spermatogenesis to terminate in post meiotic germ cells with no sperm seen in epididymis, by premature translational activation and degradation of pathways specific for polysomal mRNAs (Yang et al., 2005). In comparison, *Ybx3* null mutation causes reduced epididymal sperm count without impairing sperm differentiation (Lu et al., 2006). However, overexpression of YBX3L causes defects in sperm differentiation

and fertility (Giorgini et al., 2002). Thus, the phenotype of YBX3L over expression is much more deleterious than the knock-out.

Y-box proteins show both specific and non-specific interactions with ssRNA. Y-box proteins bind ssDNA and ssRNA specifically through the RNP-1 and RNP-2 motifs in the CSD. CSD binds bases by intercalation and hydrogen bond formation through aromatic amino acids without contacting the sugar-phosphate backbone (Sachs et al., 2012; Mayr et al., 2012). The C-terminal domain weakly interacts with the sugar-phosphate backbone non-specifically through the basic-aromatic islands.

Bouvet et al. (1995) showed that the Y-box proteins in *Xenopus laevis* oocyte, FRGY1 and FRGY2, bind a hexameric sequence - 5'AACAUC 3', which they referred to as the Y-box recognition sequence (YRS). The YRS in mammalian spermatids was further refined by Giorgini et al. (2001) who carried out an extensive mutation analysis of the wild-type mouse *Prm1* 3'UTR YRS, 5'UCCAUCA 3'. By mutating each base in the YRS and flanking bases and analysis of protein binding by RNA EMSA using protein extracts from total adult testis, they demonstrated that a seven nucleotide sequence is necessary for protein binding and that the bases outside this segment are not important. The sequence of the YRS was further examined by RNA EMSA analysis of all single nucleotide mutations by competition assays in which a radioactive probe containing mutations is competed with variable amounts of non-radioactive wild-type *Prm1* YRS. These studies demonstrate that the YRS in the mouse *Prm1* 3'UTR contains several degenerate sites, so the YRS consensus sequence can be represented as 5' [U/A/C][A/C]CA[U/C]C[A/C/U] 3'. The nucleotides in brackets are alternative bases

that exhibit maximal binding. Parallel studies using the yeast three hybrid system with cloned YBX2 and YBX3L verify the degenerate YRS consensus sequence and demonstrate that YBX2 and YBX3L bind the same YRS. However, yeast three hybrid system also showed that a G in the first position is capable of binding YBX2 and YBX3L (Giorgini et al. 2001). The disagreement between yeast three hybrid data and RNA EMSA data may be due to a technical issue – the use of RNaseT1 in RNA EMSA which digests RNA at Gs. A common feature of the YRS is the presence of CA(U/C)C but there is disagreement in the length of the sequence to which Y-box proteins bind, ranging from 4 to 8 nt (reviewed in Eliseeva et al., 2011).

Models for the regulation of mRNA translation by Y-box proteins with mRNA integrate the relatively specific binding and RNA chaperone acitivities of the cold shock domain, the relatively non-specific binding and aggregate-forming ability of the C-terminal domain, and the concentration-dependent interactions of Y-box proteins with mRNA, Figure 1.3. At low protein to mRNA ratios *in vitro*, YB-1 binds mRNA as individual molecules and the secondary structure melting activity of the cold shock domain creates mRNPs with an extended open configuration that facilitates binding of ribosomes and initiation factors and the initiation of translation. At higher ratios of YB-1 to mRNA, the B/A and acidic islands in C-terminal domain form charged zippers which package long segments of mRNA into compact mRNPs *in vitro* that are inaccessible to initiation factors and ribosomes during translational initiation.

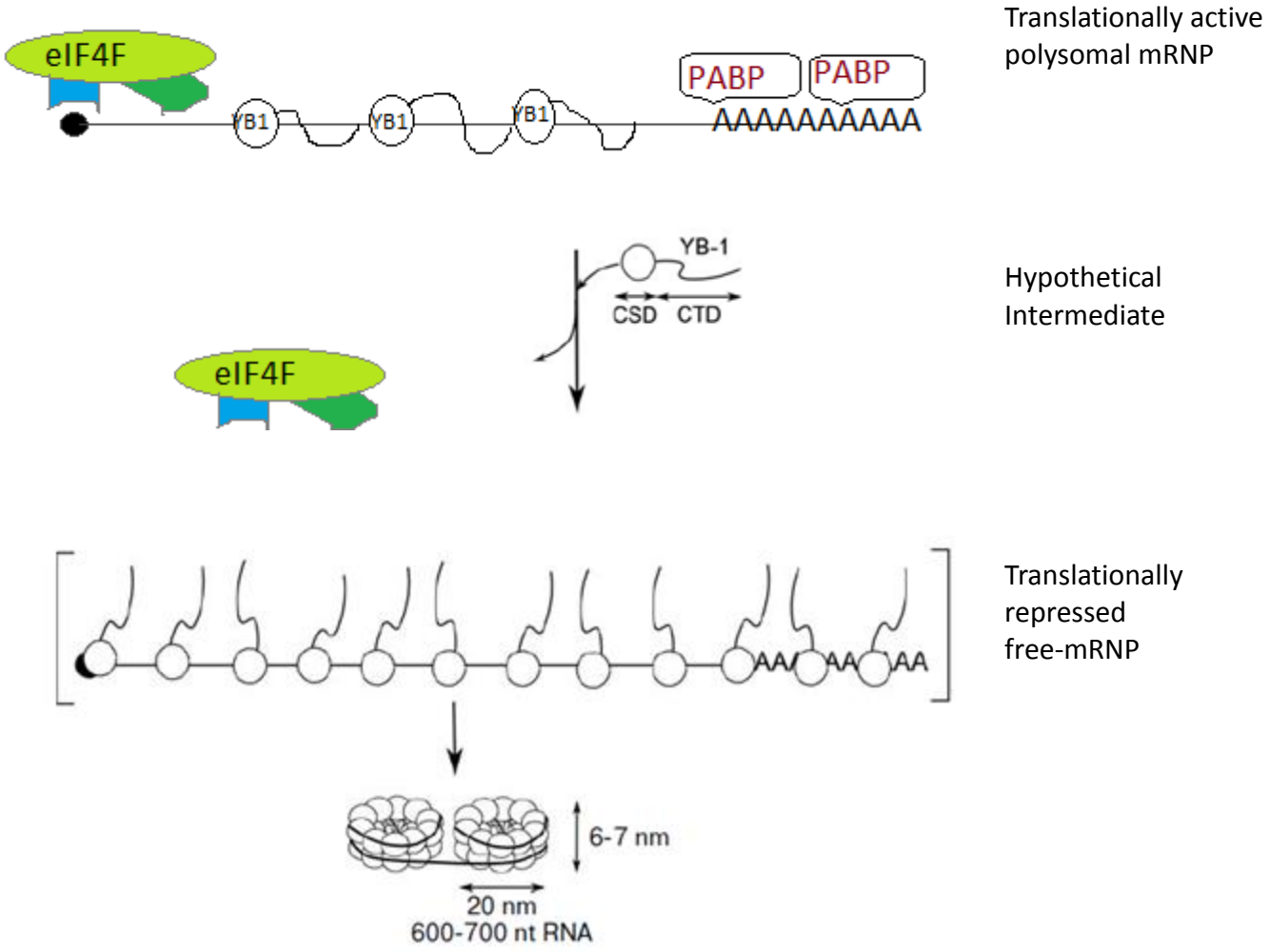


Figure 1.4 Model for translational activation and repression of mRNA translation by YB-1. At low YB-1 to mRNA ratio (top) the cold shock domain binds the mRNA, C-termuinal domain is unbound, and the mRNP is in a open configuration that is accessible to the translational apparatus. At higher YB-1 to mRNA ratios, bottom, the cold shock domain binds mRNA and C-terminal domains bind though charged zippers configuring mRNP in a compact configuration that is inaccessible to the translational apparatus. The diagram is adapted from Eliseeva et al. (2011).

1.7 Translational regulation of the sperm mitochondria-associated cysteine-rich protein (SMCP) mRNA

The sperm mitochondria-associated cysteine-rich protein (SMCP) is a constituent of the sperm mitochondrial capsule and functions in enhancing sperm motility (Nayernia *et al.*, 2002). SMCP is a relatively small (~20 kDa) structural protein with high content of cysteine and proline and is found in the keratinous outer membranes of sperm mitochondria of the mouse, rat and bull (Pallini et al. 1979; Cataldo et al., 1996). Translation of the *Smcp* mRNA is developmentally regulated in mouse spermatids. The *Smcp* mRNA is first detected by *in situ* hybridization in early spermatids at step 3, but the protein is not detected by immunocytochemistry until 6 days later at step 11 (Shih and Kleene, 1991; Cataldo *et al*, 1996). Sucrose and Nycodenz gradient analysis demonstrate that the *Smcp* mRNA is translationally repressed in free-mRNPs in early spermatids in the testes prepubertal mice, and that the mRNA shows a bimodal distribution in the testes of adult mice, with ~65% of the message present as free mRNPs and ~35% present as polysomal mRNA (Hawthorne et al., 2006; Bagarova et al., 2010).

Our lab has been studying the translational regulation of the *Smcp* mRNA for many years now. The *Smcp* mRNA is attractive to us for studies of translational regulation because it lacks the evolutionary relationship to the better studied mRNAs in the protamine-transition protein family. Thus, studies of the *Smcp* mRNA address the question whether translation of all mRNAs in spermatids is regulated by a single set of mechanisms.

Alignments of *Smcp* mRNAs from 8 species in three orders of mammals reveal high degree of conservation in both the 5'UTR and 3'UTR (Hawthorne *et al*., 2006). The conserved elements of the 5'UTR include two upstream open reading frames (uORFs) with AUGs in a strong context as defined by Kozak (1991), a 16 nt long conserved region at positions 7-22 and another 5 nt region immediately upstream of the *Smcp* translation initiation codon. As mentioned previously, uORFs are well known repressors of translation of downstream ORFs (Sachs and Geballe, 2006). The conserved features of the *Smcp* 3'UTR include two AAUAAA polyadenylation signals separated by GAGC, a 33 nucleotide long conserved sequence located directly upstream from the first poly(A) signal, and several, short conserved elements further upstream (Hawthorne *et al*, 2006).

To identify elements that regulate translation of the *Smcp* mRNA, our lab studied the translational regulation of a series of transgenes, all of which contain the *Smcp* promoter and green fluorescent protein (*Gfp*) coding region. The variables include the 5' and 3' UTRs derived from the *Smcp* and *Gfp* mRNAs and mutations in predicted regulatory elements. These transgenes are referred by the notation, $G^5G^CG^3$ or $S^5G^CG^3$, in which the exponents refer to the 5' UTR, the coding region and the 3' UTR and the capital letters refer to the source of the UTR, *Gfp* or *Smcp*. The stage of first detection of GFP fluorescence was determined by microscopic observation of squashes of living seminiferous tubules using phase contrast microscopy to identify developmental steps of spermatids, and fluorescence microscopy to determine cells that express GFP (Bagarova et al., 2010). The results of these studies are summarized in Table 1.1.

A transgene bearing the *Gfp* 5' UTR and 3' UTR is translated at the same stage, step 3, that the natural *Smcp* is transcribed, suggesting that the mRNA is translated without delay (Hawthorne et al., 2006; Bagarova et al. 2010). A transgene bearing the *Smcp* 5'UTR and *Gfp* 3'UTR is translated at step 5, corresponding to a delay of about 24 hours and reduced levels of polysomal mRNA in 21 day old prepubertal mice in which the most advanced cells are step 4 spermatids. Transgenes in which the three AUG codons in the two upstream reading frames ($S^5G^CG^3$ no-uORF1&2) or the AUG codon of the first uORF are mutated to AGG ($S^5G^CG^3$no-uORF1) are translated at step 3, the same stage as a transgene containing both *Gfp* UTRs, $G^5G^CG^3$ (Hawthorne et al., 2006a). The early detection of GFP fluorescence is accompanied by de-repressed levels of polysomal mRNA in sucrose and Nycodenz gradients. These findings suggest that translational repression by the *Smcp* 5'UTR is completely dependent on the AUG codon in the first uORF.

A transgene bearing the *Gfp* 5'UTR and *Smcp* 3'UTR, $G^5G^CS^3$, is translated in step 9 after a delay of about 4 days (Hawthorne et al., 2006; Bagarova et al. 2010). A transgene bearing mutations within the 33 nt conserved sequence just upstream of the poly(A) signal in the *Smcp* 3'UTR, $G^5G^CS^3$-mut, is translated in step 4 – 6 with low polysomal loading in 21 day old testis (Bagarova et al., 2010). Thus, mutations within the 33 nt conserved region only partially relieves repression, implying that there are additional regions within the 3'UTR that contribute towards repression. It is important to mention that the 33 nt region harbors a YRS which was also mutated in the $G^5G^CS^3$-mut transgene.

Previous studies of the mechanisms of translational regulation of the *Smcp* mRNA in transgenic mice have revealed indications that regulation of the natural mRNA involves interactions between the *Smcp* 5' UTR, 3' UTR and coding region (Hawthorne et al., 2006; Bagarova et al., 2010). First, in the presence of the *Smcp* 5' UTR alone GFP can be first detected in step 5, while in the presence of the *Smcp* 3'UTR alone GFP can be first detected in step 9. In contrast, the SMCP protein is first detected by immunocytochemistry in step 11 (Cataldo et al., 1996). Therefore, a question exists whether the 5'UTR and the 3'UTR interact producing a delay of translation of the *Smcp* mRNA until step 11. Second, the $S^5G^CG^3$ mRNA exhibits low polysomal loading in adult testis while the *Smcp* mRNA exhibits higher polysomal loading. This observation suggests that the *Smcp* 3' UTR exerts positive control that neutralizes the negative control by the *Smcp* 5'UTR, similar to a report that the *Her2* 3' UTR neutralizes repression by the uORF in the *Her2* 5' UTR (Mehta et al., 2006).

		Polysomal loading (%)[a]			
		21 dpp testis		Adult testis	
mRNA[b]	First[c]	Nycodenz	Sucrose	Nycodenz	Sucrose
Smcp	st 11	4.2±2.7 (5)	4.9±2.9 (6)	33.0±2.6 (4)	35.1±4.5 (14)[d]
$G^3G^CG^3$	st 3	ND	ND	ND	32.0[e] (1)
$G^3G^CS^3$	st 9	3.4±0.9 (3)	55.3±7.9 (4)	23.7 (1)	37.3±4.3 (3)[d]
$G^3G^CS^3$-mut	st 4–6	10.4±4.8 (5)	9.0±0.91 (2)	29.8 (1)	48.4±7.6 (3)
$S^3G^CG^3$	st 5	7.8±0.1 (2)	16.7 (1)	14.5±1.5 (5)	21.8±4.3 (4)[d]
$S^3G^CG^3$-no-uORF1&2	st 3	28.9±0.8 (3)	32.8±3.7 (2)	26.2 (1)	40.7±6.0 (7)
$S^3G^CG^3$-no-uORF1	st 4	29.8±2.9 (2)	35.1 (1)	33.9±9.0 (2)	41.9±1.4 (2)
Ldhc	PS	34.6±2.9 (8)	52.7±8.5 (9)	32.6±3.7 (3)	56.8±2.4 (10)

Table 1.1 Quantification of polysomal-loading of various mRNAs by sucrose and Nycodenz gradient analysis of 21 day and adult testis. (this table is adopted from Bagarova et al. 2010)

[a] The polysomal-loading of various mRNA was quantified by Nycodenz and sucrose gradient analysis using phosphorimage of slot/northern blots and quantitative reverse transcriptase real time PCR. In general, the polysomal loading of the transgenic *Gfp* mRNAs was quantified with RTqPCR in 21 day testes, while the *Gfp* mRNAs in adult testis and the *Smcp* and *LdhC* mRNAs in 21 day and adult testis were quantified with both RTqPCR and phosphorimaging northern blots. The polysomal loading (%) is presented as mean and standard deviation with the number of independent gradients in parentheses.
[b] mRNA species. [C] The step of spermatid that GFP or SMCP expression was first detected. [d] The LDHC protein is first detected in mid-pachytene spermatocytes.

1.8 Objectives

My thesis research focuses on two topics: first, the role of Y-box proteins as mRNA specific translational repressors and second, the analysis of cis-elements that regulate translation of the *Sperm mitochondria-associated cysteine-rich protein (Smcp)* mRNA in mouse spermatogenesis. In this section, I provide an overview of my research.

As mentioned previously, Y-box proteins are mRNA specific translational repressors. They bind ssRNA specifically through the cold shock domain (CSD) and non-specifically through the basic aromatic islands of the C-terminal domain. Bacterial CSD prefer pyrimidine rich-ligands 7-9 nt long (Max et al., 2006) (Mayr et al., 2012; Sachs et

al., 2012). Bouvet et al. (1995) showed that the Y-box proteins in *Xenopus laevis* oocyte, FRGY1 and FRGY2, bind a hexameric sequence - 5'AACAUC3', which they named the Y-box recognition sequence (YRS). In mouse spermatids, Giorgini et al. (2001) showed that YBX2 and YBX3L bind a 7 nt sequence, 5'UCCAUCA3', within the *Prm1* 3'UTR. Further, they mutated every nucleotide in UCCAUCA in RNA-EMSA and Yeast-3-hybrid assays and defined the YRS to be 5' (U/C/A)(C/A)CA(U/C)C(A/U/C)3'. The nucleotides in parenthesis are alternative bases that exhibit maximal binding. However, this work is incomplete because it did not analyze the 29 elements that differ from UCCAUCA at 2-4 sites. I have analyzed all the sequences predicted by Giorgini et al. (2001) to bind Y-box proteins in modified RNA-EMSA. The following chapter, (Chapter 2) which appears in Journal of Andrology 33 (2012), discusses the findings of my analysis of the degenerate YRS and how that information can be used to identify YRS through comparative genomics. Chapter 2 also analyzes the 5' and 3'ends of 12 mRNAs that undergo translational repression in spermatids for translational control elements.

In Chapter 3, I show that YBX2 binds an element in *Prm1* 3'UTR which is necessary and sufficient for translational repression in round spermatids. As mentioned previously, *Prm1* mRNA is transcribed in step 7 spermatids but repressed in free mRNPs until step 10 spermatids. Translational repression of the *Prm1* mRNA is controlled by the 3'UTR (Braun et al., 1989). Mutational studies in transgenic mice show that *Prm1* mRNA is repressed in round spermatids by 2 copies of a cis-element, the translational control element (TCE). However, deletion of the upstream TCE, GAAAAAUGCCACCGUC, doesn't derepress *Prm1* mRNA; therefore, the distal TCE,

GAACAAUGCCACCUGUC, must be more important for repression (Braun, 1990). The distal TCE contains an element, GCCACCU, which is predicted to bind Y-box proteins in yeast three hydrid assay although Y-box proteins don't bind the element in RNA EMSA (Fajardo et al., 1994; Giorgini et al., 2001). The discrepancy probably resulted from the use of RNase T1 in RNA- EMSA analysis which cleaves RNA at G, because rabbit YBX dramatically increases RNA degradation by RNase T1 (Evdokimova et al. 1995) . I have shown through modified RNA-EMSA that YBX2 binds GCCACCU. I have also used RNA affinity chromatography and mass spec sequencing to show that YBX2 binds the distal *Prm1* TCE *in vitro*.

In Chapter 4, I elucidate the mechanism of translational control of the *Smcp* mRNA. *Smcp* mRNA is transcribed in step 3 spermatid and stored as free-mRNPs until step 11(Cataldo et al., 1996). Previous work has shown that the translational repression of the *Smcp* mRNA is accomplished by multiple mechanisms involving both the 5' and the 3'UTR (Hawthorne et al., 2006; Bagarova et al., 2010). My research shows that the *Smcp* mRNA requires interaction between the 5'UTR and the 3'UTR for proper translational regulation. I have identified a functional YRS in the *Smcp* 3'UTR that binds YBX2 and YBX3L. I also show that the distal end of the *Smcp* 3'UTR harbors an element that binds YBX2.

CHAPTER 2

IDENTIFICATION OF POTENTIAL REGULATORY ELEMENTS IN THE 5' AND 3' UTRS OF 12 TRANSLATIONALLY REGULATED MRNAS IN MAMMALIAN SPERMATIDS BY COMPARATIVE GENOMICS

Tamjid A. Chowdhury and Kenneth C. Kleene*

*Corresponding author

Kenneth.kleene@umb.edu

Department of Biology

University of Massachusetts Boston

100 Morrissey Blvd.

Boston, MA 02125

This chapter appears in appears in Journal of Andrology 33 (2): 244-56 -

http://onlinelibrary.wiley.com/doi/10.2164/jandrol.110.012492/full

2.1 Abstract

To facilitate identifying translational control elements by studies of mutations in transgenic mice, a database of orthologous 5' and 3' ends of 12 mRNA species from 13-23 mammals that undergo delayed translational activation in spermatids was constructed for the *Acev2*, *Akap3*, *Akap4v2*, *Gapdhs*, *Odf1*, *Prm1*, *Prm2*, *Prm3*, *Smcp*, *Spata18*, *Tnp1* and *Tnp2* mRNAs. This database, available here, was searched for conserved sequences in conserved positions and known translational control elements. Numerous potential mRNA-specific elements were identified, including upstream open reading frames, conserved sequences upstream and downstream of the poly(A) signal, and non-canonical and multiple poly(A) signals. RNA electrophoresis mobility shift assays demonstrate that Y-box proteins bind 30 of the 36 permutations of the degenerate Y-box recognition sequence (YRS), [UAC][CA]CA[UC]C[ACU], and this information was used to identify hundreds of YRSs in the UTR database. Collectively, these findings suggest that the distal ends of both UTRs are particularly well-conserved, implying that translation of each mRNA is regulated by mechanisms involving the poly(A) binding protein and the closed-loop. In addition, the 5' flanking regions of all 12 genes have conserved, gene-specific sequences and configurations of elements that resemble the binding site of the testis-specific isoform of cyclic AMP response element modulator, and all 12 genes lack retrogene paralogues demonstrating the efficacy of mechanisms that limit the proliferation of retroposons in the male germ line. This study illustrates the power of comparative genomics in identifying novel hypothetical regulatory elements for analysis with biochemical and *in vivo* genetic approaches.

2.2 Introduction

Developmental delays in activation of mRNA translation are especially prevalent in mammalian haploid spermatogenic cells, spermatids, because chromatin remodelling causes the cessation of transcription before the end of the two week long haploid phase. Since late spermatids synthesize many new proteins during the final stages of sperm differentiation, a large number of mRNAs are transcribed in early spermatids, and stored in an inactive state before translation is activated at specific stages in late spermatids (Kleene, 2003). Premature translation of the *Prm1* and *Tnp2* mRNAs in transgenic mice produces deformed sperm and decreased male fertility, demonstrating the importance of delayed translation (Lee et al., 1995; Tseden et al., 2007).

mRNA species that undergo developmental delays in translational activation in spermatids have been mainly documented in rats and mice, with a smaller number of mRNA species in human (Dadoune, 2003). Nevertheless, developmental regulation of mRNA translation in spermatids is thought to occur in most, if not all, mammals, based on evidence that the replacement of histones by protamines is universal, and results in the cessation of transcription before the end of spermiogenesis (Balhorn, 2007).

Elucidating the mechanisms that regulate translation in spermatids requires mutations in regulatory elements to establish that factors target specific mRNAs directly and to assess the magnitude of the effect of the element on translation. Transgenic approaches are essential because the patterns of gene expression in spermatogenic cells including constituents of the translational apparatus differ dramatically from those in somatic cells (Kleene, 2003), and a cell culture system that supports the differentiation of spermatids has not been developed. Progress has been slow because transgenic

approaches are laborious, and after 20 years of research, only four elements that regulate the timing of translation in spermatids have been identified by analyses of mutations in transgenic mice. Two elements are in the *Prm1* 3' UTR: a highly conserved translational control element (TCE) and a Y-box protein recognition sequence (YRS) (Giorgini et al., 2001, 2002; Zhong et al., 2001). In addition, translation of the *Smcp* mRNA is delayed by upstream open reading frames (uORFs) in the 5' UTR and a YRS ~36 nt upstream of the poly(A) signal (Bagarova et al., 2010). Significantly, the *Prm1* TCE, *Prm1* YRS and *Smcp* YRS delay translation in positions close to the poly(A) signal.

Comparative genomics of the complete or partial sequences of ~30 genomes from 16 orders of eutherian mammals may expedite the process of identifying candidate sequences for analysis of mutations in transgenic mice. A previous study identified conserved sequences in the *Odf1, Prm1, Prm2, Tnp1* and *Tnp2* mRNAs (Kleene and Bagarova, 2008). However, the fact that four of these mRNAs belong to the protamine-transition protein gene family leaves the question unanswered whether the same sequences are present in unrelated mRNAs. The present study constructs a database of the orthologous 5' and 3' ends of seven additional mRNA species in mammalian species that undergo delayed translational activation in spermatids: the *Acev2, Akap3, Akap4v2, Gapdhs, Prm3, Smcp,* and *Spata18* mRNAs. Orthologous genes carry out similar functions in different mammalian species, and are therefore under similar selective pressures (Koonin, 2005). The name of each mRNA, the cellular localization of each protein and the stages at which each mRNA and protein are first detected in mice and rats are compiled in Table 2.1. These mRNAs are referred to here by NCBI nomenclature to avoid confusing alternative names.

This database, including the five mRNAs studied earlier, was searched for known translational control elements such as YRSs, upstream reading frames (uORFs), the *Prm1* TCE, and potential new regulatory elements, focusing on elements that exhibit both highly conserved sequences and positions. The results are presented as highlighted CLUSTAL alignments which depict the position of each element, a format that is easily visualized by molecular biologists.

Table 2.1 mRNAs undergoing developmental delays in translation in mammalian spermatids

Name of mRNA/Species of Mammal/ Location of Protein[1]	mRNA[2]	Protein[3]	Sucrose Gradient[4]
Acev2 (M) angiotensin converting enzyme, variant 2, membrane in mid-piece	PS	step 10	ND
Akap3 (M) A kinase 3 activating protein, fibrous sheath	~step 1	step 4	ND
Akap4v2 (M) A kinase 4 activating protein variant 2, fibrous sheath	~step 1	step 14	ND
Gapdhs (M) glyceraldehyde 3' phosphate dehydrogenase, spermatid-specific, fibrous sheath	step 4	step 12-13	ND
Odf1 (R) outer dense fiber 1, outer dense fiber	step 6	step 15	Yes
Prm1 (M) protamine 1, basic chromosomal	step 7	step 10	Yes
Prm2 (M) protamine 1, basic chromosomal	step 7	step 13	Yes
Prm3 (M) protamine 3, acidic, cytoplasmic	~step 3	step 9	ND
Smcp (M) sperm mitochondria associated cysteine-rich protein, mitochondrial capsule	step 3	step 11	Yes
Spata18 (R) gap between the flagellar axoneme and outer dense fibers	step 7	step 15	ND
Tnp1 (M) transition protein 1, basic chromosomal	step 7	step 11	Yes
Tnp2 (M) transition protein 2, basic chromosomal	step 7	step 10	Yes

[1]Name of mRNA, function and/or cellular location of protein, and species of mammal in which the stages of expression have been studied (M, mouse; R, rat).
[2]Step of spermiogenesis in which the mRNA is first detected, normally by *in situ* hybridization. The approximate stages of first detection of the *Akap3* and *Prm3* mRNAs were determined by northern blot analysis of RNAs extracted from testes of staged pr2010ertal mice.
[3]Step of spermiogenesis in which the protein is first detected by immunocytochemistry.
[4]Sucrose gradient analysis demonstrating that the mRNA in primarily in free-mRNPs in early spermatids and associated with polysomes in late spermatids. ND, not determined.
The references for the cellular location each protein, the stage of detection of mRNAs and proteins, and sucrose gradient analyses are as follows: *Acev2* (Howard et al., 1990; Métayer et al., 2002; Langford et al., 1993); *Akap3* (Brown et al., 2003); *Akap4v2* (Brown et al., 2003); Gapdhs (Bunch et al.,1998; Welch et al., 1992, 1995); *Odf1* (Morales et al., 1994; Burmester and Hoyer-Fender, 1996); *Prm1* (Mali et al., 1989; Kleene, 1989, Yan et al., 2003); *Prm2* (Mali et al., 1989; Kleene, 1989, Yan et al., 2003); *Prm3* (Grzmil et al., 2008); *Smcp* (Kleene, 1989; Shih and Kleene, 1992; Cataldo et al., 1996; Hawthorne et al., 2008), *Spata18* (Iida et al., 2004, 2006), *Tnp1* (Mali et al., 1989; Kleene, 1989; Yan et al., 2003) and *Tnp2* (Kleene, 1989; Shih and Kleene, 1992; Yan et al., 2003).

2.3 Materials and methods

2.3.1 Strategy for the identification and analysis of orthologous 5' and 3' ends of genes

A database containing the 5' and 3' ends of orthologous genes encoding mRNAs that are translationally regulated in spermatids was constructed in several stages. (1) Translationally regulated mRNAs were identified from the literature, and conserved sequences near the ends of the proteins were identified in CLUSTALW alignments. (2) Short conserved sequences at or near the amino- and carboxy-termini were used as TBLASTN queries to identify candidate genes in various mammals in releases 48-50 at ENSEMBL as described earlier (Kleene and Bagarova, 2008). The filter was turned off for low complexity queries. Orthologues were initially identified by E-values and the position of start and stop codons in the corresponding genomic sequence. Hits lacking start or stop codons at the normally conserved positions always lacked all other conserved features of orthologues. This strategy identifies sequences adjacent to the UTRs and eliminates the poor selectivity of low complexity queries. (3) The positions of introns and exons were deduced from conserved splice sites, often correctly predicted by ENSEMBL. (4) The structure of each gene and the presence of the 5' end and 3' end in the same piece of DNA were confirmed in two ways: the nucleotide sequence of the 5' or 3' end of the each gene was used to search the corresponding mammalian genome at ENSEMBL with BLAT, and the coding exons in a segment of DNA containing the entire gene (10-60,0000 nt) were identified with the corresponding full length mouse protein with NCBI TBLASTN using the BLAST2 option. Similarly, the segments of DNA identified by nucleotide queries from one end were searched with queries from the other

end with BLASTN. The genes identified here were usually complete in species in which the genomes have been extensively sequenced, although complete genes were often identified in incomplete genomes. (5) Testis ESTs in various species were searched with NCBI BLASTN (word size 7 and low complexity filter off) with queries spanning the expected 5' and 3' termini to determine the positions of putative transcription start and poly(A) sites and to confirm the positions of splice sites. (6) The inferences of orthology were based on conserved protein sequences (usually supported by E-values), intron positions, and conserved sequences in the 5' and 3'UTRs, and 5' flanking regions. (7) The 5'UTRs and 5' flanking regions and 3'UTRs and 3' flanking regions were used as queries with BLAT to determine the location of each sequence in scaffolds and chromosomes in ENSEMBL Release 52. (8) The 5' and 3'UTRs were searched for the potential regulatory features discussed in the Results using DNA Pattern Find (www.bioinformatics.org/sms2/dna_pattern.html) and FASTA35, downloaded from William Pearson's home page. MFOLD was used at Michael Zucker's homepage.

The procedures used to identify the 5' and 3' ends of the *Acev2*, *Akap3*, *Akap4v2*, *Gapdhs*, *Smcp* and *Spata18* genes are described in the Supplemental Methods.

2.3.2 RNA electrophoresis shift assays (RNA-EMSA)

RNA-EMSAs were carried out as described previously with minor modifications (Fajardo et al., 1994; Bagarova et al., 2010). Plus and minus strand oligonucleotides encoding the first 37 nucleotides of the *Prm1* 3' UTR were purchased from Invitrogen. The sequence of the plus strand follows:

ACGTACGTACGAATTCTAGATGCACAGAATAGCAAGTCCATCAAAACTCCTG CAAGCTTAATTGGCCTGG. In the various probes, the *Prm1* 3'UTR YRS

(underlined) was replaced by permutations of the degenerate YRS ([TAC][CA]CA[TC]C[ACT]) that differ at 2-4 sites from TCCATCA (Giorgini et al. 2001). The oligos were annealed, digested with *Eco* RI and *Hind* III, ligated into the *Eco*RI and *Hind* III sites of pGEM3 (Promega-Biotec). The sequence of the insert was verified, the plasmid was linearized with *Hind* III and probes were synthesized with the T7 bacteriophage RNA polymerase (New England Biolabs) and α-[^{32}P]-rUTP (Perkin Elmer). After removal of unincorporated nucleotides on a BioGel P6 (Bio-Rad) column, the cpm of each probe was determined by scintillation counting, and 10^5 cpm was used in each reaction. Sequence-specific complexes were formed by incubation with cytoplasmic extracts of adult testis, *E. coli* tRNA, RNase T1 and heparin, and the sizes of the complexes were determined by UV-crosslinking and SDS-PAGE on a 3 cm 5% stacking and 20 cm 10% separating 30:1 acrylamide:bis-acrylamide gel.

2.3.3 3' Rapid amplication of cDNA ends (3'RACE)

The polyadenylation sites of several mRNAs were determined using a 3'RACE kit (Invitrogen). 2-5 μg of total adult testis RNA purified with the Trizol reagent (Invitrogen) was reversed transcribed with an oligo dT primer, GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTT. The 3' ends of the following mRNAs were PCR-amplified with *Taq* polymerase with GGCCACGCGTCGACTAGTAC as the reverse primer and a forward primer specific for each mRNA: *Ace*v2, CAGCAGAGGAGTGTCCCATA; Akap3, CACTGCCGTGGAGAAAGG; *Smcp*, GGTCAGGCTAAGACTATGTTGTA; and *Spata18*, TGTCCTCGTAATCATTTTGGAA. The PCR products were cloned into pCR2.1 (Invitrogen), and at least 10 independent 5'RACE inserts were sequenced.

Male CD-1 mice were purchased from Charles River Laboratories, Wilmington, MA, and maintained in the UMass Boston Animal Care Facility. Procedures for the use of adult male mice in the preparation of the cytoplasmic extract and total testis RNA have been reviewed and approved by UMass Boston Institutional Animal Care and Use Committee.

2.4 Results

2.4.1 Identification of orthologous translationally regulated mRNAs

Although the number of mRNAs that exhibit temporal delays in translational activation in spermatids is probably large, many mRNAs are not suitable for comparative genomics. Desirable features include: (1) Well-documented translational regulation, preferably including precise determination of the stage of first detection of the mRNA and protein by *in situ* hybridization and immunocytochemistry, respectively, and sucrose gradient analysis demonstrating negligible levels of polysomal mRNA in early spermatids and substantial levels of polysomal mRNA in late spermatids. (2) A single-copy gene encoding a distinctive, conserved protein to facilitate identifying orthologues and minimize the possibility of confusing orthologues and paralogues. (3) Identification of the transcription start site by primer extension and/or S1 nuclease protection to determine the exact 5' terminus of the 5'UTR, a segment that is important in translational regulation. (4) A compact gene without large introns in the 5'UTR. (5) An abundant mRNA that is expressed as a single structure only in spermatids. Abundant mRNAs will also usually be represented by many expressed sequence tags (ESTs), which are useful in determining the structure of the UTRs, while the alternative transcripts of a single gene

may exhibit different patterns of translational activity. The mRNAs studied here meet these criteria to varying degrees; deficiencies that affect interpretation are noted below.

All of the genes were initially identified with TBLASTN searches of the ENSEMBL genomic sequence database using amino and/or carboxy-terminal queries. The similarities in the N-terminal and C-terminal segments of the ACEV2, AKAP3, AKAP4, GAPDHS, SMCP and SPATA18 orthologues are demonstrated by CLUSTAL alignments in Supplemental figure 1. The sequences of the complete ODF1, PRM1, PRM2, PRM3, TNP1 and TNP2 proteins have been reported previously (Kleene and Bagarova, 2008; Grzmil et al., 2008). A database of orthologous 5' and 3' ends including the complete UTRs and short flanking sequences was constructed, Supplemental figure 3, by deleting introns which are only present in the *Akap3*, *Akap4v2* and *Smcp* 5' UTRs. Supplemental figure 2 presents CLUSTAL alignments of the 5' and 3' ends of orthologous genes showing the positions of putative regulatory elements identified below. Figure 2.1 illustrates the position of these elements in representative mouse or human 5' and 3' UTRs.

The identification of orthologous genes is based on similarities of the amino- and carboxy-terminal peptides or the entire protein, the positions of splice sites and start and stop codons, conserved sequences in the 5' flanking region, 5'UTR and 3'UTR and the absence of paralogues. As reported earlier (Kleene and Bagarova, 2008; Grzmil et al., 2008), the *Odf1*, *Prm1*, *Prm2, Prm3, Smcp*, *Tnp1* and *Tnp2* genes encode proteins with distinctive, conserved features. Supplemental Figure 1 shows CLUSTAL alignments of the amino acid sequences that identify the 5' and 3' ends of the *Acev2*, *Akap3*, *Akap4v2*, *Gapdhs*, *Smcp* and *Spata18* genes. The *Akap3*, *Akap4*, and *Gapdhs* genes are members of

gene families, but orthologues can be distinguished from paralogues by spermatogenic cell-specific amino acid sequences, and conserved gene-specific sequences in the 5' flanking region, 5' UTR and 3' UTR. In addition, all of the *Gapdhs* 3' UTRs are in a syntenic location, ~500 nt upstream of the transmembrane protein 140 gene.

Mouse *Acev2* 5'UTR
TCTGCTTTCCTGCGGCCATG

Human *Akap3* 5'UTR
GATATCACCAATTGTGATTCTGATGTCACTCTCCAGTCCCAGCAAGGGGGAG
GTACATGGAAGGCCACAGGAAGAAACAAGATCTTGAGCTGAGCAAGAACAT
CCCAGCATCTTCATTGACTTTAAAAGTATATTCTGGAGTCTTCCGTGGTTCAC
TATTCCAGTACTACAGAGATTCCTTATATTACATGGCAGGAGGGGGGTAAAC
TGAGGGATAGTGAAGACAACAATAAATTAATCAAGAGCTTTCCTCATATCTC
AGAACCTATCCTCTGTAAGAATG

Mouse *Akap4* 5'UTR
ATCAGTCTGGTCTAACAGCTGACCGGGGTGGCAGCCAGCTGCAAGTGCCTAA
GAACTTGGCACTGCCCCCTTCCATCTAAAGGGGCACATCTCACTTCTGGGTGA
CACACACTCAGTCAAAGGTACAAAACAACTCTATCATCAAGATG

Mouse *Gapdhs* 5'UTR
ACACCTCAGTAACACCACGGAGGGGGGCCAAGGCAGCCAGGCCATGAGATC
TTAGGCCATG

Mouse *Odf1* 5'UTR
TTTTAAAGGAGGCCTCTGAGAAGAGCTTAGAACAATTTTTTCCTCTGAGTGCC
ATTTCCCAAAGGTACTCACAGAACAATAAGGTGTGACCATAATG

Mouse *Prm1* 5' UTR
ACAGCCCACAAAATTCCACCTGCTCACAGGTTGGCTGGCTCGACCCAGGTGG
TGTCCCCTGCTCTGAGCCAGCTCCCGGCCAAGCCAGCACCATG

Mouse *Prm2* 5'UTR
ATCATCACCACCAAGAGCAGGTGGGCAGGCTTTCGTCCCTCCTCCTCCAATCC
AGGTCAGCTGCAGCCTCAATCCAGAACCTCCTGATCTCCTGGCACATG

Mouse *Smcp* 5'UTR
GTCAGAAGACTTTGACTTCTGATAGCCATGGACTCACTAGACTGCTGAGGAA
GACCCAGCATCTATTCAATCTGCTGAAACATCCAGGAAACTACTTTTAACACC
GAGAATCAAGTATGGAAATGCTGAACTAAGAAGAGCCCAAGGAAGAACTGT
GTTGCCAGATCAGGAACTCCAACTCTAAAGAAGATG

Human *Spata18* 5'UTR
GTTTAATAATCGCCAGGGTATCTATGGCCGGGCTCAGGCGGCTGCTGGGGAG
CCAGGAGACCGCGCGGGACGGCGGATGAGGCGCGGCGGCTGCGGCCCAGGG
CACCTCCCCTCTGGCTTCCCGAACCCGGCCAGGTCCGACCCGAGGGGGAGGA
TGGAAACACCTGCCGCGCTCTGAGCCCCCCAGAAGAGAACACCCTTCCCGCC
ATATCACCCCACGGTCCTGCGGAGGCCACCGCCTGGTCCCCCCAAGTCTCCAT
CGCGCAGCGTGGGGCCGAGAGGAATAGTGAGCGATG

Mouse *Tnp1* 5'UTR
GCAAAGCCCCTCATTTCGGCAGAAAGTACCATG

Mouse *Tnp2* 5'UTR
AAAGCCGGGCCTGCTGGGAGGAGGAGGAGGAGGAAGTCTCTGCCCCGAGTG
TGGCCTCCCATG

Mouse *Acev2* 3'UTR
TGAGGTGACCCTGCTGGCAAGGCCAGCAGAGGAGTGTCCCATAAGAAACTGG
ATGGGGGACATGGTGTTTGAGTGGAACATACCCTAGCTGAGCCTTCCTCCTTT
GCTGTCCCCATCCACTCTGCCCCACCCCCACCCCCCAGGCCCAGCCCCCATTC
TCTGGATACCCAGTGTCTAACACCGACTTCCCTGCCCAGTCTCTGTGAATACA
ATTAAAGGTCCTCCCCCA

Mouse *Akap3* 3'UTR
TGATTGGGGCCTACCCTGAGTTCCCTCAGCGGGCCGAGTCCCCGCCCCCTCAG
CCCCCTCCATGCCCCACAGAGCCCTAAAGTCCCCTCCATGCCACGCACACTA
GACATGCCATCTAACGCTACTCACTGGATTTTGCAGATTTTCTTGTCCATGCG
AGCAAGGACATAAATTAAAAGATTACAGTTAAAGGGCA

Mouse *Akap4* 3'UTR
TAAGCTGAGAATTCCTTTGACTCCCCTCCATCCATCCTCCCCCCCAGCAGCAA
TTCCACCCCAGCTGGAGCCACCCTCACCATCAGGCTGGTGAACTGCACAATT
GGGATCACATTTACCAATACATCTGAGCAGTTGCACTGTGAAAATACTGGGT
GCCCTCCTGGGCAACATGAATAAAAAAATTCA

Mouse *Gapdhs* 3'UTR
TAACACAAAAGGCCCCTCCTTGCTCCCCTGCGCACCTCGCGTTCCTGACTTCG
GCTTCCACTCAAAGGCGCCGCCACCGGGTCAACAATGAAATAAAAACGAGA
ATGCGCAC

Mouse *Odf1* 3'UTR
TAAGGTGTATGTAAGAACTTATGTCTTAGCAGAAGTCAGTACTCCAGCCAGG
CAGCTCTTCAGCATTTCTCGTCCCCTTCCCAGAGCCCGTGTAGTTCCAGTGTG
TAGGAAACTTAATACAGAATTGCATCTGA

Mouse *Prm1* 3'UTR
TAGATGCACAGAATAGCAAGTCCATCAAAACTCCTGCGTGAGAATTTACCAG
ACTTCAAGAGCATCTCGCCACATCTTGAAAAATGCCACCGTCCGATGAAAAA
CAGGAGCCTGCTAAGGAACAATGCCACCTGTCAATAAATGTTGAAAACTCAT
CCC

Mouse *Prm2* 3'UTR
TAAGCCTCCCCAGGCCTGTCCATTCTGCCTGGAGCCAAGGAAGTCACTTGCCC
AAGGAATAGTCACCTGCCCAAGCAACATCATGTGAGGCCACACCACCATTCC
ATGTCGATGTCTGAGCCCTGAGCTGCCAAGGAGCCACGAGATCTGAGTACTG
AGCAAAGCCACCTGCCAAATAAAGCTTGACACGAGATTC

Mouse *Smcp* 3'UTR
TAAACTGTCCCTGACACCATGCCCTTTTTCAAAGGGTATAGGATTACTACAGG
TCAGGCTAAGACTATGTTGTAAAGATGCTGTTTTCACAATAACCAACAAGTCC
ACTCAACCATAAGCTACCATTTCGACCTAACTGTAGGCTACTATTGCAACTGG
AAATGGAAGGTAGAAAAGGATAGAAACATCTTGTCTAGTGATCCTGACATTT
AGATAGCAAAGAAATAAAAGAGCAAATAAAAAGA

Human *Spata18* 3'UTR
TAAAAGCACCAGACCTGCTCCTTTGACCCAGTGCGTGGAAACAGCTGCTTTCT
CCAGTGCCGCCATCTGTCTTCTGTGTCTGCCTCAGACCTCACTTAAGATAATG
TCAAAAGGCAATTCTGTGTATCACCCCACACAGAGAGTTAAATGTTTTGGCTT
GGCGCATTTGTAACTTT

Human *Tnp1* 3'UTR
TGAGCCCCCAGCGGGCTCTGCCCTGGTGCGCTTCACACAGCACCAAGCAGCA
ACAAGAACAGCAGAAGGGGAACTGCCAAGGAGACCTGATGTTAGATCAAAG
CCAGAGAGGAGCCTATGGAATGTGGATCAAATGCCAGTTGTGACGAAATGAG
GAATGTATATGTTGGCTGTTTTTCCCCAACATCTCAATAAAACTTTGAAAGCA
G

Mouse *Tnp2* 3'UTR
TGACGCACTCCAGGATGTTCCTGTGTCCATTTGATCCCAAAATGAGATAGCCA
TCACTAGGGGACTGTTGGGATGATGTCACAGGAACATGTCACTGCAGCAATT
TCTATGCAACATGGATTAAAGCTTGTACCCTGGAAGACT

Figure 2.1 Positions of putative regulatory signals in the 5' UTRs and 3' UTRs of selected human and mouse mRNAs. The human and mouse 5' and 3' UTRs were selected to provide the best general representation of the presence and position of conserved elements in the orthologous UTRs of diverse species. The highlighted elements are as follows: red, the ends of 5'UTRs (transcription start sites and translation start codons, and the ends of 3' UTRs (translation stop codons and poly(A) sites; grey, poly(A signals differing by no more than one base from AAUAAA or AUUAAA; green, uORFs; magenta, *Pm1* TCE elements; khaki, pumilio/CFlm UGUA elements; yellow, YRS. The complete set of 36 YRS are highlighted yellow, and the differences in intensity of complex formation determined by RNA-EMSA here and Giorgini et al. (2001) are symbolized by underlines: ++, double underline; moderate, +, single underline, -/+, dotted underline. The boxed transcription start sites and poly(A) sites were identified by primer extension/S1 nuclease protection and 3' RACE respectively, while the unboxed sites were determined by BLASTN alignments of the 5' and 3' ends with ESTs.

The identification of orthologues is greatly simplified by the observation that none of these genes have spawned retroposons, aka. processed pseudogenes, (Kleene and Bagarova, 2008; Grzmil et al., 2008), a class of genes that is created by synthesizing a reverse transcriptase copy of a mature mRNA (Weiner et al., 1986). The absence of *Gapdhs* retroposons is especially striking because the gene encoding the *Gapdh* mRNA that is expressed in spermatogonia and somatic cells has generated tens to hundreds of retroposons in various species (Gibbs et al., 2004; KC Kleene, data not shown). Two virtually identical copies of several genes or exons were identified, which may represent recent duplications of genomic DNA or assembly errors. All of these genes lack the hallmarks of retroposons: the remnant of the poly(A) tail and the absence of introns and 5' and 3' flanking regions derived from the intron-containing progenitor gene (Weiner et

al., 1986). Of course, the caveat exists that retroposons could not be detected in the unsequenced regions of incompletely sequenced genomes.

2.4.2 Transcription start sites

The transcription start site marks the boundary between the 5' UTR and flanking region. Primer extension and/or S1 nuclease protection identifies discrete transcription start sites in the mouse or rat *Acev2, Gapdhs, Odf1, Prm1, Prm2, Smcp, Tnp1* and *Tnp2* genes (Hawthorne et al., 2006; Howard et al., 1990; Welch et al., 1995; additional references in Kleene and Bagarova, 2008). Supplemental figure 2 also shows that a large number of human *Acev2, Akap4v2, Prm1, Prm2* and *Tnp1* ESTs terminate at virtually the same 5' base, corresponding to start sites that were identified by primer extension analyses of rodent mRNAs, suggesting that transcription starts at a single site, and that the corresponding human cDNA libraries have a high proportion of full-length 5' ends. The 5' termini of the ESTs from other mammals lie upstream and downstream of those determined biochemically, presumably due to well-known difficulties in copying the 5' ends of mRNAs, and inferring start sites from a small number of ESTs.

2.4.3 Gene-specific CRE-sites

We reported previously that the 5' flanking regions of the *Odf1, Prm1, Prm2, Tnp1* and *Tnp2* genes contain gene-specific elements in conserved positions that differ by one to two bases from the consensus sequence of the binding site for the testis-specific isoform of cyclic-AMP response element modulator, CREMτ, TGACGTCA (Delmas and Sassone-Corsi, 1994; Kleene and Bagarova, 2008). The seven new genes analyzed here also consistently exhibit gene-specific CRE-like elements in their 5' flanking regions: *Acev2*, TGAGGTCA; *Akap3*, two elements separated by 13 nt, usually CGATATCA and

TGATGTCA; *Akap4v2*, TGACTCA or TGACCCA; *Gapdhs*, two or three elements, usually separated by 10-nt, ~1/3 match the CRE-consensus and 2/3 differ at one base from the CRE-consensus; *Smcp*, TGAGGTCA or TGGTGTCA; *Prm3*, the CRE-like element (underlined) is fused to a conserved 5' sequence, CTTTGTGATGTCA; and *Spata18*, the CRE-like element is fused to a 3' sequence, CCACGTCAAGGTTTGTT. The conserved positions and sequences of the CRE-like elements support the inference of orthology. The vast majority of CRE-like elements in conserved positions in these genes are not TGACGTCA, ~254/268, implying that the consensus sequence is inaccurate.

2.4.4 Upstream open reading frames (uORFs)

The principal translation start codon begins the coding sequence of the principal protein encoded by that mRNA (Sonenberg and Hinnebusch, 2009). Principal start codons are usually the AUG codons closest to the 5' cap, and are normally in a strong context, defined by a purine in the –3 position (the A of the ATG is designated +1), or an adequate context, defined by a -3 pyrimidine and a +4 G (Kozak, 1991). Most ribosomes initiate at AUG codons in a strong context, about half initiate at AUG codons in an adequate context and few ribosomes initiate at AUG codons in a weak context lacking both a -3 purine and a +4G (Kozak, 1991). The *Tnp2* principal start codon is in an adequate context, while those of the remaining mRNAs are in a strong context.

The *Akap3*, *Gapdhs*, *Smcp* and *Spata18* 5'UTRs contain uORFs located between the transcription start site and the principal start codon (Figure 2.1). Translation of uORFs represses translation of principal ORFs by decreasing the proportion of ribosomes that initiate at downstream AUG codons, and the strength of repression depends on many factors such as the sequence, length and position of the uORF and transacting factors

(Sonenberg and Hinnebusch, 2009). The description of the uORFs below points out features that may affect the strength of repression.

The *Spata18* 5'UTR contains a seven codon uORF, and the *Smcp* 5'UTR usually contains two short uORFs, six and five codons, all of which are headed by AUG codons in a strong context. The *Gapdhs* uORF encodes three amino acids beginning with an AUG codon in a strong context in most mammals, but in higher primates the uORF encodes four amino acids beginning with a AUG codon in a weak context. However, the higher primate *Gapdhs* uORFs may be translated despite their weak contexts, a prerequisite for repression of downstream ORFs (Sonenberg and Hinnebusch, 2009), because MFOLD predicts that the 5' end of the *Gapdhs* coding sequence contains a stem loop with ΔG of –9-14.1 kcal 15 nt downstream from the uORF start codon. A stem loop with similar stability in a similar position enhances initiation at start codons in an adequate context (Kozak, 1991). The *Gapdhs* uORF could impose an especially strong block to *Gapdhs* translation by virtue of its proximity 3-6 nt upstream of the principal initiation codon (Sonenberg and Hinnebusch, 2009). The uORF, which is present in all of the *Gapdhs* 5' UTRs and absent from the 5'UTR of the *Gapdh* mRNA that is expressed in somatic cells, is a consistent difference between these paralogues (not shown).

All of the *Akap3* 5'UTRs contain uORFs, but the number and length of the uORFs is unclear, because the transcription start site is not known precisely. All of the 5' ends of *Akap3* genes have a AUG codon in an adequate context at nt ~62 which begins an ORF of 24 to 78 codons, but initiation at this AUG may be reduced by its proximity to the transcription start site (Kozak, 1991). However, all of the *Akap3* 5' UTRs have one or two additional downstream AUG codons in a strong or adequate context, which would

likely be translated by ribosomes that did not initiate at AUGs closer to the cap. The first uORF in macaque, bull and tree shrew and the second uORF in rabbit, bull, dog, horse, bat and tree shrew terminate downstream of the *Akap3* start codon. Long uORFs and uORFs that overlap the principal start codon can create especially strong blocks to the translation of principal ORFs (Sonenberg and Hinnebush, 2009). The mouse and rat *Akap3* 5' UTRs contain an exon with a uORF that is absent in other species.

2.4.5 Y-box recognition sequences (YRS)

The mouse *Prm1* 3' UTR contains a seven nt YRS, UCCAUCA, that binds Y-box proteins MSY2 and MSY4 (Giorgini et al., 2001). Analysis of the single base mutations at each position reveals that UCCAUCA is one of seven elements containing four degenerate sites that exhibit equivalent protein binding, described by a consensus sequence in which degenerate sites are in brackets, [UAC][CA]CA[UC]C[ACU] (Giorgini et al., 2001). However, that study did not analyze the 29 elements that differ at 2-4 sites from UCCAUCA.

We addressed this question using RNA electrophoretic mobility shift assays (RNA EMSAs) with total adult testis extracts. The probes in these RNA EMSAs contain the first 37 nt of the *Prm1* 3'UTR in which various mutant YRS were substituted for UCCAUCA. Sequence-specific complexes were selected by treatment with *E. coli* tRNA, heparin and RNase T1, and the complexes were UV-crosslinked to form covalent bonds to enable estimation of their molecular weights with SDS-PAGE. Giorgini et al. (2001) demonstrated that the bases surrounding UCCAUCA do not affect binding.

In agreement with UV-cross-linked complexes reported by Fajardo et al. (1994), Figure 2.2 reveals that UCCAUCA and other YRSs form two complexes which contain

proteins of apparent MW 53 kDa and 51 kDa after subtracting the ~5.4 kDa 16 nt minimum RNase T1 fragment. The complexes of different sizes presumably correspond to various Y-box proteins in spermatids: MSY2, MSY2A, MSY4 (aka MSY3L) and MSY3S (Gu et al., 1998; Davies et al., 2000; Mastrangelo and Kleene, 2000). The intensity of the complexes was divided into three categories as summarized in Panel 2B, strong (++), moderate (+) and negligible (-/+). Six of the YRS that formed negligible complexes lacked the base, U, in the α[^{32}P]-rUTP labeled probes, but were rich in C or A and C. However, the 16 nt RNase T1 fragment including the *Prm1* 3'UTR YRS contains two Us, and UCCAUCA, but not the C- and AC-rich YRSs, formed complexes when the probes were labeled with [^{32}P]-rCTP (not shown).

We searched the UTR database for YRSs with the DNA Pattern Find algorithm with queries for all 36 YRS, and annotated the YRS according to strength of complex formation (Figure 2.2). This search identified many potential YRSs in the various UTRs, several of which are strongly conserved and occupy positions that have been implicated in translational control. The *Tnp1* mRNA contains a YRS immediately upstream of the poly(A) signal in 14 of 17 species, AACAUCU, which matches that of the FRGY2 YRS that binds *Xenopus laevis* Y-box protein 2, FRGY2, in SELEX assays (Bouvet et al., 1995), and the YRS in the *Smcp* 3'UTR (Bagarova et al. 2010). All of the *Prm1* 5' UTRs and a subset of *Prm2* and *Prm3* 5'UTRs have YRSs at or slightly downstream from the transcription start site, another position that is often involved in repressing translation (Kozak, 1991; Goossen and Hentze, 1992; Levy et al., 1991).

In addition, three mRNAs have YRSs in positions that may affect the translational regulation by uORFs. The first *Akap3* uORF contains a YRS in 13/14 species, and the

Spata18 uORF contains a YRS in 7/16 species. A YRS in a uORF could prolong translation of the uORF, diminishing reinitiation at downstream start codons (Sonenberg and Hinnebusch,.2009). 14/16 *Spata18* 5' UTRs have one or two putative YRSs between the uORF and the *Spata18* start codon, and 7/17 *Smcp* 5' UTRs have a YRS between the first and second uORF, a position that might block scanning by small ribosomal subunits after translating uORF1.

Since the binding of Y-box proteins is enhanced by two adjacent *Prm1* YRSs (Giorgini et al., 2001), it is notable that closely spaced or overlapping YRSs are present in the all of the *Akap4v2* 5' UTRs and a subset of *Spata18* 5'UTRs and 3' UTRs. Most of the *Akap3, Akap4v2, Prm1, Prm2, Smcp,* and *Spata18* mRNAs have putative YRSs in both UTRs, usually well upstream of the poly(A) signal, which may promote interactions between the 5'UTR and 3' UTR. Finally, the *Odf1* and *Tnp2* 5'UTR and 3'UTR, the *Acev2* and *Tnp1* 5'UTR and the *Gapdhs* 3'UTR are nearly or completely devoid of YRSs.

2.4.6 Gene-specific poly(A) signals

All 12 mRNAs exhibit strongly conserved mRNA-specific poly(A) signals. The *Gapdhs, Prm1, Prm2, Prm3* and *Tnp1* 3'UTRs contain canonical AAUAAA signals (Figure 2.1), while the *Acev2* and *Tnp2* 3'UTRs contain the most common variant, AUUAAA, and the *Odf1* 3'UTR contains a perfectly conserved, atypical variant, AAUACA. The *Spata18* 3'UTR in 12 mammals contains another atypical variant, GUUAAA, but the rat and mouse *Spata18* 3'UTRs lack this signal, and contain two, contiguous copies of another non-canonical signal 28 nt upstream, AGUAAA, that is functional based on 3'RACE described below.

Four mRNAs have conserved multiple, contiguous or nearly contiguous poly(A) signals. The *Akap4* 3'UTR contains conserved multiple poly(A) signals consisting of an upstream canonical AAUAAA signal and a downstream A-rich segment often containing one or two additional signals. The *Acev2* 3'UTR usually contains a AAUACA immediately upstream of a AUUAAA. The 3'UTR of the *Akap3* mRNA contains two poly(A) signals in an A-rich region, but the upstream signal is either AAUAAA or AUUAAA, and the downstream signal is always AUUACA, possibly related to AAUACA. The *Smcp* poly(A) signals are usually separated by GAGC, whereas the *Akap3* and *Akap4* poly(A) signals are separated by A-rich sequences.

The presence of two conserved poly(A) signals raises the question whether both signals are used to specify the position of poly(A) sites, the base to which the poly(A) tail is added (Tian et al., 2005). To answer this question, 3'RACE was used determine the exact position of the poly(A) sites in the *Acev2, Akap3, Smcp* and *Spata18* 3'UTRs in mouse testis. The *Acev2, Akap3,* and *Smcp* 3'UTRs exhibit a strong preference for single poly(A) sites 16-19 nt and 4-11 downstream of the first and second poly(A) signals, respectively. This suggests that only one poly(A) signal is used, presumably the upstream poly(A) signal, because poly(A) sites are usually more than 14 nt downstream from the poly(A) signal (Tian et al. 2005). In contrast, the double *Spata18* poly(A) signals in mouse are used with similar efficiency giving rise to three poly(A) sites.

Careful inspection of the poly(A) signals reveals short conserved sequences associated with canonical poly(A) signals of other mRNAs such as the UGUU immediately following the *Prm1* poly(A) signal, the As surrounding the *Gapdhs* poly(A) signal, and GAGUC which is either immediately upstream or downstream of the *Prm3*

poly(A) signal in different species. The *Odf1* 3' UTR contains two perfectly conserved elements, the non-canonical poly(A) signal, AAUACA, and the element, UGCA, five nt down stream. The fact that these two sequences are the only conserved sequences in the entire guinea pig *Odf1* 3'UTR suggests that they may function in post-transcriptional gene regulation.

In addition, the sequence, UGUA, which binds the polyadenylation factor CFIm and the translational regulator pumilio (Xu et al., 2007; Sartini et al., 2008) is present in conserved positions near the transcription start site and 3 nt downstream of the poly(A) signal of the *Tnp2* mRNA, 26 nt upstream of the *Tnp1* poly(A) signal, and 16 and 19 nt upstream and downstream of the *Spata18* poly(A) signal, and less-well conserved positions in the *Akap4* 3' UTR and *Odf1* 5' UTR. These UGUA elements likely function in polyadenylation because CFIm subunits are expressed at high levels in testis and bind UGUA elements in CHIP assays (Sartini et al. 2008).

Since cytoplasmic polyadenylation is an important mechanism of developmental regulation of mRNA translation in oocytes and spermatocytes (Radford et al. 2008), it might be used by mammalian spermatids. This possibility seems unlikely because the mRNAs studied here lack cytoplasmic polyadenylation elements, UUUUUAU and closely related variants (Radford et al., 2008), close to their polyadenylation signals.

2.4.7 The Prm1 TCE and other conserved elements

The highly conserved 34 nt segment upstream of the mouse *Prm1* 3' poly(A) signal has been divided into two elements, a downstream 17 nt TCE, and an upstream element partially complementary to the TCE, CST (Zhong et al., 2001). Mutation

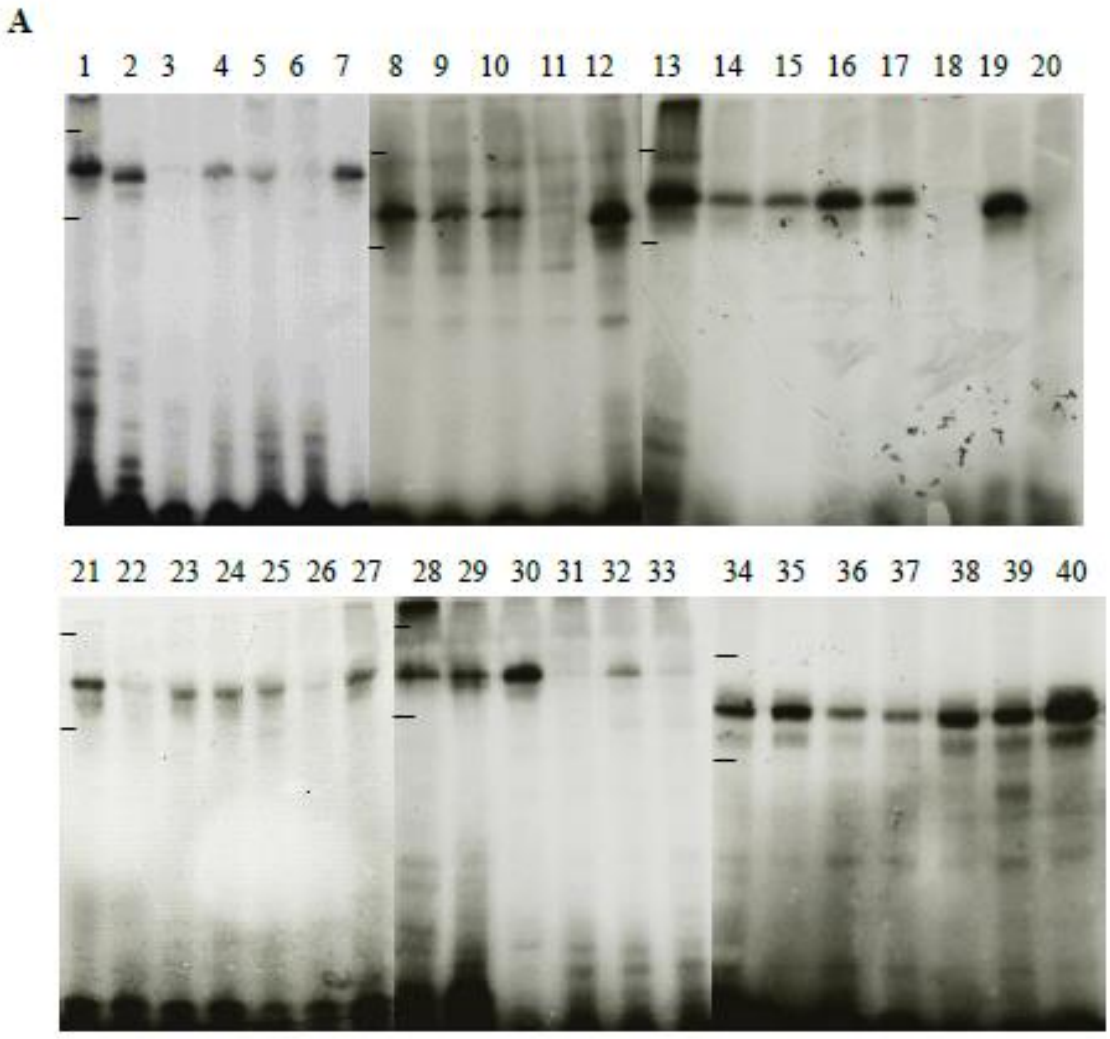
analysis in transgenic mice demonstrates that the *Prm1* TCE is sufficient to delay hGH translation, but mutation of the CST has no apparent effect (Zhong et al. 2001).

Is the *Prm1* TCE present in other mRNAs? Sensitive FASTA searches with the mouse *Prm1* TCE, GAACAATGCCACCTGTC as query (wordsize = 1 and no limit on the similarity of the hits to the query) identify sequences with 0-2 mismatches immediately upstream of all *Prm1* poly(A) signals. A second copy of the *Prm1* TCE with 2-4 mismatches and/or indels is present ~43-59 nt upstream of virtually all *Prm1* poly(A) signals. *Prm1* TCE-like sequences are also present immediately upstream of the *Prm2* poly(A) signal in mouse and rat, three mismatches and two mismatches and a one nt indel, respectively. We suspect that the *Prm1* TCE-like sequences with lower similarity in other mRNAs are spurious, because the sequences and positions are not conserved, or the elements are in the 5' and 3' flanking regions (not shown). This reservation also applies to the *Prm1*-like TCE sequences immediately upstream of the *Prm2* poly(A) signal in human and other species, ≥4 mismatches and indels and weak conservation even in closely related pimates.

The *Smcp, Tnp1* and *Tnp2* 3'UTRs also contain conserved elements immediately upstream of the poly(A) signal. Like the situation with the *Prm1* TCE, FASTA detects sequences with small numbers of indels/mismatches in the same location in orthologous mRNAs, while less similar sequences are present inconsistently and in variable positions in non-homologous mRNAs (not shown).

We did not search our database for the Y-H element, which is purported to regulate translation of the *Prm2* and other mRNAs (Han et al., 1995), because Li and Baraban (2004) have shown that recombinant translin/trax binds clusters of G instead of the

specific sequence of the Y-H element. They also contend that the configuration of Gs in the Y-H element has been insufficiently characterized to be used as a reliable query.



B

SEQUENCE	#	IC	SEQUENCE	#	IC	SEQUENCE	#	IC
AACACCA	3	+	CACACCA	3	-/+	UACACCA	2	++
AACACCC	4	+	CACACCC	4	-/+	UACACCC	3	-/+
AACACCU	4	++	CACACCU	4	+	UACACCU	3	++
AACAUCA	2	++	CACAUCA	2	+	UACAUCA	1	++
AACAUCC	3	++	CACAUCC	3	+	UACAUCC	2	++
AACAUCU	3	++	CACAUCU	3	++	UACAUCU	2	++
ACCACCA	2	+	CCCACCA	2	-/+	UCCACCA	1	++
ACCACCC	3	+	CCCACCC	3	-/+	UCCACCC	2	-/+
ACCACCU	3	++	CCCACCU	3	+	UCCACCU	2	+
ACCAUCA	1	++	CCCAUCA	1	++	UCCAUCA	0	++
ACCAUCC	2	++	CCCAUCC	2	+	UCCAUCC	1	++
ACCAUCU	2	++	CCCAUCU	2	++	UCCAUCU	1	++

Figure 2.2 RNA-EMSA analysis of YRS sequences. Panel A. The 29 seven nt sequences of the degenerate YRS, [UAC][CA]CA[UC]C[ACU] (Giorgini et al. 2001), which differ at 2-4 sites from the *Prm1* 3' UTR YRS, UCCAUCA, were analyzed by RNA-EMSA using cytoplasmic extracts of adult testes, UV-crosslinking, SDS-PAGE and autoradiography with x-ray film to determine the size and relative labeling of the complexes. The probes were synthesized as α[^{32}P]-rUTP-labeled T7 bacteriophage RNA polymerase trancripts of nt 1-37 of the *Prm1* 3'UTR in which the UCCAUCA was replaced by 2-4 mutations at degenerate sites, and equal numbers of cpm of each probe were used in each experiment. The sequences of the YRS in the probes are as follows: Lane 1, *Prm1* 3'UTR YRS, UCCAUCA; Lane 2, AACAUCU; Lane 3, CACACCA; Lane 4, CACAUCC; Lane 5, CCCACCU; Lane 6, UACACCC; Lane 7, UACACCU; Lane 8, AACACCU; Lane 9, AACACCC; Lane 10, CACACCU; Lane 11, CACACCC; Lane 12, AACAUCU; Lane 13, UCCAUCA; Lane 14, AACACCA; Lane 15, AACAUCC; Lane 16, ACCAUCU; Lane 17, ACCACCU; Lane 18, CACACCC; Lane 19, CACAUCU; Lane 20, CCCACCC; Lane 21, UCCAUCA; Lane 22, UCCACCU; Lane 23, UACAUCC; Lane 24, UACACCA; Lane 25, CCCAUCU; Lane 26, CACAUCA; Lane 27, ACCAUCC; Lane 28, UCCAUCA; Lane 29, ACCAUCA; Lane 30, ACCAUCU; Lane 31, CCCACCA; Lane 32, CCCAUCC; Lane 33, UCCACCC, Lane 32, CCCAUCC; Lane 33, UCCACCC; Lane 34, UCCAUCA, Lane 35, AACAUCA; Lane 36, ACCACCA; Lane 37, ACCACCC; Lane 38, UACAUCU; Lane 39, UCCAUCC; Lane 40, UCCAUCU. Panel B. The relative intensity of labeling of each UV-crosslinked complex (IC) that exhibits virtually identical mobility to those with UCCAUCA were divided into three categories: similar in intensity to those form with UCCAUCA, ++; +, reduced intensity compared with UCCAUCA, and -/+. The lowest level of complex labeling is judged to be -/+ because long exposures usually detected faint bands with similar mobility to those with the UCCAUCA positive control. Each probe was analyzed at least twice and the designations of IC were based on varying exposures of x-ray film and multiple gels. Panel B also records the number of sites (#) at which each YRS differs from UCCAUCA. The 6 YRS that differ at one site from UCCAUCA were analyzed by Giorgini et al. (2001). The positions of the 66 and 45 kDa size markers are indicated by bars in the first lane of each gel.

2.5 Discussion

The contention that the 5' and 3' ends of the genes studied here are orthologues is based on the absence of paralogues and a variety of conserved protein and DNA sequences. In general, the elements discussed here are considered likely to have regulatory functions if they are present at the same positions in the UTRs of multiple species or the elements have been reported to have regulatory functions in any mRNA.

However, two problems with interpreting conserved sequences should be emphasized at the outset. First, developmental regulation of mRNA translation in spermatids is regulated by multiple mechanisms that determine the duration and strength of the initial repression and fine-tune the period and rate of active translation, any of which could be regulated by a specific element (Braun et al., 1989; Mali et al., 1989; Shih and Kleene, 1992; Fajardo et al., 1997; Bagarova et al., 2010). In addition, the absence of a phenotype of mutating the *Prm1* CST indicates that highly conserved sequences can have elusive functions (Zhong et al. 2001). Second, it is well known that orthologous genes are not under identical selective pressures (Koonin, 2005), and therefore may not bear exactly the same conserved sequences. Indeed, this study identifies numerous sequence differences in the UTRs of various mammalian species, some of which may result in differences in post-transcriptional regulation. For example, the low levels of *Prm2* mRNA in some mammals may reduce the selective advantage of translational delay (Kleene and Bagarova, 2008). All of these uncertainties underscore the necessity of evaluating the functions of conserved sequences by studying mutations in transgenic

mice. This very important point applies to most of the findings in this study and will not be repeated.

Although the focus of this study is translational regulation, two important features of gene expression in spermatids noted previously apply to the seven new genes studied here (Zhong and Kleene, 1999; Kleene and Bagarova, 2008). First, a majority of all 12 genes have conserved CRE-like elements consistent with the idea that CREMτ has an important role in transcription in spermatids (Delmas and Sassone-Corsi, 1994). However, the conserved gene-specific differences in the CRE-like elements imply that the consensus sequence, TGACGTCA, derived from studies of recombinant CREMτ does not accurately reflect the diversity of elements *in vivo*. Second, these 12 genes have not spawned retroposons, arguing that piRNAs and methylation protect the genome of the male gametes by effectively quenching the amplification of retrogenes (Aravin et al., 2008).

Translation of uORFs usually decreases translation of the principal ORF, but the strength of repression depends on the sequence and configuration of the uORF and trans-factors (Sonenberg and Hinnebusch, 2009). Studies in transgenic mice demonstrate that uORFs contribute to the developmental regulation of the *Smcp* mRNA by briefly repressing translation in early spermatids and that the repression is neutralized by the *Smcp* 3'UTR in late spermatids (Bagarova et al., 2010). The uORFs in the *Gapdhs*, *Spata18* and *Akap3* 5' UTRs may also play a role in temporal regulation of mRNA translation, providing the effects of the uORFs are modulated by trans-factors.

Y-box proteins are distinguished by a central cold shock domain and a carboxy terminal domain containing alternating clusters of basic-aromatic and acidic amino acids

(reviewed in Skabkin et al., 2006). Y-box proteins have pleiotropic functions in transcriptional and post-transcriptional gene regulation including the architecture of mRNPs, mRNA stability and the repression of mRNA translation. Y-box proteins are hypothesized to control the timing of translation of many mRNAs in spermatids, because decreased levels of Y-box proteins in late spermatids correlate with the activation of many dormant mRNAs (Oko et al., 1996; Davies et al., 2000).

The *Prm1* 3'UTR YRS, UCCAUCA, is one of seven degenerate elements, differing at one site, [UAC][CA]CA[UC]C[ACU], that exhibit similar affinities for mouse Y-box proteins MSY2 and MSY4 (Giorgini et al., 2001). The present study demonstrates that 23 of the 29 elements that differ at 2-4 sites from UCCAUCA form complexes. We identify hundreds of YRS in the orthologous UTR database. Although many of these YRSs are likely to function in determining the structure of mRNPs (Skabkin et al., 2006), two are likely to regulate mRNA translation in spermatids by different mechanisms. The first is a FRGY2 YRS, one nt upstream of the *Tnp1* poly(A) signal, a position similar to those of the *Prm1* YRS and TCE that repress translation in early spermatids. The second is a YRS in the *Spata18* uORF, which may mediate a novel mechanism of developmental regulation: high Y-box protein levels in early spermatids could create a block to translation of the *Spata18* ORF by prolonging translation of the uORF (Sonenberg and Hinnebusch, 2009), which is released when Y-box protein levels decrease in late spermatids.

Eight of these mRNAs contain conserved mRNA-specific non-canonical poly(A) signals and contiguous or nearly contiguous canonical and/or non-canonical double poly(A) signals. Non-canonical poly(A) signals are thought to be unusually common in

spermatids and to specify upstream alternative poly(A) sites upstream of canonical poly(A) signals in somatic cells (Wang et al., 2006; Liu et al., 2006). The present findings are inconsistent with this idea because the non-canonical poly(A) signals are conserved and the databases lack ESTs with longer 3' UTRs and canonical poly(A) signals in somatic cells.

As far as we are aware this is the first report that non-canonical and multiple, adjacent poly(A) signals are conserved. The evolutionary pressures that limit the divergence of these poly(A) signals are unclear, and include functions in post-transcriptional gene regulation in the cytoplasm, or in regulating pre-mRNA stability in the nucleus (Kazerouninia et al., 2010).

The conserved sequences near the ends of the 3' UTRs identified here suggest a novel regulatory mechanism and future experiments. Studies of the *Prm1* and *Smcp* mRNAs in transgenic mice concur that the 3' UTR is the primary determinant of delayed translation (Braun et al., 1989; Bagarova et al., 2010), and mutations indicate that the regulatory elements must be positioned near the 3' terminus of both 3' UTRs (Zhong et al., 2001; Giorgini et al., 2002; Bagarova et al., 2010). The present findings reveal that the distal ends of the 3' UTRs of most of the mRNAs studied here are particularly well conserved including sequences upstream of the poly(A) signal, downstream of the poly(A) signal, and the poly(A) signals themselves. The latter two classes of sequences are rarely considered to function in translational control. That these conserved sequences are usually mRNA-specific implies that translation of each mRNA may be regulated by its own element, or by a smaller set of highly variable elements that cannot be recognized at present. The position of these putative elements near the 3' termini of the 3'UTRs

implies that mRNA-specific mechanisms may interact with the global regulatory mechanism involving the poly(A) binding protein and PABP-interacting protein 2a (Yanagiya et al., 2010) in a variant of the closed loop model which plays an important role in many forms of translational regulation (Sonenberg and Hinnebusch, 2009). We are using RNA EMSAs to identify proteins that bind to the conserved regions at the distal ends of various 3' UTRs as the first step in identifying sequences for studies of mutations in transgenic mice.

Funding

This research was supported by Grant MCB-0642128 from the National Science Foundation.

CHAPTER 3

YBX2 BINDS THE TRANSLATION CONTROL ELEMENT THAT REPRESSES *PRM1* mRNA TRANSLATION IN SPERMATIDS

This chapter was prepared as a manuscript to be submitted to Reproduction

3.1 Abstract

The protamine 1 (*Prm1*) mRNA exemplifies a wide-spread pattern of developmental regulation of mRNA translation in which mRNAs are transcribed in early haploid spermatogenic cells, spermatids, stored as translationally inactive free-mRNPs, and recruited onto polysomes in transcriptionally inert late spermatids. Previous studies of mutations in transgenic mice identify a 3'UTR translation control element (TCE) that appears to be necessary for total *Prm1* mRNA repression in early spermatids. However, the mechanism by which the TCE represses *Prm1* translation is unknown, because the factor that binds the TCE has not been identified. Here, modified UV-crosslinking assays, RNA-affinity chromatography and proteomics, and re-analysis of previous RNA binding assays concur that the failure to detect protein binding to the TCE is an artifact of RNase T1, and that Y-box protein 2, YBX2/MSY2, binds a *cis*-element, GCCACCU, in the *Prm1* TCE, and a similar element, AACAUCU, in the *Tnp1* 3'UTR. These findings agree with previous evidence that YBX2 is the predominant Y-box protein isoform in testis

mRNPs and the critical Y-box protein isoform for spermatid differentiation. The present and previous findings also suggest that YBX2, a well-documented global and sequence-specific translational repressor, targets specific mRNAs for repression in early spermatids by binding directly to Y-box recognition sequences close to the poly(A) signal, conditions that are fulfilled with reasonable frequency in mRNAs that are repressed in early spermatids. These findings have implications in human reproductive medicine because single nucleotide polymorphisms in the human *Ybx2* gene correlate with abnormal protamine expression and male infertility.

3.2 Introduction

Global and mRNA-specific translational controls influence the rates of protein synthesis during the post-meiotic phase of spermatogenesis in which haploid spermatids develop the highly specialized organelles of male gametes, spermatozoa (reviewed in (Kleene 2003, 2013)). The mRNA-specific developmental regulation of translation in spermatids is widely known and results from chromatin remodeling in which histones are replaced by spermatid-specific basic chromosomal proteins, transition proteins (TNP1 and TNP2) and protamines (PRM1 and PRM2), which package chromatin into a configuration that prevents transcription in late spermatids (Meistrich et al., 2003). The *Prm-Tnp* mRNAs encoding these chromatin-remodeling proteins are the most commonly studied temporally regulated mRNAs in spermatids: all four mRNAs are first detected in transcriptionally active step 7 early spermatids and are strongly repressed in free-mRNPs until the corresponding proteins are detected at least 3 days later in transcriptionally compromised step 10-11 spermatids (Kleene 1989, Mali, et al., 1989; Meistrich et al.,

2003). Delayed translational activation is necessary for sperm development since premature PRM1 and TNP2 expression in early spermatids causes abnormal spermatid development and male sterility (Lee et al., 1995; Tseden et al.. 2007).

The global repression of mRNA translation in early spermatids is less well known, but is demonstrated by proportions of polysomal mRNAs for active mRNAs in sucrose gradients, ~35-55%, that are considerably lower than those of fully active mRNAs in somatic mammalian cells, ≥85% (Kleene 2003, Mathews et al., 2007). In addition, mutations which eliminate strong translational repression of the *Prm1* and *Smcp* mRNAs in early spermatids result in levels of polysomal mRNA, ~33%, that are indicative of partial repression (Bagarova et al., 2010, Schmidt et al., 1999, Zhong et al., 2001). Evidently, the strong mRNA-specific repression in early spermatids is superimposed on default partial global repression.

Little is known about the mechanisms that regulate the timing of mRNA translation in spermatids. Several RNA binding proteins (RBPs) and miRNAs have been implicated in the regulation of *Prm-Tnp* mRNAs in early spermatids with gene knock-outs and studies of translational activity in cultured somatic cells and the rabbit reticulocyte lysate (reviewed in (Idler & Yan 2012, Kleene 2013, Nguyen-Chi & Morello 2011)). Unfortunately, these approaches have difficulty answering informative questions: Does repression result from direct or indirect interactions with putative target mRNAs? Does the factor repress translation strongly or weakly? The effects of knockouts of RBPs can be hard to connect with mRNA targets because RBPs are usually expressed for one or more weeks during spermatogenesis and potentially interact with thousands of mRNA species at different stages of spermatogenic cell development (Kleene 2013). In addition,

depletion of RBPs causes deleterious effects throughout the periods in which they are expressed in wild type meiotic and haploid cells, creating uncertainties whether the final effects on mRNA translation are direct or indirect. Studies in cultured somatic cells and reticulocyte lysates cannot elucidate the magnitude of regulatory effects of factors in spermatids which exhibit striking quantitative and qualitative differences in RBPs from other cell types in the mammalian body (Kleene 2013). Studies of mutations in *cis*-elements in transgenic mice can yield precise information about the regulatory functions of a factor, providing the mutation is specific for the binding of that factor. However, transgenic mice are considered too impractical and risky by many research groups.

The present study connects two disparate lines of research with evidence that Y-box protein 2 (YBX2), also known as MSY2, is the elusive factor that binds a *cis*-element in the *Prm1* 3'UTR that is necessary for complete *Prm1* mRNA repression in early spermatids *in vivo*.

The first line is a remarkable series of studies of deletion and point mutations of the role of the *Prm1* 3'UTR in translational repression in transgenic mice culminating with a report that repression in early spermatids requires the translational control element (TCE), GAACAAUGCCACCUGUC, which contains a putative Y-box recognition sequence (YRS), underlined (Braun et al., 1989, Zhong et al., 2001). The TCE merits special attention in studies of the mechanisms of strong translational repression in early spermatids, because it is the only *cis*-element that has been demonstrated with mutations to totally account for delayed translation of any mRNA in spermatids *in vivo* (reviewed in (Kleene 2013)). However, the pathway by which the TCE represses *Prm1* mRNA

translation is unknown because an RBP that binds the TCE has not been identified (Fajardo et al., 1994; Zhong et al., 2001).

The second line concerns the idea that Y-box proteins repress mRNA translation in early spermatids. Y-box proteins are ssDNA and ssRNA binding proteins which are characterized by a variable N-terminal alanine- and proline-rich domain, a highly conserved cold shock domain, and a variable C-terminal domain containing alternating 15-30 amino acid islands rich in basic-aromatic amino acids and acidic amino acids (reviewed in (Eliseeva et al., 2011)). Y-box proteins bind mRNA specifically and non-specifically, mediated primarily by the cold shock and C-terminal domains, respectively (Bouvet et al., 1995, Manival et al., 2001). The elements to which Y-box proteins bind specifically are known as Y-box recognition sequences, YRSs. Y-box proteins function in virtually every aspect of gene expression, from transcription to degradation including melting RNA-secondary structure and global repression of translation by packaging mRNAs into mRNPs that are inaccessible to the translational apparatus (Eliseeva et al., 2011). However, mRNA-specific post-transcriptional regulation by Y-box proteins is also demonstrated with point mutations in YRSs that release translational repression, decrease stability and alternative splicing of many individual mRNA species (Eliseeva et al., 2011; Giorgini et al., 2001; Matsumoto et al., 1996).

The mouse genome contains three genes encoding four Y-box protein isoforms, YBX1/MSY1, YBX2/MSY2, YBX3S and YBX3L/MSY4, all of which are expressed in spermatids (Eliseeva et al., 2011; Kwon et al., 1993; Mastrangelo & Kleene 2000). The *Ybx3* pre-mRNA is alternatively spliced producing two mRNA variants encoding short and long isoforms, YBX3S and YBX3L, which, respectively, have three and four

basic/aromatic and acidic amino acid islands in their C-terminal domains (Mastrangelo & Kleene 2000). YBX2 and YBX3L exhibit striking correlations with the developmental regulation of mRNA translation in spermatids because both proteins are expressed at high levels in early spermatids, their levels decrease subsequently becoming undetectable in late spermatids, and the vast majority of both proteins are present in free-mRNPs (Davies et al., 2000, Giorgini et al., 2002; Kwon et al., 1993; Oko et al., 1996; Yang et al., 2005).

The functions of YBX2 and YBX3L in mRNA-specific translational repression in early spermatids are unclear. YBX2 and YBX3L have not been demonstrated to bind the *Prm1* TCE and a YRS which binds YBX2 and YB3XL does not repress *Prm1* mRNA translation in its natural position (Giorgini et al., 2001; Zhong et al., 2001). These uncertainties extend to radically different opinions whether YBX2 and YBX3L are sequence-specific or non-specific RBPs, and whether mRNA-specific translational repression is mediated at the level of transcription by the promoter or by binding to YRSs in the *Prm1* 3'UTR (Giorgini et al., 2001; Kwon et al., 1993; Yang et al., 2005; Yu et al., 2002). The failure to identify a RBP that represses *Prm1* mRNA translation has been invoked as evidence that its repression might be mediated by a microRNA (Papaioannou & Nef 2010)

The present study formulates and affirms the hypothesis that YBX2 and YBX3L bind a previously unrecognized YRS in the *Prm1* TCE. This hypothesis originates in evidence that YBX2 and YBX3L bind a *Prm1* 3'UTR YRS, UCCAUCA (Davies et al., 2000). The bases that are necessary for YBX2 and YBX3L binding have been analyzed with point mutations at every position within and surrounding the YRS (Giorgini et al., 2001). RNA-EMSAs define the YBX2 YRS as a 7 nt element that lacks G and contains

three critical sites (underlined) and four degenerate sites (bracketed) [ACU][AC]CA[UC]C[ACU]. Most of the 36 permutations of bases at the degenerate sites have little effect on YBX2 binding (Chowdhury & Kleene 2012). The *Prm1* TCE contains a subsequence, GCCACCU, that potentially binds YBX2 and YBX3L, because it differs at one site from the degenerate YRS, the G in the first position. The likelihood that YBX2 and YBX3L binds GCCACCU is greatly increased by evidence that YBX2 and YBX3L bind a mutated YRS with a G in the first position in yeast three hybrid assays, GCCAUCA, even though they do not in RNA-EMSAs (Giorgini et al., 2001).

The present study demonstrates with UV-crosslinking RNA-binding assays and RNA affinity chromatography that YBX2 binds YRSs in the *Prm1* TCE and *Tnp1* 3'UTR, and that the inability to detect proteins that bind the TCE is an artifact of RNase T1 digestion.

3.3 Materials and methods

3.3.1 Animal research

Protocols for the maintenance and usage of mice in this study were reviewed and approved by the University of Massachusetts IACUC, Assurance # A3383-01, and are in accord with the 2011 NIH "Guide for the Care and Use of Experimental Animals". CD-1 mice were maintained on a 12 hr light - 12 hr dark cycle and were sacrificed with CO_2 hypoxia.

3.3.2 RNA binding assays

Our UV-crosslinking assays have been detailed previously (Chowdhury & Kleene 2012, Fajardo et al., 1994), and were used with minimal modification. Briefly, probes

were synthesized with the T7 bacteriophage RNA polymerase (New Egland Biolabs, Beverly MA) and α-[^{32}P]-rCTP (Perkin Elmer, Boston MA) from linearized, sequence-verified plasmids in which ds-oligos were ligated into the *Eco*RI and *Hind*III sites of pGEM3. 100,000 cpm of each probe was combined with 3 µl DEPC-H_2O and 5 µl 2X binding buffer (1X binding buffer is 20 mM HEPES, 3 mM $MgCl_2$, 40 mM KCl and 1 mM DTT, pH 7.6). Sequence-specific complexes were created by adding 1 µl of adult mouse testis cytoplasmic extract (25-50 µg/µl) and 1 µl *E. coli* tRNA (5 mg/ml), incubation for 20 min, digestion with 5U RNase T1 (Epicentre Biotechnologies, NT09100K, Madison WI) for 10 min, addition of 1 µl heparin (50 mg/ml, Sigma-Aldrich, St. Louis MO) for 10 min and UV-cross-linking with germicidal lamps. The samples were resolved on SDS-polyacrylamide gels, which were exposed to X-ray film.

3.3.3 RNA affinity chromatography and proteomics

20 µg 5'-biotinylated *Prm1* TCE RNA (Sigma-Aldrich Co., St Louis, MO) was heated to 70°C for 5 min in 400 µl 1X binding buffer containing protease inhibitor cocktail (Roche 11836170001, Indianapolis IN) and 5% glycerol, slow cooled, and incubated with ~500 µg cytoplasmic testis extract and 5 µg *E. coli* tRNA for 30 min at 25° C. The reactions were treated with 2 µl of heparin (200 mg/ml) for 10 min, and incubated with pre-washed streptavidin-agarose (Pierce 20347, Rockford IL) on a rotating disc for 2 hr at 4°C. After five 15 min 1 ml washes with 1X binding buffer with protease inhibitors, proteins were released by boiling for 5 min in 50 µl SDS sample buffer, separated by SDS-PAGE and visualized by silver staining (Pierce 24600, Rockford IL). Bands of interest were excised and identified with trypsin digestion and

mass spectrometry sequencing at the Taplin Mass Spectrometry Facility (Harvard Medical School, Boston, MA).

3.4 Results

Studies of RBPs in mouse testis extracts with RNA-EMSAs and non-denaturing electrophoresis, UV-crosslinking and SDS-PAGE, and Northwestern blots consistently detect a distinctive closely spaced doublet (Chowdhury & Kleene 2012, Davies et al., 2000; Fajardo et al., 1994; Giorgini et al., 2001, 2002; Kwon et al.,1993): a major slower migrating complex and a minor faster migrating complex. The presence of YBX2 and YBX3L in the major complex has been established with gel mobility super shifts, immunoprecipitation of native and UV-crosslinked complexes, and northwestern and western blots (Davies et al., 2000; Giorgini et al., 2002; Kwon et al., 1993). The mobilities of YBX2 (38.0 kDa) and YBX3L (38.8 kDa) in SDS PAGE, ~52 kDa, are slower than those expected from their MWs due to anomalous mobility of Y-box proteins (Davies et al., 2000; Eliseeva et al., 2011; Kwon et al., 1993).

To determine whether the *Prm1* TCE binds YBX2/YBX3L, short RNA probes were reacted with adult testis extracts and *E. coli* tRNA, treated sequentially with RNase T1 and heparin, UV-crosslinked and resolved with SDS-PAGE. The probes were labeled with α-[^{32}P]-rCTP, because Y-box protein binding requires a C-rich subsequence, CA[CU]C (Bouvet et al., 1995; Giorgini et al., 2001; Wei et al., 2012). Similar protocols are widely used with RNA-EMSAs and UV-crosslinking assays to detect sequence-specific RBPs in total cell extracts (Walker et al., 1998). RNase T1 decreases non-specific background by degrading probe that is not protected by an RBP, tRNA and

heparin compete with non-specific, electrostatic binding of basic amino acids, and UV covalently crosslinks amino acids that are in direct contact with bases. The combination of UV-crosslinking and heparin in our protocol also favors base-specific binding by the cold shock domain, because heparin suppresses non-specific binding by mouse YBX2 and UV-crosslinking of the C-terminal domain of FRGY2, the *Xenopus laevis* YBX2 orthologue, to poly(U,C) (Kwon et al., 1993; Ladomery & Sommerville 1994).

The probes in lanes 1-3 (Figure 3.1) contain high affinity YRSs which have been defined with many mutations: UCCAUCA in the *Prm1* 3'UTR, and AACAUCU in the *Tnp1* and *Smcp* 3'UTRs, referred to here as the FRGY2 YRS, because it binds *X. laevis* YBX2 (Bouvet et al., 1995; Chowdhury & Kleene 2012, Giorgini et al., 2001). The sequences of the probes are listed in the legend to Figure 3.1. All three probes form the strong and weak complexes described above, but the mobility of complexes vary slightly, consistent with the predicted number of bases (enclosed in brackets in the legend to Figure 3.1) in the cross-linked RNA fragments created by cleavage with RNase T1 after G-residues flanking the YRS in each probe.

UV-crosslinking did not detect complexes with the *Prm1* TCE (lane 4, Figure 3.2) in agreement with the failure of RNA-EMSAs to detect complexes with the corresponding segment of the *Prm1* 3'UTR (Fajardo et al., 1994). A G→U mutation in the first base of the putative TCE YRS, GCCACCU→UCCACCU, lane 5 (Figure 3.2), produces strong complexes, consistent with the degenerate YRS (Chowdhury & Kleene 2012). Thus, the G in the putative TCE YRS is incompatible with detecting complexes.

A report that a YRS containing a G in the first position, GCCAUCA, interacts with YBX2 and YBX3L in yeast three hybrid assays (Giorgini et al., 2001) suggests that

RNase T1, which digests after G-residues, might artifactually eradicate binding of YBX2 and YBX3L to GCCACCU. This hypothesis is intimated by evidence that rabbit YBX1/YB-1 greatly increases RNA degradation by RNase T1 (Evdokimova et al., 1995).

We predicted accordingly that binding of YBX2 and YBX3L to the TCE might be detectable with a protocol in which the probe is reacted with testis extract, treated with heparin and UV-crosslinked before RNase T1 digestion. This prediction was fulfilled by a complex with the *Prm1* TCE (lane 7, Figure 3.1) that is similar in intensity to, and slightly smaller and more diffuse, than that formed with the *Prm1* YRS (lane 6, Figure 3.1). These differences can be explained by the smaller distance between RNase T1 cleavage sites, 7 nt vs. 16 nt, and partial protection of UV-crosslinked TCE from RNase T1. The complex with the *Prm1* TCE is abrogated by a GCCACCU→GCACGAU mutation (lane 8, Figure 3.1), that drastically reduces binding of YBX2 to the *Prm1* YRS *in vitro* and eliminates translational repression in transgenic mice (Giorgini et al., 2001). Due to high background, the small weak complex is not detected.

To identify the proteins that bind the TCE, protein extract was incubated with biotinylated TCE-probe, treated with heparin, and complexes were captured with streptavidin-agarose. After washing, bound proteins were eluted in SDS-sample buffer and separated by SDS-PAGE. Silver staining (Figure 3.2, lane 1) detects a single, prominent band migrating at ~52 kDa, which contains YBX2 and YBX3L based on mass spectrometry sequencing of tryptic peptides. However, spectral counts indicate that YBX2 is more abundant than YBX3L, and the average precursor intensity of the four most abundant YBX2 and YBX3L peptides reveals that YBX2 is ~13-fold more abundant than YBX3L. The 52 kDa band is eliminated by a GCCACCU→GCACGAU

mutation in the TCE that eliminates binding in UV-crosslinking assays. In combination, the UV-crosslinking assays and RNA-pulldowns with wildtype and mutant TCEs both demonstrate that the abundant ~52 kDa protein that binds the TCE YRS is primarily YBX2.

The proteins in the weak lower complex also appear to be Y-box proteins because the intensity of these complexes with mutant and wildtype YRSs varies in parallel with that of YBX2 in RNA-EMSAs and UV-crosslinking assays (Figure 3.1 and (Chowdhury & Kleene 2012, Davies et al., 2000; Giorgini et al., 2001)). The weak small complex potentially contains YBX1 (35.7 kDa), YBX3S (30.7 kDa), a scarce alternatively spliced isoform of YBX2 (MSY2a, 31.5 kDa), or partially degraded YBX2 (Gu et al., 1998; Tafuri et al., 1993; Yang & Yen 2013, Yu et al., 2001). We detect YBX2, YBX3 and YBX1 in weak, small bands in RNA affinity chromatography (data not shown), but the peptide sequences do not distinguish between large and small variants of YBX2 and YBX3.

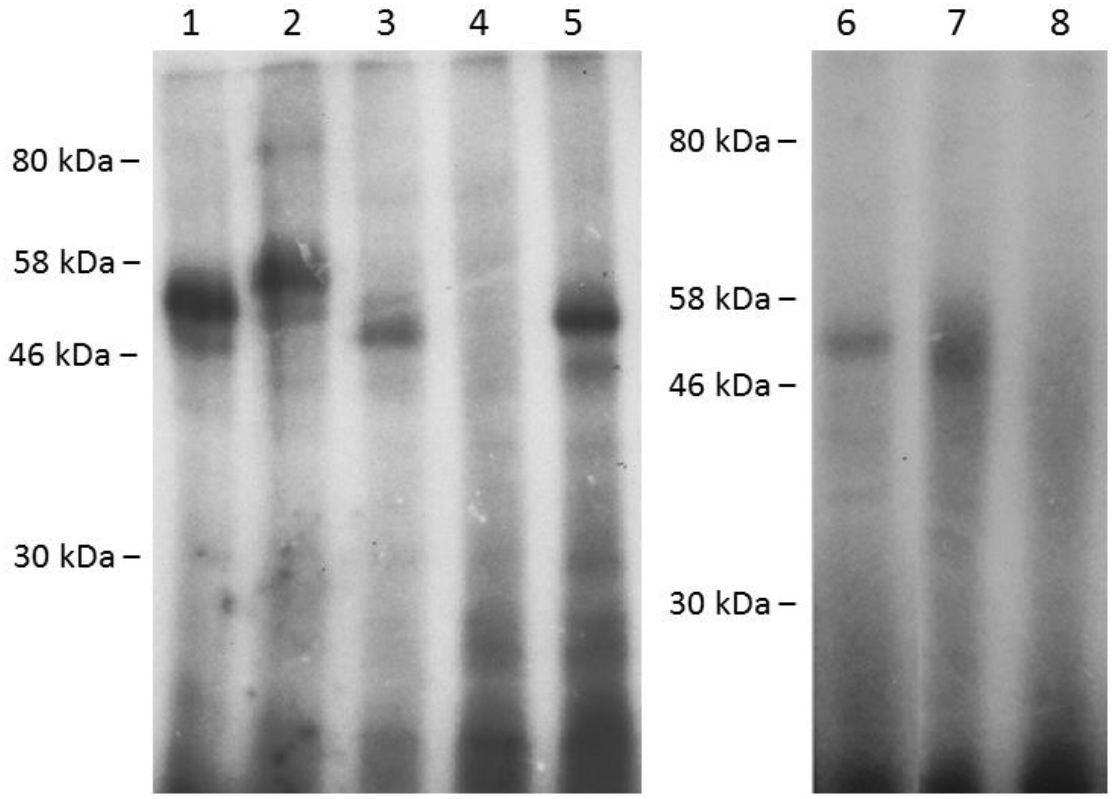


Figure 3.1 UV-crosslinking analysis of YRSs in the *Prm1*, *Smcp* and *Tnp1* 3'UTRs. [^{32}P]-labeled RNA probes were incubated with total testis cytoplasmic extracts, and sequence specific complexes were formed through the sequential use of *E.coli* tRNA, RNaseT1 and heparin. The complexes were UV cross linked, resolved by SDS-PAGE, and visualized by autoradiography. The samples in lanes 6-8 were treated with heparin and UV-cross linked before digestion with RNase T1. *Bona fide* and predicted wildtype YRSs are single underlined, and mutated bases are double wavy underlined. The number of nucleotides flanking the YRS resulting from RNase T1 digestion is enclosed in brackets. Lane 1, *Prm1* 3' UTR wild-type YRS [16 nt] AGAUGCACAGAAUAGCAAGUCCAUCAAAACUCCUG; Lane 2, *Tnp1* 3'UTR [29 nt] GAAUUCCCCCAACAUCUCAAUAACAUUUUGAAAACAAAUAAAAUUGUGA; Lane 3, *Smcp* wild type YRS [10 nt] GAAGGUAGAAAAGGAUAGAAACAUCUUGUCUAGUGAUCCUGACAUUUAGAU; Lane 4, *Prm1* wild-type TCE [7 nt] GAACAAUGCCACCUGUCAAUAAAU; Lane 5, *Prm1* TCE with G→U mutation in YRS [14 nt] GAACAAUUCCACCUGUCAAUAAAU; Lane 6, *Prm1*, 3' UTR wild type YRS [16 nt] AGAUGCACAGAAUAGCAAGUCCAUCAAAACUCCUG; Lane 7, *Prm1* wild type TCE [7 nt] GAACAAUGCCACCUGUCAAUAAAU; Lane 8, mutated *Prm1* TCE, GAACAAUGCACGAUGUCAAUAAAU.

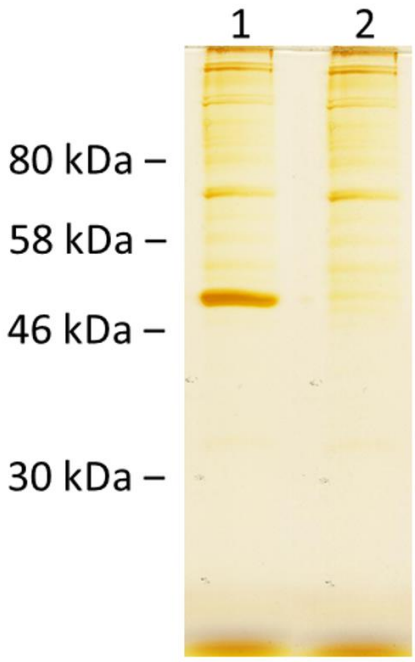


Figure 3.2 Identification of proteins binding to the *Prm1* TCE with RNA-affinity chromatography and mass spectrometry sequencing. Cytoplasmic extracts were reacted with 5' biotinylated *Prm1* TCE, treated with heparin, bound to streptavidin-agarose, and proteins were eluted with SDS-sample buffer. Lane 1, proteins bound to biotinylated TCE, GAACAAUGCCACCUGUCAAUAAAU; Lane2, proteins bound to mutant biotinylated TCE, GAACAAUGCACGAUGUCAAUAAAU.

Table 3.1 Positions of YRSs relative to canonical and non-canonical poly(A) signals and poly(A) sites in translationally regulated mRNAs in mammalian spermatids

mRNA (Mammal)[1] 3T3U Sequence[2]
Dazap1 long (*Mm*)[3]
UGACCAGUUUGACCCGGUUUGAAUAAAACAGCGUGUUUGGAUCAGA
Dazap1 short (*Mm*)[3]
UUUUCUCUGACCCAUCAGCACAAUAAAAACACGUCACUGGUUCAACAACA
*Dibl*5 (*Rn*)[3]
AGCAGGGUUAGCAGAAACAUCAAUAAAUCAUUCAAACUGCA
Gapdhs (Mm)
CGCGCCACCGGGUCAACAAUGAAAUAAAAACGAGAAUGCGCACA
Odf2 var5 (Mm)
GAGCUAUCAUCAGUGCUGUGAAAUAAAAGUCUGGUGUGCCA
Pgk2 (*Mm*)[3]
GGAAACAUUCUCAUGUCAACUAUUAAAGAAGUGAGCUAAGUAAGUU
Prm1 (*Mm*)[3]
UAAGGAACAAUGCCACCUGUCAAUAAAUGUUGAAAACUCA
Prm2 (*Mm*)[3]
ACUGAGCAAAGCCACCUGCCAAAUAAAGCUUGACACGAGA
Smcp (*Mm*)[3]
CUGACAUUUAGAUAGCAAAGAAAUAAAAGAGCAAAUAAAAAGA
Tnp1 (*Mm*)[3]
GCUGUUUCUCCCCAACAUCUCAAUAACAUUUUGAAAACAAAUAAAAUUGUGA
Tnp2 (*Mm*)[3]
AGCAAUUUCUAUGCAACAUGGAUUAAAGCUUGUACCCUGGAAGACUA
Ybx2 (*Mm*)
UCAUGUGCCACCUGAGCCUCCAGUAAAAACAAAAGCAGGCUUUCA

[1]The name of the mRNA and species of mammal from which the sequence is derived. [2]The sequence of the 3' terminus of the 3' UTR of the mRNA was reported in the indicated GenBank reference sequences. Mouse (*Mus musculus*) *Dazap1*, DAZ associated protein 1 short and long mRNA variants (NM_133188.2); rat (*Rattus norvegicus*) *Dibl5*, diazepam binding inhibitor-like 5 mRNA (NM_021596.2); mouse *Gapdhs* (XM_006539545.1), mouse *Odf2* transcript variant 5 (NM_001177661.1), mouse *Pgk2*, phosphoglycerate kinase 2 (NM_031190.2); mouse *Prm1*, protamine 1 mRNA (NM_013637.4), mouse *Prm2* mRNA, (NM_008933.1); mouse *Smcp*, sperm mitochondria-associated cysteine rich protein mRNA (NM_008574.3); mouse *Tnp1* mRNA (NM_009407.2); mouse *Tnp2* mRNA (NM_013694.4), mouse *Ybx2* mRNA (NM_016875.2). Canonical and non-canonical poly(A) signals are bold underlined, poly(A) addition sites based on 3' RACE or analysis of expressed sequenced tags are boxed (Chowdhury & Kleene 2012, Kleene & Bagarova 2008, Yang & Yen 2013), degenerate YRSs are double underlined.

[3]References citing evidence for developmental patterns of translational regulation (Chowdhury & Kleene 2012, Kleene 2013, Yang & Yen 2013).

3.5 Discussion

The present findings reveal that YBX2 is the predominant RBP that binds GCCACCU in the *Prm1* TCE and is therefore a promising candidate for a factor that represses *Prm1* mRNA translation in early spermatids. This statement summarizes several insights that are supported by and clarify previous studies.

First, the identity of the Y-box protein isoforms in complexes with testis extracts has been confused by high levels of expression and virtually identical sizes of YBX2 and YBX3L (Davies et al., 2000, Kwon et al., 1993). However, the proteomics evidence here that YBX2 is the predominant Y-box protein in complexes with the TCE agrees with supershift assays of complexes containing a single copy of *Prm1* YRS, UCCAUCA: YBX2-antibody produces a strong supershift while YBX3-antibody produces a negligible supershift (Davies et al., 2000). Interestingly, YBX2 and YBX3L bind similar sequences in yeast three hybrids (Giorgini et al., 2001), but in testis extracts YBX2 binds strongly to a single *Prm1* YRS, while YBX3L binds strongly only to a double YRS (Davies et al., 2000). It has been suggested that YBX3L binds as a heterodimer with YBX2 (Davies et al., 2000), but a YBX3L:YBX3S heterodimer has not been considered.

Second, the binding of YBX2 to GCCACCU concurs with reports that mouse YBX2, FRGY2 and human YBX1 bind strongly to degenerate YRSs containing a G in the first position in RNA binding assays without RNase T1 (Bouvet et al., 1995; Giorgini, et al., 2001; Wei et al., 2012). We suggest that digestion with RNase T1 diminishes

YBX2 binding to GCCACCU by decreasing the size of the YRS below a critical length, because cold shock domains bind more strongly to 7 nt ssRNAs than they do to 6 nt ssRNAs (Mayr et al., 2012; Sachs et al., 2012). We envision the YRS as a 7 nt element, [ACGU][AC]CA[UC]C[ACU], in which most permutations at the degenerate sites bind strongly (Chowdhury & Kleene 2012). The possibility that YBX2 binds YRSs containing G at other positions in the absence of RNase T1 can easily be examined with purified recombinant YBX2. It is also conceivable that YBX2 binds sequences that diverge from the degenerate YRS, because interactions of cold shock domains and ssRNA involve two flexible molecules which can assemble in different configurations with diverse elements (Cléry et al., 2013).

Third, gene knockouts suggest that YBX2 is the critical Y-box protein isoform for *Prm1* mRNA repression. Specifically, the *Ybx2*-null mutation blocks the differentiation of late spermatids and produces male infertility (Yang et al., 2007), while the *Ybx3*-null mutation causes depletion of meiotic and haploid spermatogenic cells that increase with the age of the mouse accompanied by minor defects on spermatid differentiation and subfertility (Lu et al., 2006).

The idea that YBX2 specifically represses *Prm1* mRNA translation is enigmatic because the levels of YBX2 are sufficiently high to repress all mRNAs. YBX2 is the most abundant RBP in testis mRNPs and represents ~0.7% of total protein in adult testis (Herbert & Hecht 1999, Yang et al., 2005), but the YBX2 levels in early spermatids are even higher because YBX2 is most abundant in late meiotic cells and early spermatids and is less abundant or undetectable in other testicular cells (Oko et al., 1996). The levels of YBX2 in early spermatids can be placed in perspective by considering that YBX1 and

YBX2 comprise ~0.1% and 2% of protein in cultured mammalian cells and mouse oocytes, which correspond, respectively, to 5-10 and 73 molecules of protein for each mRNA (Davydova et al., 1997, Yu *e*t al., 2001). The ratio of YBX2 to mRNA in early spermatids cannot be calculated precisely, because the amount of total mRNA in early spermatids is unknown. However, the ratio likely exceeds the 20:1 ratio at which FRGY2 globally represses translation in *Xenopus* oocytes (Matsumoto et al., 1996). The *Ybx2*-null mutation increases the polysomal loading of two translationally active mRNAs in early spermatids, *Ybx3* and *Acr*, from ~20-30% to <50%, a level well below that of fully active mRNAs, ≤85% (Yang et al., 2007). The partial activation of the *Yb3x* and *Acr* mRNAs implies that YBX2 alone is not sufficient for the partial global repression in early spermatids, and that additional factors are required (Kleene 2013). Unfortunately, the effect of the *Ybx2*-null mutation on the strong *Prm-Tnp* mRNA repression in early spermatids has not been studied with sucrose gradients and immunohistochemistry.

Evidence that the *Prm1* TCE represses *Prm1* translation (Zhong et al., 2001) disagrees with ideas that repression by YBX2 is determined during transcription by the presence of a double-strand DNA Y-box promoter element (Yang et al., 2005). Specifically, microarray analyses reveal that YBX1/YB-1 exhibit no preference for double stranded Y-box promoter elements (Dolfini & Mantovani 2013; Eliseeva et al., 2011; Zasedateleva et al., 2002). More importantly, studies of mutations in transgenic mice consistently implicate mRNA sequences, not promoters, in mRNA repression in early spermatids (Bagarova et al., 2010; Braun et al., 1989; Fajardo et al., 1997; Giorgini et al., 2001, 2002; Hawthorne et al., 2006; Lee et al., 1995; Schmidt et al., 1999; Tseden et al., 2007; Zhong et al., 2001) .

The mRNA specific repression by the *Prm1* TCE YRS implies that the properties of certain YRSs, such as binding affinity, number and position, select mRNAs for strong repression. The importance of position is indicated by findings that transposing the TCE from its natural position immediately upstream of the poly(A) signal to the middle of the *Prm1* 3'UTR or the *Prm1* 5' UTR inactivates its ability to repress translation in early spermatids (Robert E. Braun 2013, personal communication). Similarly, the *Prm1* YRS, UCCAUCA, in an unnatural position, 16 nt upstream of the poly(A) signal, partially represses translation in early spermatids, while the *Prm1* YRS in its natural context, 110 nt upstream of the poly(A) signal, does not (Fajardo *et al*. 1997, Zhong *et al*. 2001). In addition, FRGY2 YRSs in the *Smcp* 3' UTR 37 nt upstream of the poly(A) signal and *Smcp* 5' UTR account for little or no translational repression in transgenic mice (Bagarova *et al*. 2010). The similar position-dependence of both *Prm1* YRSs implies that strong repression by YBX2 requires unidentified additional factor(s) which potentially bind the 3' poly(A) tail, canonical and non-canonical AAUAAA polyadenylation signals, or unrecognized short, degenerate elements.

Table 3.1 lists 11 mRNA species which illustrate various relationships between translational repression, poly(A) signals and YRSs at the 3' termini of their 3' UTRs, referred to below as 3T3Us. This information should facilitate designing mutations in *cis*-elements and RNA binding studies in elucidating the mechanisms of repression by YBX2. The mouse *Gapdhs*, *Odf2*, *Prm1*, *Prm2*, *Smcp*, *Tnp1* and *Tnp2* mRNAs and rat *Dbil5* mRNAs are strongly repressed in early spermatids and actively translated in late spermatids (Kleene 1989, Pusch *et al*. 2000). The *Dbil5*, *Prm1* and *Prm2* 3T3Us have degenerate YRSs $\leq$4 nt upstream of canonical AAUAAA poly(A) signals. The FRGY2

YRS in the *Tnp1* 3T3U is one nucleotide upstream of a non-canonical poly(A) signal, AAUAAC, and 20 nt upstream of a canonical AAUAAA poly(A) signal. The failure of overexpressed YBX3L to repress the *Tnp1* mRNA needs to be clarified (Giorgini et al., 2002).

Four mRNAs in Table 3.1 potentially contain unrecognizable 3T3U YRSs. Overexpression of YBX3L in transgenic mice represses two mRNAs with 3T3U YRSs (*Prm1*, *Prm2*) and three mRNAs that lack obvious 3T3U YRSs (*Gapdhs*, *Odf2* and *Tnp2*) (Giorgini et al., 2002). We speculate that overexpression of YBX3L enables repression by binding to single YRS elements as a monomer because the YBX3S isoform was not overexpressed, which limits the formation of YBX3L:YBX3S heterodimers. It is notable that the *Tnp2* 3T3U lacks an obvious YRS, because the *Tnp2* mRNA is evolutionarily related to and repressed during the same period as the *Prm1*, *Prm2* and *Tnp1* mRNAs (Meistrich et al., 2003). Although the *Smcp* 3T3U contains two canonical poly(A) signals and lacks an obvious YRS, the *Smcp* 3T3U binds YBX2 in RNA affinity chromatography (TAC and KCK, unpublished). It should be easy to determine whether the *Gapdhs*, *Odf2*, *Smcp* and *Tnp2* 3T3Us contain non-cognate YRSs by analyzing mutations with recombinant YBX2.

YBX2 also appears to repress mRNAs that do not undergo delayed activation in late spermatids. The *Dazap1* pre-mRNA is polyadenylated at two sites producing 3'UTRs of different length (Yang & Yen 2013). Both *Dazap1* mRNA variants are fully active in early meiotic cells which do not express YBX2, but the short variant is partially repressed and the long variant is weakly repressed after the appearance of YBX2 in pachytene meiotic cells (Oko *et al*. 1996, Yang & Yen 2013). Interestingly, repression of

the short *Dazap1* variant correlates with a 3T3U YRS four nt upstream of its poly(A) signal and binding of YBX2 and YBX1 by its 3' UTR, and active translation of the long variant correlates with the absence of a 3T3U YRS (Table 3.1). The *Prm1* TCE YRS in the *Ybx2* 3' UTR, 7 nt upstream of a non-canonical poly(A) signal, may create an autoregulatory loop that represses the *Ybx2* mRNA.

The *Pgk2* mRNA is expressed at high levels from preleptotene meiotic cells through early spermatids in the absence of detectable PGK2 protein and is activated in late spermatids ((Danshina et al., 2010) and references therein). The *Pgk2* mRNA seems to be a clear example of an mRNA that is not repressed by YBX2, because it lacks a 3T3U YRS, is repressed in early meiotic cells before the appearance of YBX2, and is not derepressed in *Ybx2*-null mice (Yang et al., 2007).

Understanding the mechanisms by which the YBX2 and the *Prm1* TCE repress translation requires delineating the pathway by which factors bound to the 3T3U block binding of the small ribosomal subunit to the m7G 5' cap at the 5' end of the mRNA (Jackson et al., 2010). This pathway potentially includes additional unidentified factors that bind YBX2 and other 3T3U elements, and protein kinases and helicases that potentially modulate binding of YBX2 to mRNA (Herbert & Hecht 1999; Tsai-Morris et al., 2004; Weston & Sommerville 2006, Zhong et al., 1999). However, the position dependence of both *Prm1* YRSs indicates the necessity of analyzing mutations in *cis*-elements in transgenic mice. The prevalence of repression by YBX2 would be clarified by determining which 3T3Us in Table 3.1 bind YBX2 and whether mutation of the YRSs in several mRNAs releases repression. Understanding of the mechanism by which the *Prm1* TCE represses translation would be increased by studies of mutations to determine

whether the YRS is the only sequence in the 17 nt *Prm1* TCE that is necessary for strong repression, whether the combination of the *Prm1* TCE-YRS and poly(A) signal is able to repress translation at the proximal end of the 3' UTR, and the configurations of YRSs and canonical and non-canonical poly(A) signals that produce strong translational repression. Including measurements of poly(A) length in these studies could illuminate the basis of the enigmatic correlation between long poly(A) tails and translational repression and short poly(A) tails and active translation of the *Dazap1*, *Prm-Tnp* and *Smcp* mRNAs (Kleene 1989, Yang & Yen 2013). The single most important idea here is the necessity of studies of mutations in *cis*-element in spermatids in elucidating mechanisms. Finally, understanding the mechanisms by which YBX2 regulates translation in spermatids may lead to understanding of the functional correlations between single nucleotide polymorphisms in the human *Ybx2* gene with male infertility and abnormal protamine expression (Hammoud et al., 2009).

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this research.

Funding

This research was supported by a Doctoral *Dissertation* Research *Grant* to T.A. Chowdhury from the University of Massachusetts Boston.

Acknowledgements

We thank Dr. Alexey Veraksa for discussions about quantification by mass spectrometry, and Reviewer 1 for perceptive comments about Y-box proteins.

CHAPTER 4

IDENTIFICATION OF CIS-ELEMENTS AND RNA-BINDING PROTEINS THAT CONTROL THE TIMING OF SPERM MITOCHONDRIA-ASSOCIATED CYSTEINE-RICH PROTEIN MRNA TRANSLATION IN TRANSGENIC MICE

Please note that this chapter is written as a manuscript containing experiments performed by Danielle Cullinane and myself. My work contains the analysis of the $S^5G^CS^3$ transgene in seminiferous tubule squashes and sucrose and Nycodenz gradient analysis, studies of the protein-RNAbinding studies UV-crosslinking, RNA-affinity chromatography mass spectrometry. Danielle Cullinane analyzed the developmental expression of the $G^5G^CS^2$-mut2 transgene with seminiferous tubule squashes and sucrose and Nycodenz gradients.

4.1 Abstract

The sperm mitochondria-associated cysteine-rich protein mRNA is translationally repressed in early spermatids and translationally active in late spermatids. Previous studies in transgenic mice have demonstrated that the *Smcp* 5' and 3' UTRs alone account for partial repression. Here, we demonstrate that *Smcp* 5' and 3' UTRs are

required for full translational repression. We further demonstrate that replacement of the 16 nt downstream of the first AAUAAA polyadenylation signal with the 17 nt downstream of the early SV40 polyadenylation signal results in abrogation of translational repression by the *Smcp* 3' UTR. UV-crosslinking RNA binding assays and RNA affinity chromatography, and mass spectrometry sequencing identify proteins that interact with 3' termini of the *Smcp* and SV40 3'UTRs that potentially repress and activate translation.

4.2 Introduction

Translational regulation is crucial for controlling gene expression during the development of post-meiotic, haploid spermatogenic cells, spermatids, because transcription in late spermatids ceases due to chromatin remodeling (Kierszenbaum and Tres 1975; Kleene 1996, 2003, 2013; Meistrich et al., 2003). In the absence of transcription, delayed activation of mRNA translation is utilized to synthesize such proteins as protamine 1 (PRM1) and the sperm mitochondria-associated cysteine-rich protein (SMCP) in the final stages of sperm differentiation (Chowdhury and Kleene, 2012). The mRNAs encoding these proteins are transcribed in early haploid cells, round spermatids, and stored as translationally inactive messenger ribonucleoprotein particles (free-mRNPs) for several days to a week before the mRNA is translated in transcriptionally inactive late haploid cells, elongated spermatids (Kleene 1989, reviewed in Kleene, 2003, 2013). Repression of mRNA translation in round spermatids is necessary for normal sperm development since premature activation of translation of the

Prm1 and *transition protein 2* (*Tnp2)* mRNAs in round spermatids in transgenic mice leads to deformed spermatozoa and reduced male fertility (Lee et al. 1995; Tseden et al. 2007). Although the delayed activation of translation is thought to be a wide-spread phenomenon involving mRNA species encoding hundreds of proteins in the specialized organelles of the spermatozoon, the developmental patterns of translational activity of few mRNAs have been described carefully (reviewed in Kleene 2013).

mRNA-specific translational regulation usually involves *cis*-elements which bind *trans*-factors, either RNA binding proteins (RBPs) or small non-coding RNAs, which activate or repress translation (reviewed in Jackson et al. 2010; Groppo and Richter 2009, Kleene, 2013). More than 20 RNA-binding proteins and several microRNAs have been implicated in developmental regulation of mRNA translation in spermatids based on studies of knockout mice and overexpression of specific RNA binding proteins (reviewed in Idler and Yan 2009, Paronetto & Sette, 2010; Kleene 2013). Unfortunately, the complexity of post-transcriptional regulation and unusual features of mammalian spermatogenesis have stymied defining how much these factors affect the timing and efficiency of translational activity of individual mRNAs (reviewed in Kleene, 2013).

The complexity of post-transcriptional regulation is illustrated by findings that RNA binding proteins and microRNAs interact with many, sometimes thousands of, mRNA targets (Bartel 2009, Morris and Keene 2010; Kishore et al., 2010; Ascano et al., 2012). This creates questions in identifying which mRNA targets are regulated strongly, weakly or negligibly by specific miRNAs and RNA binding proteins. This challenging situation is further complicated in spermatogenesis by findings that all RNA binding proteins studied to date are expressed for prolonged periods, one to several weeks, and

the knockouts of genes encoding RNA binding proteins often have deleterious phenotypic effects in multiple stages spermatocytes, round spermatids, elongating and elongated spermatids (Kleene 2013). Consequently, it is unclear whether the knocked-out or overexpressed factors have direct effects on the translational activity of a specific mRNAs, or indirect effects mediated regulation of one or more upstream factors (Zhong et al., 1999; Dass et al., 2007; Kleene, 2013).

Another major problem is that the mechanisms of translational regulation in spermatids exhibit unique features. For example, the *Prm1* mRNA exemplifies unique regulatory mechanisms in which translationally repressed mRNAs have long poly(A) tails, active mRNAs have shortened poly(A) tails, and excess cytoplasmic poly(A) binding protein, PABPC1, represses translation instead of activating translation (Kleene 1989, Yanagiya et al., 2010; Yang and Yeh, 2012). In addition, a number of spermatid specific factors have been implicated in translational control in spermatids including all four isoforms of Y-box proteins and a cytoplasmic isoform of poly(A) polymerase (Mastrangelo and Kleene, 2000; Kashiwabara et al., 2000). Unfortunately, a culture system which supports the differentiation and DNA transfection of spermatids is unavailable (Hunter et al. 2012; Kleene 2013). Thus, transgenic mice are the only system for analyzing the regulatory effects of mutations in cis-elements. Since transgenic mice are expensive and laborious, the vast majority of studies of cis-elements have been carried out in cell-free translation and culture systems based on somatic cells which have dubious relevance to the atypical mechanisms of translational control in spermatids (Kleene, 2013).

At present, the *Prm1* and *Smcp* mRNA are the only mRNAs in which mutations in mRNAs have been analyzed in transgenic mice to identify cis-elements. A remarkable series of studies of deletion and point mutation in transgenic mice identify a translational control element (TCE) in the *Prm1* 3'UTR that is "necessary and sufficient" for translational repression in round spermatids (Braun et al., 1989; Braun 1990; Fajardo et al., 1997; Zhong et al., 2001). The translational repressor, Y-box protein 2 (YBX2/MSY2) was recently identified as the factor that binds the TCE (Chapter 2, Chowdhury & Kleene, in review). However, the identification of one *cis*-element and one factor in one mRNA species is no basis for making general statements about factors and elements that regulate hundreds of mRNA species.

The present study uses transgenic mice to study the developmental regulation of the *Smcp* mRNA. SMCP is a structural protein found in the keratinous capsule surrounding mammalian sperm mitochondria (Cataldo et al., 1996; Ursini et al.,1999). The *Smcp* mRNA is evolutionarily different from protamine and transition protein mRNAs which are commonly used in studies of translational regulation in spermatogenesis (Hawthorne et al., 2006A). Therefore, studies of the *Smcp* mRNA begin to address the question whether all the mRNAs in spermatogenesis are regulated by the same set of *cis*-elements and *trans*-factors as the *Prm1* mRNA.

The *Smcp* mRNA is synthesized in step 3 spermatids, and is stored as a translationally inactive free-mRNP for about 6 days before the mRNA is active in translation in step 11 spermatids as demonstrated by the appearance of the SMCP protein (Shih and Kleene, 1992; Cataldo et al., 1996). Previous studies show that the *Smcp* mRNA is regulated by multiple mechanisms involving both the 5'UTR and 3'UTR using

the EGFP reporter in transgenic mice (Hawthorne et al.,2006; Bagarova et al., 2010). However, the *Smcp* 5' UTR alone delays GFP expression until step 5, the *Smcp* 3'UTR alone delays GFP expression translation until step 9, and we have identified a mutation in the *Smcp* 3'UTR that results in partial, not complete, release of translational repression. Clearly, our studies to date have not identified all of the elements and factors that repress *Smcp* mRNA translation in round spermatids.

The present study continues our goals of identifying the *cis*-elements and *trans*-factors that are necessary and sufficient for translational repression of the *Smcp* mRNA from step 3 to step 11 spermatids. We have analyzed two new transgenes. The first contains the *Smcp* 5' UTR and 3'UTR to test the proposition that both UTRs are necessary to delay translational activation until step 11. The second transgene mutates a highly conserved segment in the *Smcp* 3'UTR downstream of the first AAUAAA polyadenylation signal to search for additional elements that repress translation in round spermatids.

We also use UV-crosslinking RNA binding assays, RNA affinity chromatography and mass spectrophotometry sequencing to demonstrate that YBX2 binds two sites in the 3' terminus of the *Smcp* 3'UTR, and that multiple proteins bind a segment of a transgenic mRNA that does not undergo delayed translational activation.

4.3 Materials and methods

4.3.1 Construction of the $S^5G^CS^3$ and $G^5G^CS^3$-mut 2 transgenes

The $S^5G^CS^3$ transgene was constructed from $G^5G^CS^3$ and $S^5G^3G^3$ transgenes constructed previously (Hawthorne et al., 2006). Briefly, plasmids containing the $G^5G^CS^3$ and $S^5G^3G^3$ transgenes were digested with *Bsrg I* and *Afl II*, and the large $S^5G^3G^3$ and small $G^5G^CS^3$ fragments were purified by agarose gel electrophoresis and a Gene Clean II kit (Bio101), and the small $G^5G^CS^3$ fragment was ligated into the large $S^5G^3G^3$ fragment.

The $G^5G^CS^3$-mut2 transgene was constructed from the $G^5G^CS^3$ and $G^5G^CG^3$ transgenes in several steps. A *Swa* I site was inserted overlapping the upstream *Smcp* poly(A) signal with overlap extension PCR in the $G^5G^CS^3$ transgene (Higuchi et al., 1988). The *Swa I-Afl II* fragment from the $G^5G^CG^3$ transgene was inserted into the *Swa I-Afl II* sites of the $G^5G^CS^3$. Finally, the Swa I site was reversed to that of the original *Smcp* 3'UTR with a second round of overlap extension PCR.

The plasmids were electroporated into *E. coli* DH5α, plated on LB agar containing 50 μg/ ml kanamycin, and the sequence of the transgene was verified by sequencing on both strands, the small *Xho* I and *Afl* II fragment containing the transgene is purified with agarose gel electrophoresis, extracted with a NucleoTrap kit (Clontech), filtered, and adjusted to 50 ng/μl in 0.1 mM EDTA, 5 mM Tris-HCl (pH 7.4). One-cell C57BL/6 X SJL F2 embryos are injected and implanted into pseudopregnant females at the University of Massachusetts Medical Center Transgenic Core facility and tail-biopsies were analyzed to determine which pups contain the transgene. After weaning, the founders are transferred to the UMass Boston Animal Care Facility. Transgenic founders are bred to C57BL/6 X SJL mice of the opposite sex to produce lines. To

identify transgenic mice, 5 mm is excised from the end of the tail of 10-21 day old pups in accord with NIH guidelines for genotyping transgenic mice, and the DNA is purified with a DNAeasy Blood and Tissue kit (Qiagen). The presence of transgenes is assayed by PCR using *Gfp*-specific primers (Hawthorne et al., 2006).

4.3.2 Analysis of GFP fluorescence in squashes of seminiferous tubules

The stage of GFP expression was analyzed in living spermatogenic cells as described previously (Bagarova et al., 2010) and is based on techniques described by Kotaja et al. (2004). Briefly, adult mice were put down by CO_2 asphyxiation and the testes were dissected out. Following the removal of *tunica albuginea*, seminiferous tubules were teased apart in phosphate buffered saline and visualized with a dissecting microscope to identify segments of potential interest based on the banding pattern of the tubules under transillumination (Kotaja et al., 2004). The stages of spermatids were identified in one cell thick squashes of 0.5 mm tubule segments by visualization with phase contrast microscopy at 1000X using an Olympus BX51 microscope equipped with a Plan Fluorite 100X phase objective (NA 1.3), 100 W mercury burner, and SPOT XPLORER monochrome camera, SPOT advance image processing software (Diagonistic Instruments, Sterling Heights, MI, USA). EGFP fluorescence was excited at 470 nm and visualized at 525 nm and photographed at a manual setting of 3 sec and γ=1, and are depicted as the grayscale images that were actually recorded by the camera. ImageJ (downloaded from NIH) was used to quantify the pixel intensity with GFP fluorescence associated with various cell types.

4.3.3 Sucrose and Nycodenz gradient analysis

Cytoplasmic extracts of adult testes or 21/25 day old testes were prepared by dissecting testes (1 testis for adult mouse and 2 testes for 21/25 day mice), removing the *tunica albuginea* and homogenizing the testes in 300 ul HNM buffer (20 mM HEPES, pH 7.4, 0.1 M NaCl, 3 mM $MgCl_2$) containing 0.5% Triton X100 and 1 unit/µl RNasin Plus (Promega Biotech). The nuclei were pelleted by centrifugation at 13,000 X g for 2 minutes, and 250 ul of the supernatant was layered on either a 3.8 ml linear 15-40% sucrose gradient in HNM buffer (w/w) or a 3.8 ml 20-60% (w/v) Nycodenz gradient (Accurate Scientific Chemical Corporation, Westbury, NY, USA) prepared by layering 760 ml of 60, 50, 40, 30, and 20% Nycodenz (Accurate Scientific Co.) in HNM (w/v) in polyallomer centrifuge tubes for the Beckman SW60 rotor. Sucrose gradients were centrifuged for 80 min at 35,000 rpm at 4°C, and ~0.4 ml fractions were collected onto 0.3 g guanine thiocyanate, and RNA was extracted as described previously (Kleene et al., 2010). Nycodenz gradients were centrifuged for 24 hr at 37,000 rpm at 4°C, and 0.2 ml fractions were collected, and RNA was extracted as for sucrose gradients with adjustments for the smaller volume of fractions. RNA was extracted from each fraction of sucrose or Nycodenz gradients using techniques that result in recovery of equal amounts of RNA from each fraction as described by Kleene et al. (2010).

4.3.4 Northern blot and quantitative reverse transcriptase real-time PCR

For northern blots, RNA was extracted from each fraction of sucrose or Nycodenz gradients as described by Kleene et al. (2010), and denatured in formaldehyde and formamide (Hawthorne et al. 2006B). RNA samples were loaded onto a 1% agarose gel with 2.2M formaldehyde in 10 mM phosphate buffer (Kleene 1989), electrophoresed for

about 4 hr at 35 volts, transferred to nitrocellulose membrane in 20X SSC and hybridized to ^{32}P labeled DNA probes. DNA probes were generated through PCR with specific primers listed in table 2 and labeled with [α^{32}P]-dCTP with Random Primers DNA Labeling System kit (Invitrogen). Northern blots were hybridized to [^{32}P]-labeled cDNA probes overnight in 5X SSPE, 0.2% SDS, 10X Denhardt's solution and 100 μg/ml denatured sonicated salmon sperm DNA at 65°C, and washed five times for 10 min in 0.2 SSPE, 0.2% SDS at 65°C, and quantified in Molecular Dynamics Storm Model 840 phosphoimager. Quantitative Reverse Transcriptase real-time PCR (RT-qPCR) was carried out as described by Bagarova et al. (2006).

4.3.5 UV-crosslinking RNA binding assays

UV-crosslinking RNA binding assays were carried out as described by Chowdhury and Kleene (2010). Plus and minus strands oligonucleotides corresponding to various segments of *Smcp* 3'UTR and 3' terminus of the $G^5G^CS^3$-mut2 3'UTR were purchased from Life Technologies (Grand Island, NY). The oligos were annealed, digested with *Eco* RI and *Hind* III, and ligated into the *Eco*RI and *Hind* III sites of pGEM3 (Promega-Biotec) downstream of the T7 promoter. The sequence of the insert was verified through sequencing at Massachusetts General Hospital DNA Sequencing Facility (Cambridge, MA). The plasmid was linearized with *Hind* III and probes were synthesized with the T7 bacteriophage RNA polymerase (New England Biolabs, Beverly MA) and α-[^{32}P]-rUTP (Perkin Elmer, Boston MA). Probes were extracted twice with phenol:chloroform, chromatographed on a Biogel P6 column (Bio-Rad), ethanol-precipitated, and dissolved in DEPC-treated H_2O. The cpm of each probe was determined by scintillation counting, and 10^5 cpm was used in each reaction.

RNA probes were combined with 3 µl DEPC-treated H_2O and 5 µl 2X Binding Buffer (40 mM HEPES, 6 mM $MgCl_2$, 80 mM KCl and 1 mM DTT, pH7.6) denatured by heating at 70°C for 5 mins, renatured by slow cooling to room temperature. Following renaturation, sequence specific complexes were created by the following incubations at room temperature, ~25^{o}C: (1) incubating the samples with 1 µl of cytoplasmic extract of adult testis (25-50 µg/µl) and *E. coli* tRNA (5 mg/ml) for 20 min, (2) digestion with RNase T1 (5U) for 10 min, (3) treatment with 1 µl heparin (50 mg) for 10 min. The samples were irradiated with UV using two Sylvania G15T8 germicidal bulbs at a distance of 8 cm for 8 min on ice, and mixed with 12 µl 2X SDS sample loading buffer, boiled for 4 min and resolved on SDS-polyacrylamide gels containing a 3 cm 5% stacking gel and a 20 cm 10% separating gel. Gels were fixed in methanol: H_2O: acetic acid (5:4:1), dried, and autoradiographed at -80°C with an intensifier screen.

4.3.6 RNA affinity chromatography

5'-biotinylated RNA probes were purchased from Sigma-Aldrich Corp. (ST Louis, MO). 20 ug of 5'-biotinylated RNA probes were mixed with 400 ul binding buffer (20 mM HEPES, 3 mM $MgCl_2$, 40 mM KCl, 5% glycerol, 1 mM DTT, pH 7.6), heated to 70°C for 5 minutes and slow cooled to room temperature before incubating with 1 mg total cytoplasmic testis extract and 5 µg of tRNA for 30 min. The samples were then treated with 2 µl of heparin (200 mg/ml) for 10 min, incubated with pre-washed streptavidin agarose (Pierce 20347, Rockford IL) on a rotating disc for 2 hr at 4°C. After five 1 ml washes with 1X Binding Buffer (with protease inhibitor), bound proteins were released by boiling in 2X SDS sample buffer, resolved by 10% SDS-PAGE and visualized by silver staining. Protein bands of interest were excised from the gel, and

identified with mass spectrometry sequencing at the Taplin Mass Spectrometric Facility (Boston, MA).

4.4 Results

Previous studies in transgenic mice show that neither the *Smcp* 5'UTR nor the *Smcp* 3'UTR alone are sufficient to delay GFP expression until step 11 (Hawthorne et al, 2006B; Bagarova et al, 2010). To determine whether proper translational repression of the *Smcp* mRNA requires interactions between the 5'UTR and the 3'UTR, we analyzed a new transgene, $S^5G^CS^3$, containing 518 nt of *Smcp* 5' flanking region, the entire *Smcp* 5'UTR, the Gfp coding region (720 nt) derived from the pEGFP plasmid and the entire *Smcp* 3'UTR. The promoter of $S^5G^CG^3$ transgene directs expression of the *Gfp* mRNA in early spermatids at the same transcription site and in the same cells as the natural *Smcp* mRNA (Hawthorne et al. 2006B). 11 founders were identified, and the expression of GFP was analyzed in 3 founders and 4 lines derived from the founders.

The second transgene was designed to identify elements in the *Smcp* 3'UTR that repress translation. The randomization of the sequence 6-38 nt upstream of first *Smcp* poly(A) signal in the $G^5G^CS^3$-mut1 transgene resulted in partial release of translational repression and abrogation of the binding of the translational repressor YBX2 (Bagarova et al. 2010). Since previous studies of mutant 3' UTRs of the *Prm1* transgene revealed that elements at the distal end of the 3'UTR mediate translational repression (Fajardo et al., 1997; Zhong et al. 2001; Giorgini et al., 2001), we reasoned that this might also apply to the *Smcp* 3'UTR. Since the mutation in the $G^5G^CS^3$-mut1 transgene covered the

conserved segment of the *Smcp* 3'UTR, we studied a mutation in the segment of the *Smcp* 3'UTR downstream of the upstream poly(A) signal which also contains two sequences which are conserved in many species of mammals, the downstream AAUAAA poly(A) signal and the GAGC between the two poly(A) signals (Chowdhury and Kleene 2012). We therefore replaced the sequence of the *Smcp* 3'UTR downstream of the first poly(A) signal with the corresponding sequence in the eGFP plasmid which was originally derived from the early SV40 tumor virus polyadenylation signal (Kessler et al., 1986; Wilusz and Shenk 1988). We refer to these 23 nt as the SV40 early poly(A) segment because it contains an AAUAAA poly(A) signal and 16 nucleotides between poly(A) signal and the poly(A) addition site which have no known function. We assumed that this sequence would lack *cis*-elements because the replacement of the *Smcp* 3'UTR with the pEGFP 3'UTR results in partial and total loss of translational repression in the $G^5G^CG^3$ and $S^5G^CG^3$ transgenes (Hawthorne et al., 2006; Bagarova et al., 2010). In addition, there are very few reports of cis-elements in the short 15-30 nt segments of 3'UTRs between the poly(A) signal and the polyadenylation site (Tian et al. 2005), and an exhaustive literature search found no studies of protein binding and effects of the early SV40 signal on post-transcriptional gene expression.

$G^5G^CS^3$

UAGAAACAUCUUGUCUAGUGAUCCUGACAUUUAGAUAGCAAAGAAAUAAAAGAGCAAAUAAAAAG
A

$G^5G^CS^3$-mut1

UAGAAAAGAUGAGAUCGUACUUCGAUAUUACCCUCAUGUUAAAGAAAUAAAAGAGCAAAUAAAAA
GA

$G^5G^CS^3$-mut2

UAGAAACAUCUUGUCUAGUGAUCCUGACAUUUAGAUAGCAAAGAAAUAAAGCAUUUUUUUCACUG
CA

Figure 4.1 Sequence of the 3' terminus of the natural and mutant *Smcp* 3'UTRs in transgenes. The FRGY2 YRS sequence is highlighted yellow, AAUAAA canonical poly(A) signals are highlighted grey, and the bases highlighted red are the poly(A) addition sites determined with 3'RACE (Chowdhury and Kleene 2012). The underlined sequence in the $G^5G^CS^3$-mut1 transgene is randomized and eliminates the CATC element that is essential for binding YBX2. The double underlined sequence in the $G^5G^CS^3$-mut2 transgene is derived from the early SV40 poly(A) signal in the pEGFP plasmid (Kessler et al. 1986).

4.4.1 Developmental expression of GFP fluorescence in $S^5G^CS^3$ mice

The developmental expression of GFP fluorescence was determined with phase contrast and fluorescence microscopy of single-cell layer thick squashes of short segments of living seminiferous tubules (Kotaja et al., 2004). Mouse spermatids are divided into 16 developmental steps based on cell associations and the morphology of the acrosome, nucleus and tail which have been described thoroughly by Russell et al. (1990) and Kotaja et al. (2004). Remarkably, the morphology of spermatogenic cells is more clearly visualized with phase contrast microscopy in squashes than with stained paraffin-embedded sections (Kotaja et al. 2004). GFP-expressing cells were identified initially by enhancing the brightness and contrast in ImageJ (Bagarova et al., 2010). In general, GFP positive cells exhibited fluorescence throughout their cytoplasm while the fluorescence of GFP-negative cells, pachytene spermatocytes and Sertoli cells, was not more intense than the background.

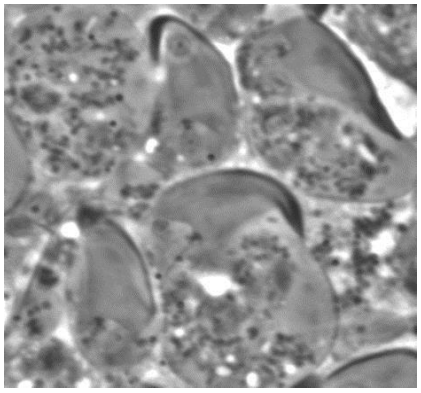
The critical stages for the initial GFP expression in $S^5G^CS^3$ transgenic mice are spermatids in steps 10, 11 and 12. Step 10 and 11 spermatids have clear nuclei because chromatin remodeling and nuclear condensation mediated by the replacement of histones by transition proteins and protamines has not yet begun (Meistrich et al., 2003). According to Russell et al. (1994) step 10 spermatids are identified by a sharp angle between the ventral and caudal surfaces and a rounded dorsal angle, whereas step 11 spermatids are characterized by sharp dorsal and ventral angles. In contrast, step 12 spermatids exhibit much longer nuclei which are darkened by chromatin remodeling and nuclear condensation. GFP fluorescence was not detectable in step 10 spermatids, and was clearly noticeable in step 11 and 12 spermatids (Figure 4.2).

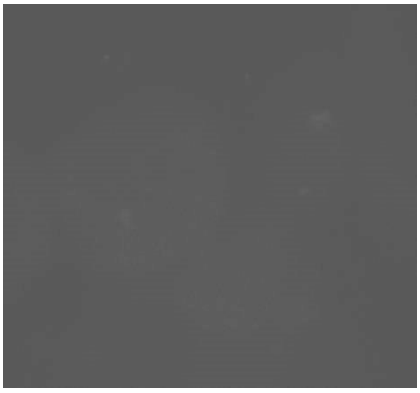
GFP expression in step 1-8 $G^5G^CS^3$-mut2 spermatids was difficult to visualize in squashes of seminiferous tubules from adult mice because intense fluorescence of step 13-15 spermatids obscures the low levels of GFP fluorescence in early spermatids. The developmental expression of GFP fluorescence was easily analyzed in 25-28 day $G^5G^CS^3$-mut2 spermatids which lack intensely fluorescent elongated spermatids. GFP fluorescence was not detected in step 1 $G^5G^CS^3$-mut2 spermatids which are distinguished by the absence of acrosomes, and was first detected in step 3 spermatids which are characterized by a circular acrosome with a central, dark acrosomal granule. As reported previously (Bagarova et al. 2010), GFP fluorescence is absent from the acrosomes, demonstrating that the GFP-protein is present in these cells and is excluded from the acrosome (Figure 4.2).

The pixel intensity of unenhanced fluorescing cells was matched with corresponding phase contrast images to identify the exact steps of spermatids that showed

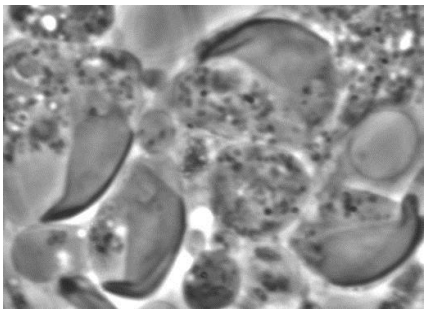
fluorescence, and the average pixel intensities over 10 cells were quantified with ImageJ (Figure 4.2), and a two sided unpaired t-test was carried out with the pixel intensities of the fluorescent spermatids and the background fluorescence in pachytene spermatocytes and Sertoli cells and cell free areas. In general, the pixel intensities of pachytene spermatocytes, Sertoli cells, step 1 $S^5G^CS^3$ and step 10 $S^5G^CS^3$spermatids were not statistically different from those of cell free areas ($P > 0.2$). The average pixel intensities of GFP-fluorescent step 3 and 4 $G^5G^CS^3$-mut2 and step 11 and 12 $S^5G^CS^3$ spermatids were ~1.5-2.5-fold greater than those of non-fluorescing pachytene spermatocytes or Sertoli cells (P values <0.0001). The initial detection of GFP expression in GFP-fluorescent step 3 $G^5G^CS^3$-mut2 and step 11 $S^5G^CS^3$ spermatids was observed in three and five independent lines and founders, respectively.

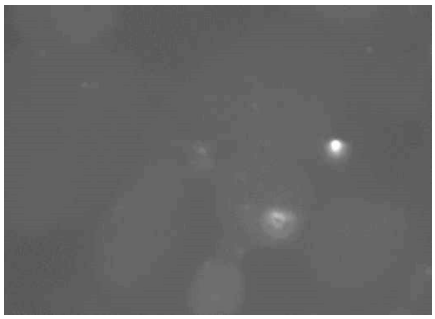
$S^5G^CS^3$ Step 10



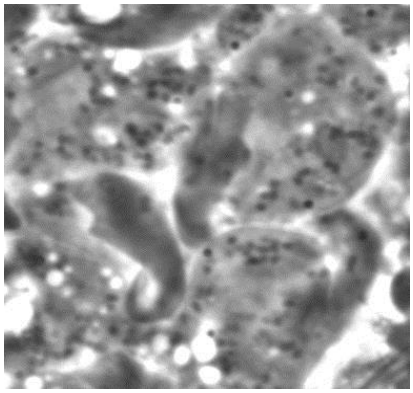


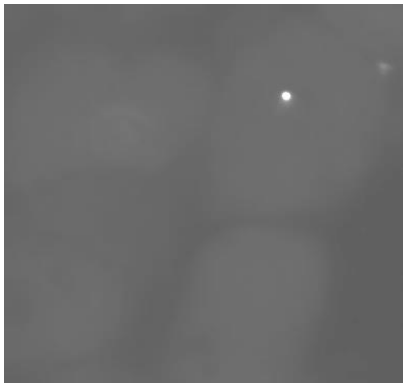
$S^5G^CS^3$ Step 11



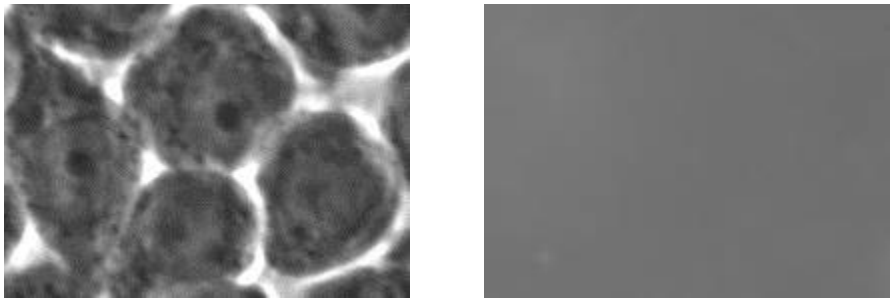


$S^5G^CS^3$ Step 12

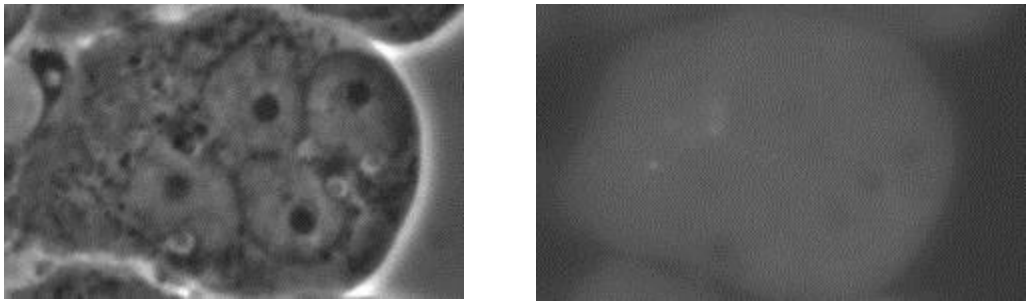




$G^5G^CS^3$-mut2 Step1



$G^5G^CS^3$-mut2 Step 3



$G^5G^CS^3$-mut2 Step 4

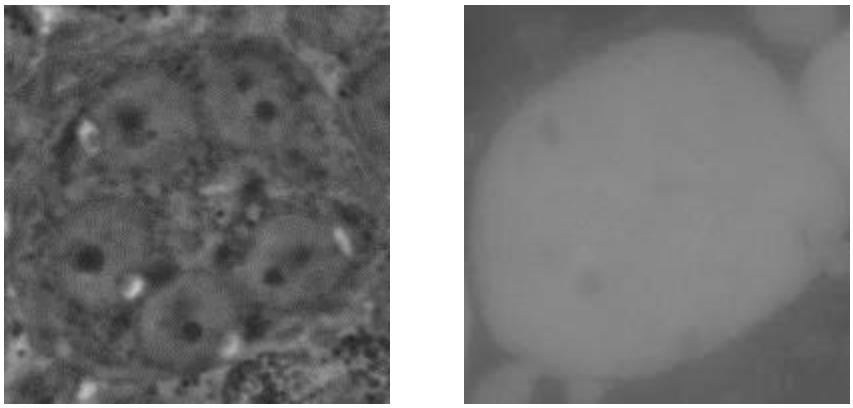


Figure 4.2 Stage of first detection of GFP fluorescence in $S^5G^CS^3$ and $G^5G^CS^3$-mut2 transgenes in round spermatids. Squashes of 0.5 mm microdissected segments of seminiferous tubules were visualized with phase contrast microscopy to identify cell types and fluorescence microscopy to detect GFP expression. The contrast and brightness were enhanced to facilitate the visualization of GFP fluorescence.

4.4.2 Sucrose and Nycodenz gradient analysis of translational activity

We also measured polysomal loading, the proportion of mRNA that is active in translation, of selected mRNAs in $S^5G^CS^3$ and $G^5G^CS^3$-mut2 mice by sedimentation on Nycodenz and sucrose gradients. Sucrose gradients separate free-mRNPs from polysomes by differences in sedimentation velocity determined primarily by the number of ribosomes associated with the coding region (Mathews et al., 2007; Arava et al., 2003; Kleene et al., 2010). Gradients containing Nycodenz, a non-ionic iodinated derivative of benzoic acid, separate free-mRNPs from polysomes by differences in buoyant density (Bagarova et al., 2010; Kleene et al., 2010). We studied the polysomal loading of the *Smcp-Gfp* $S^5G^CS^3$ mRNA, the *Smcp* mRNA and the testis specific isoform of lactate dehydrogenase C (*Ldhc)* mRNA. The *Ldhc* mRNA shows constant polysomal loading in pachytene spermatocytes, round spermatids, prepubertal and adult testes (Bagarova et al., 2010; Kleene et al., 2010), therefore it was used as an internal control for polysome integrity and RNA recovery. The *Smcp* mRNA, on the other hand, is almost absent from polysomes in the round spermatids as it is repressed in free-mRNPs, and shows modest levels in polysomes (~35%) in adult testis (Hawthorne et al., 2006; Bagarova et al., 2010). Thus, the *Smcp* mRNA serves as an example of an mRNA that is repressed in step 1-4 round spermatids and is actively translated in elongated spermatids in adult testis.

The use of both sucrose and Nycodenz gradient analysis provides a rigorous study of polysomal loading. Although sucrose and Nycodenz gradients have been used interchangeably to measure translational activity (Tafuri et al., 1993; Herbert and Hecht 1999; Kleene et al., 2010), differences in polysomal loading with the two techniques were reported previously (Bagarova et al., 2010) but were not observed here. mRNA from

cytoplasmic extracts of adult $S^5G^CS^3$ testes, 21 day *post partum* (dpp) $S^5G^CS^3$ prepubertal testes and 25 dpp $S^5G^CS^3$ prepubertal testes were quantified through phosphor-imaging of Northern blots and real time RT-qPCR following sedimentation on sucrose or Nycodenz gradients. Adult testes contain all 16 steps of spermatids, while the most advanced cells in 21 dpp and 25 dpp testes are step 4 and step 8 round spermatids, respectively, where endogenous *Smcp* mRNA is repressed in free mRNPs (Kleene et al., 2010). 25 dpp testes were only used for Northern blots since the levels of the *Smcp* and *Gfp-Smcp* mRNAs are too low to be detected in Northern blots in the pair of testes from a single 21 dpp transgenic mouse (Hawthorne et al., 2006).

In sucrose gradient analysis, *Smcp-Gfp* mRNA in extracts of 21 and 25 day testes sediment in fractions 7 through 10 which contain free-mRNP and monosomes, similar to endogenous *Smcp* mRNA (Figure 4.3 and 4.5). This result implies that the $S^5G^CS^3$ mRNA, just like *Smcp* mRNA, is repressed in round spermatids. In Nycodenz gradient analysis, the $S^5G^CS^3$ mRNA in 21/25 day testes extracts sediments with free-mRNPs in fractions 4 through 7, mimicking the polysomal profile of the *Smcp* mRNA and confirming that the $S^5G^CS^3$ mRNA is repressed in round spermatids (Figures 4.4 and 4.5). By contrast, both $S^5G^CS^3$ and *Smcp* mRNAs are bimodally distributed between free mRNPs and polysomes in adult testes in both sucrose and Nycodenz gradients (Figure 4.3, 4.4 and 4.5). We found that the mean polysomal loadings measured with Northern blots and RT-qPCR were virtually identical, therefore the results of both the assays were pooled and the mean and standard deviations are summarized in Table 4.1. The proportions of *Smcp* mRNA and $S^5G^CS^3$ mRNA associated with polysomes in 21/25 day testes and adult testes are comparable (Table 4.1). Thus, both the mRNAs undergo

similar modes of developmental regulation of mRNA translation. The *Ldhc* mRNA, on the other hand, shows constant polysomal loading in 21/25 dpp and adult testes in both sucrose and Nycodenz gradients (Table 4.1). Our polysomal loading data for *Smcp* and *Ldhc* mRNA is in agreement with previously published data (Hawthorne et al., 2006; Bagarova et al., 2010; Kleene et al., 2010).

In contrast, the $G^5G^CS^3$-mut2 mRNA exhibits high levels of polysomal mRNA in sucrose and Nycodenz gradients in 21 day testis (round spermatids) and adult testis (round and elongated spermatids) (Figure 4.6). Once again, the *Ldhc* mRNA, the internal control for integrity of polysomes in the gradients and RNA recovery during extraction exhibits high levels of polysomal mRNA in both 21 day and adult mice. In addition, the *Smcp* mRNA exhibits low levels of polysomal mRNA in 21 day testis and high levels of polysomal mRNA in adult testis.

In summary, the sucrose and Nycodenz gradient analysis support findings based on developmental GFP expression that the $S^5G^CS^3$ mRNA is translationally repressed in round spermatids and that the $G^5G^CS^3$-mut2 mRNA is translationally active in round spermatids, and that both mRNAs are translationally active in elongated spermatids.

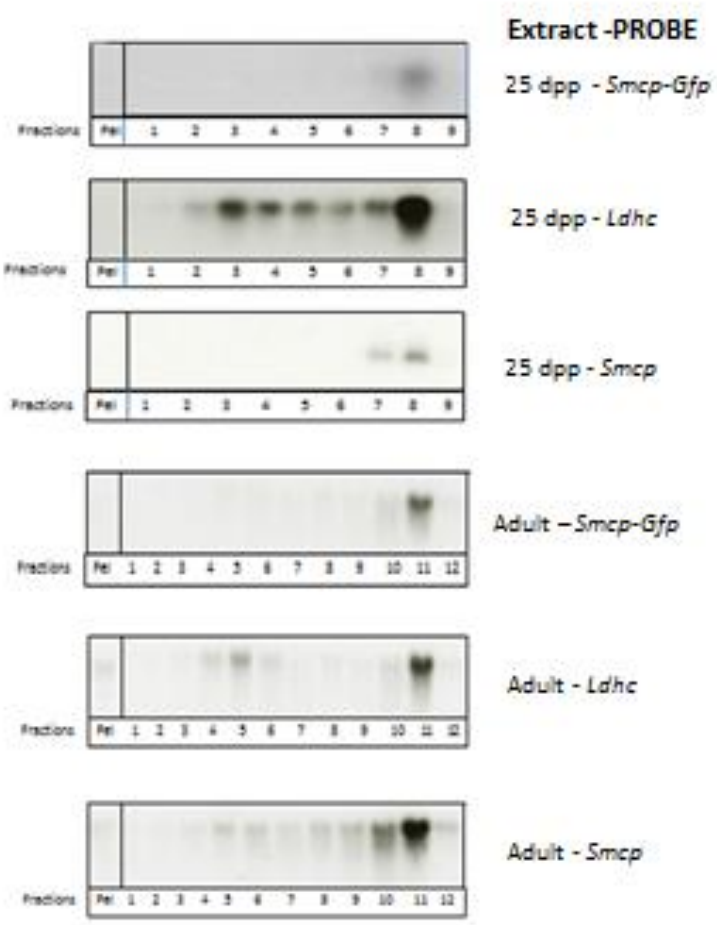


Figure 4.3 Northern blot analysis of translational activity of the $S^5G^CS^3$, *Smcp* and *Ldhc* in 25 dpp and adult testis in sucrose gradients. Cytoplasmic extracts of 25 dpp and adult testis were sedimented on sucrose gradients, fractions were collected from the bottom, RNA was extracted from each fraction, transferred to a nitrocellulose filter and hybridized to α[^{32}P]-labeled DNA probes for the *Gfp* coding region, *Smcp* coding region and *Ldhc* mRNAs. The column at right of each panel identifies the source of the cytoplasmic extract and the probe. In the 25 dpp gradient, polysomes sediment in fractions pellet and 1-6, and free-mRNPs sediment in fractions 7 through 9. In the adult gradient, polysomes sediment in the pellet and fractions 1-8, and free-mRNPs sediment fractions 9-12.

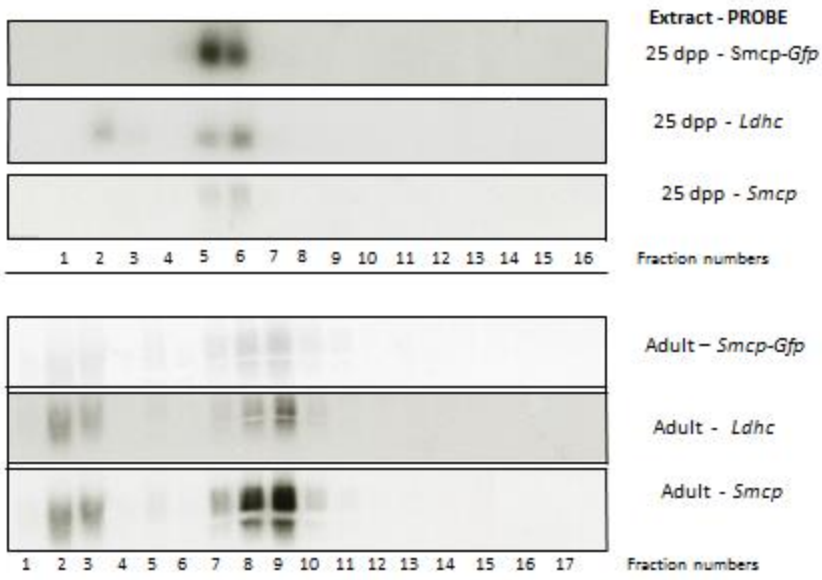


Figure 4.4 Northern blot analysis of translational activity of the $S^5G^CS^3$, *Smcp* and *Ldhc* in 25 dpp and adult testis in Nycodenz gradients. Cytoplasmic extracts of 25 dpp and adult testis were sedimented on sucrose gradients, fractions were collected from the bottom, RNA was extracted from each fraction, transferred to a nitrocellulose filter and hybridized to α[^{32}P]-labeled DNA probes for the *Gfp* coding region, *Smcp* coding region and *Ldhc* mRNAs. The column at right of each panel identifies the source of the cytoplasmic extract and the probe. In the 25 dpp gradient, polysomes sediment in fractions 1-3, and free-mRNPs sediment in fractions 5-6. In the adult gradient, polysomes sediment in fractions 1 and 3, and free-mRNPs sediment in fractions 7-10.

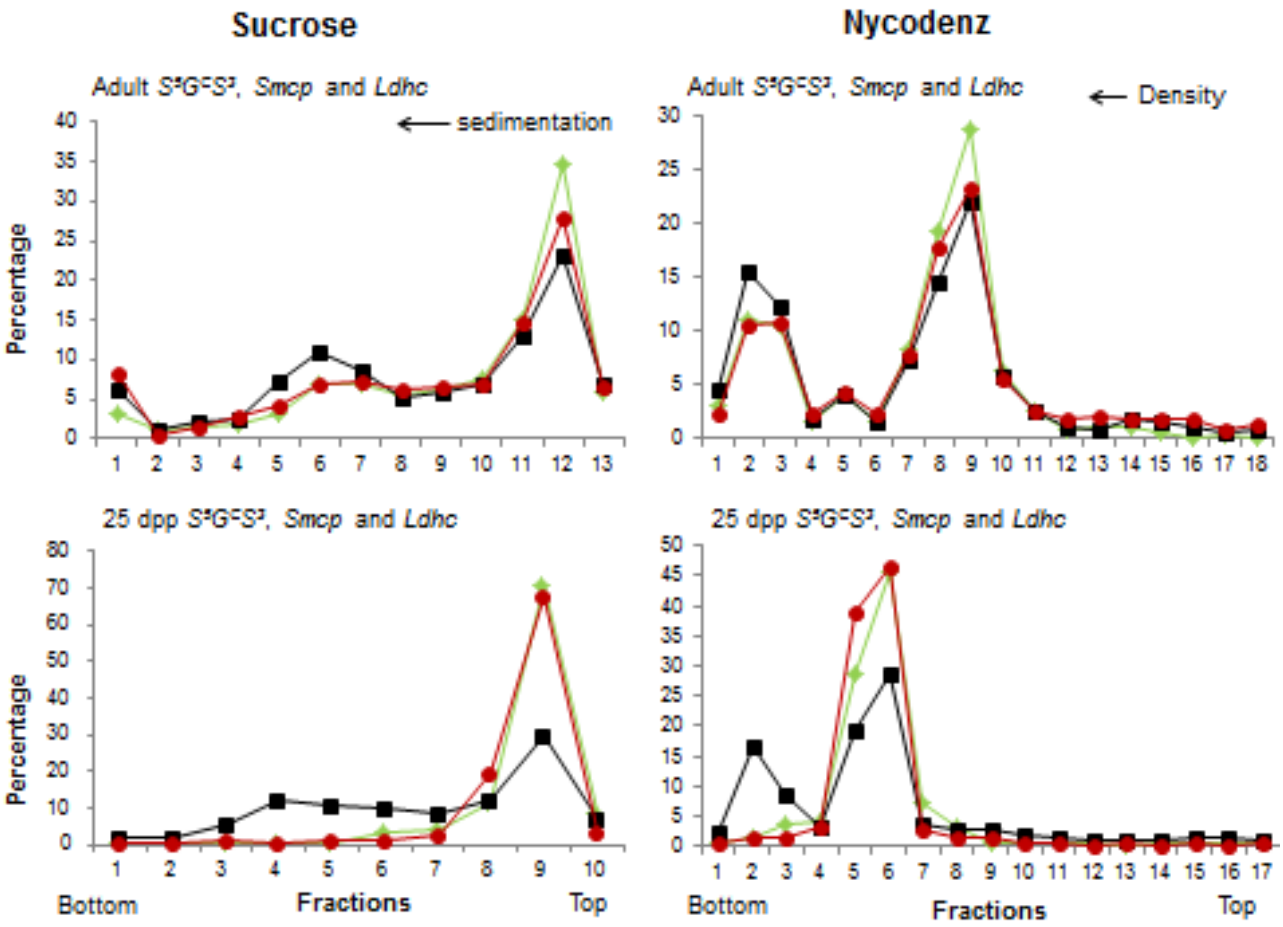


Figure 4.5 Quantitative analysis of the distribution of the $S^5G^CS^3$, *Smcp* and *Ldhc* mRNAs in the free mRNP and polysome regions of Nycodenz and sucrose gradients from 25 day old and adult $S^5G^CS^3$ transgenic mice. Cytoplasmic extracts were sedemented on Nycodenz and sucrose gradients fractions were collected from the bottom, RNAs were extracted using techniques that recover virtually identical proportions of RNA from each fraction (Kleene et al. 2010), and the amounts of specific mRNAs in each fraction were quantified by phosphorimaging of northern blots. The results are depicted as graphs of the percentage of total RNA on the gradient in each fraction. Green lines and parallelograms depict the $S^5G^CS^3$ mRNA, red circles and lines depict the *Smcp* mRNA, and black lines and squares depict the *Ldhc* mRNA.

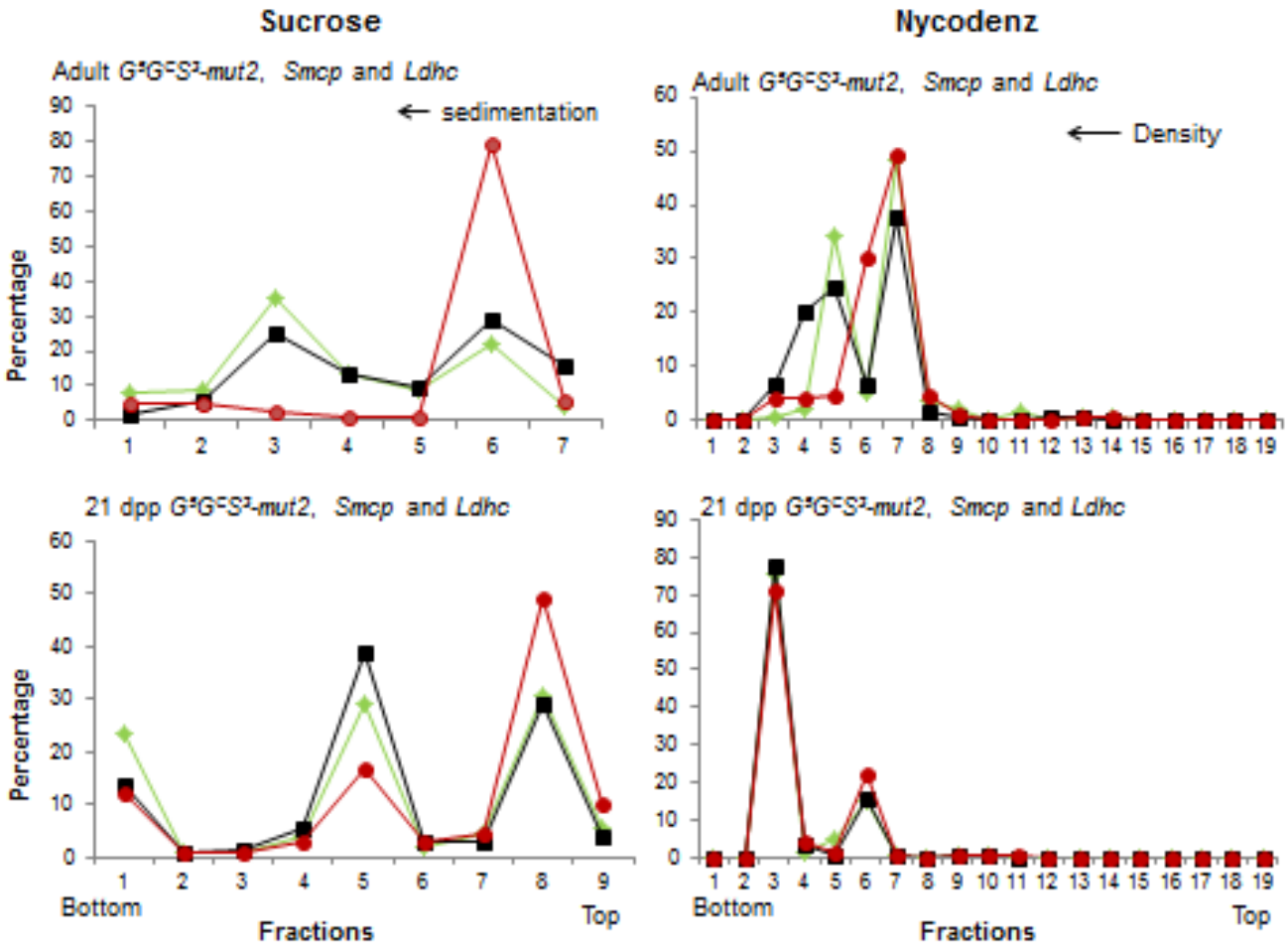


Figure 4.6 Quantitative analysis of the distribution of the $G^5G^CS^3$-mut2, *Smcp* and *Ldhc* mRNAs in the free mRNP and polysome regions of Nycodenz and sucrose gradients from 25 day old and adult $G^5G^CS^3$-mut2 transgenic mice. Cytoplasmic extracts were sedemented on Nycodenz and sucrose gradients fractions were collected from the bottom, RNAs were extracted using techniques that recover virtually identical proportions of RNA from each fraction (Kleene et al. 2010), and the amounts of specific mRNAs in each fraction were quantified by phosphorimaging of northern blots. The results are depicted as graphs of the percentage of total RNA on the gradient in each fraction. Green lines and parallelograms depict the $G^5G^CS^3$-mut2 mRNA, red circles nd lines depict the *Smcp* mRNA, and black lines and squares depict the *Ldhc* mRNA.

Table 4.1 Quantification of polysomal loading by sucrose and Nycodenz gradient analysis in $S^5G^CS^3$ mice.

Polysomal Loading (%)[a]				
	21/25 dpp testis		Adult Testis	
mRNA	Sucrose	Nycodenz	Sucrose	Nycodenz
Smcp-Gfp	6.94 ± 3.4(3)	7.82 ± 2.1(3)	41.1 ± 10.6(4)	24.7 (1)
Ldhc	52.29 ± 10.2(2)	55.02 (1)	52.5 ± 4.9 (3)	32.1 (1)
Smcp	4.5 (1)	3.33 (1)	38.2 ± 1 (2)	31.4 (1)

[a]The polysomal loading of various mRNAs in Nycodenz and sucrose gradients was quantified with phosphorimage analysis of northern blots and RT-qPCR. The polysomal loading (%) is presented as mean and S.D. with the number of independent gradients in parentheses.

4.4.3 Proteins binding to the 3'termini of the *Smcp* and $G^5G^CS^3$-mut2 3'UTRs

To identify proteins in total testis extracts that bind to the 3' terminus of the *Smcp* 3'UTR, bacteriophage RNA polymerase transcripts labeled with α[^{32}P]-rUTP were reacted with total testis extracts in the presence of RNase T1, heparin and *E.coli* tRNA to reduce non-specific binding of proteins to the probe (Walker et al., 1998), the reactions were UV-irradiated at 254 nm to covalently crosslink proteins and RNA, and the complexes were analyzed with SDS-PAGE to identify the sizes of the complexes consisting of proteins with RNase T1 fragments of the probe. Second, proteins were identified directly by RNA affinity chromatography and mass spectrometry analysis in which 5' biotinylated RNA probes were incubated with cytoplasmic protein extracts, treated with heparin to reduce non-specific binding, and protein–RNA complexes were captured with streptavidin-agarose resin. After extensive washing, the bound proteins were eluted by boiling in SDS sample buffer and resolved by SDS-PAGE. The negative

control consisted of testis extract which was incubated with streptavidin beads without biotinylated RNA, washed and eluted with the same procedure.

At the outset it should be noted that UV-crosslinking assays and RNA-affinity chromatography identify different sets of proteins. UV-crosslinking assays detect complexes with RNA fragments of probes that end with a guanosine and contain uridine, because RNase T1 digests after G-residues and the RNA probes were labeled with $\alpha[^{32}P]$-rUTP. In addition, UV-crosslinking at 254 nm preferentially co-valently crosslinks pyrimidines and aromatic amino acids (phenyalanine, tryptophan, tyrosine) and cysteine, and does not crosslink amino acids to the ribose-phosphate backbone or amino acids in one protein to amino acids in another proteins (Ule et al., 2005). Some of these proteins bind RNAs through a sequence-specific mode while others bind bases with little or no specificity. In contrast, the proteins bound to biotinylated-RNA probes in RNA-affinity chromatography potentially contain a mixture of proteins that bind RNA bases specifically and non-specifically and to the ribose-phosphate backbone and bind biotin, streptavidin and agarose non-specifically. In addition, the proteins captured with RNA affinity chromatography under non-denaturing conditions potentially include proteins that interact with RNA-binding proteins that bind the biotinylated RNA probes, but do not themselves bind the biotinylated RNA probes directly. It should also be noted that the sizes of proteins in SDS-PAGE in UV-crosslinking assays are expected to be slightly larger than those of the same proteins in affinity chromatography by virtue of covalent linkage of small RNase T1 fragments. For these reasons, the most promising proteins are those that exhibit sequence-specific binding with UV-crosslinking and sequence specific binding in RNA-affinity chromatography with similar molecular weights.

In general, we have focused our UV-crosslinking and RNA-affinity chromatography assays on the distal ends of the *Smcp* and $G^5G^CS^3$-mut2 mRNAs, because replacement of the 16 nt at the 3' terminus of the *Smcp* 3'UTR with the 17 nt at the 3' terminus of the $G^5G^CS^3$-mut2 mRNA results in loss of translational repression in round spermatids. In theory, the RNA binding assays are the first step in determining whether the abrogation of translational repression in the $G^5G^CS^3$-mut2 results from the deletion of negative translation control elements from the 3' terminus of *Smcp* 3'UTR or the insertion of previously unrecognized positive regulatory elements contained in the 3' terminus of the SV40 early polyadenylation segment (Kessler et al. 1986; Wilusz and Shenk 1988).

Figure 4.7, Panel A, Lane 1 displays the complexes with the segment of the *Smcp* 3'UTR analyzed previously that contains the AACAUCU YRS (Bagarova et al. 2010). In agreement with that study and Chowdhury and Kleene (2012) (Chapter 2), we observe the prominent complex containing YBX2, flanked by lower intensity complexes. Panel A, Lane 2, displays complexes formed with the 3' terminus of the *Smcp* 3'UTR, which displays a weaker complex of the same size as YBX2 in lane 1. Although there is no obvious degenerate YRS in this segment, the element GACAUUU, differs at one base from the degenerate YRS (underlined), and is the most promising element to which YBX2 binds. The idea that YBX2 binds this element was confirmed with a two base mutation, GCUAUUU, which eliminates binding (Panel B, lane 2). Panel A, lane 3, displays complexes formed with the 3' terminus of the $G^5G^CS^3$-mut2 3'UTR. In contrast to our expectation that this probe would not bind Y-box proteins, we observed complexes of the same size as YBX2 and other Y-box proteins and a series of complexes with

greater and lower mobility. Note that the lanes which contain AAUAAA poly(A) signals, Panel A, lanes 2 and 3, and Panel B, lanes 1 And 2, display no trace of the 160 kDa subunit of CPSF which binds AAUAAA (Murthy and Manley, 1995). A striking observation is that these short probes bind multiple proteins.

Figure 4.8, lane 1 shows proteins that bound 5'-biotinylated *Smcp* YRS probe, AAGGAUAGAAACAUCUUGUCUAGUGAUCCUG (YRS underlined), in RNA affinity chromatography. Bands marked as SY-1 and SY-2 were of similar sizes as the complexes observed in UV-crosslinking study (Figure 4.7, lane 1). Mass spectrometry sequencing identified the proteins in SY-1 to be YBX2 and YBX3L. However, spectral counts show that YBX2 is more abundant than YBX3L. Additionally, comparisons of the average precursor intensity of the 5 most abundant peptides show that YBX2 is 10 fold more abundant than YBX3L. SY-2 was identified as YBX1.

Figure 4.8, lane 2 shows proteins that interact with the 3' terminus of the $G^5G^CS^3$-mut2 3'UTR. Mass spectrometry sequencing identified several proteins in G-1 and G-2. However, comparisons of the average precursor intensity of the 5 mos abundant peptide shows that Heat shock protein 1A (HSP1A) and Far upstream binding protein (Fubp1) are the most abundant proteins in G-1 and HnrnpD (aka AUF-1) is the most abundant protein in G-2. We are currently in the process of establishing the identity of all the other proteins through mass spectrometry sequencing. RNA affinity chromatography shows that several proteins interact with the 3'terminus of the *Smcp* 3'UTR (figure 4.8, lane 4). The most important of these proteins is S3E4 because a protein of similar size was observed in UV-cross-linking study as well (figure 4.7, lane 2). Mass spectrometry sequencing identified the protein to be YBX2. Mass spectrometry sequencing identified

S3E1 to be Fubp1, S3E2 to be Calcium Response Factor (CRF), S3E5 to be Tubulin 2α and S3E6 to be HnrnpD/AUF1. Figure 4.8, lane 3 and lane 5 shows the negative control: testis cytoplasmic extract incubated with streptavidin agarose resin. Mass spectrometry sequencing identified C-1 as Tubulin – β4 and C-2 and C-3 as EEF1A1.

The information from mass spectrometry sequencing of RNA affinity chromatography purified and SDS-PAGE fractionated proteins is compiled in Table 4.2. All of these proteins identified with high confidence based on multiple spectra and perfect matches with proteins in the databases. The beads only control consistently recovered two bands of about 46 and 55 kDa. The predominant 46 kDa band is translation elongation factor eEF1A1, while the upper 55 kDa band contains beta tubulins. Notably, it appears that the levels of the tubulins can be reduced with extensive washing. Tubulins were often present at lower levels in the bands containing proteins purified by RNA affinity chromatography, presumably as non-specific contaminants.

The proteins purified with RNA affinity chromatography using RNA probes contained a mixture of proteins with diverse functions. The most interesting of these here are RNA binding proteins. Three prominent bands migrating at about were 52 kDa analyzed by mass spectrometry sequencing, SY1, Prm1-TCE (chapter 3), and S3E4. In all three bands, YBX2 was the predominant protein, at least 11-14 fold more abundant than YBX3L. The very prominent band purified with the 3' terminus of the $G^5G^CS^3$-mut2 3' UTR containing the early SV40 polyadenylation segment has not been analyzed with mass spectrometry sequencing but probably also contains high levels of YBX2.

Other RNA binding proteins were purified in great quantities, far upstream RNA element binding protein1, NP_476513.2 (642 AAs), in G1 and S3E1 fractions. As far as

we are aware, this is the first detection of FUBP1 in testis. In somatic and malignant cells, FUB1 binds both single stranded DNA and RNA and promotes malignancy, cell migration and inhibits apoptosis. At the molecular level, FUBP1 can inhibit mRNA translation and promote mRNA degradation. FUBP1 has been reported to bind AU-rich mRNA elements but the RNA-binding specificity has not been characterized carefully. The functions and properties of FUBP1 have been reviewed in Zhang and Chen (2013).

The other abundant protein in G2 and S3E6 is HnrnpD, popularly known as AUF1, is a well-known single strand RNA binding protein that also interacts with AU-rich elements and normally promotes mRNA degradation and represses mRNA translation (reviewed in White et al., 2013).

The protein named RNA binding protein 14, abundant in band S3E1, is a member of the huge family of RNA-recognition motif RNA binding proteins that bind specific sequences of bases in single stranded RNA. Virtually nothing is known about its patterns of expression, function or binding specificity.

The absence of several proteins among the RNA binding proteins mentioned above is notable. Cytoplasmic poly(A) binding protein, a highly abundant protein that binds the poly(A) tail, oligo(A) sequences in the UTRs of some mRNAs, and a poorly defined element that is not A-rich was not one of the abundant proteins in any fraction. RNA binding proteins that bind AAUAAA polyadenylation signals are of interest because YRSs to which YBX2 binds appear to be just a few nucleotides apart. However, the 160 kDa subunit of CPSF was not detected, nor was tristetraprolin (NP_035886.1) , another AU-rich element that specifically binds AAUAAA detected (Emmons et al., 2008). In addition, the sequence of seven contiguous Us in the SV40 early

polyadenylation unit at the 3' terminus of the $G^5G^CS^3$-mut 2 3'UTR is an obvious target to which the AU-rich translational promoter binds, ELAV1/HuR. ELAV1/HuR is thought to be an important regulator of mRNA translation in spermatogenic cells (Chi et al. 2011), but was only detected as a minor component of the G2-fraction. Conceivably, HnrpD and FUBP1 bind oligoU or AAUAAA sequences, but I am unaware of evidence supporting these ideas.

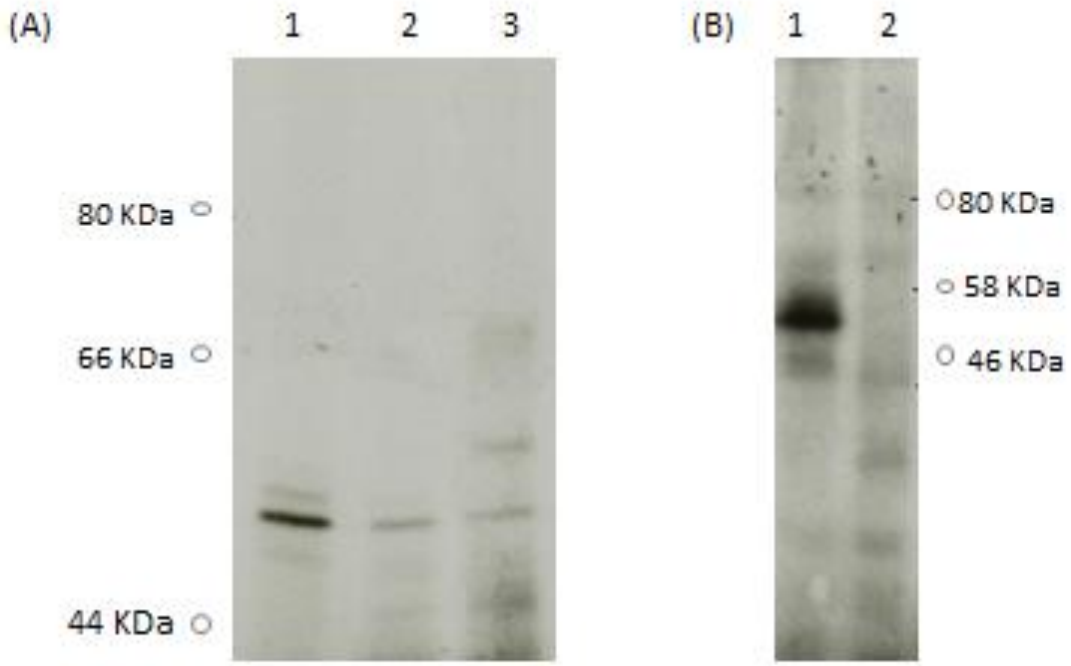


Figure 4.7 UV-crosslinking analysis of *Smcp* 3'UTR and $G^5G^CS^3$-mut2 3'UTR. α[^{32}P]-rUTP-labeled RNA probes were reacted with total testis cytoplasmic extract and sequence specific complexes were formed through the sequential treatment with *E.coli* tRNA, RNaseT1 and heparin. The complexes were UV crosslinked, resolved by SDS-PAGE and visualized by autoradiography. Canonical AAUAAA polyadenylation elements are highlighted grey, high affinity and low affinity YRSs are single underlined, and mutations are double underlined. (A) 7.5 % SDS-PAGE. Lane 1, *Smcp* 3'UTR YRS, GAAGGUAGAAAAGGAUAGAAACAUCUUGUCUAGUGAUCCUGACAUUUAGAU. Lane 2, *Smcp* 3'UTR 3' terminus, UGUCUAGUGAUCCUGACAUUUAGAUAGCAAAGAAAUAAAAGAGCAAAUAAAAAG. Lane 3, 3' terminus of $G^5G^CS^3$-mut2 3'UTR, GACAUUUAGAUAGCAAAGAAAUAAAAGCAUUUUUUUUCACUGC.

(B) 10% SDS-PAGE. Lane 1, *Smcp* 3'UTR 3' terminus, UGUCUAGUGAUCCUGACAUUUAGAUAGCAAAGAAAUAAAAGAGCAAAUAAAAAG.
Lane 2, *Smcp* 3'UTR 3' terminus with mutations, UGUCUAGUGAUCCUGCUAUUUAGAUAGCAAAGAAAUAAAAGAGCAAAUAAAAAG

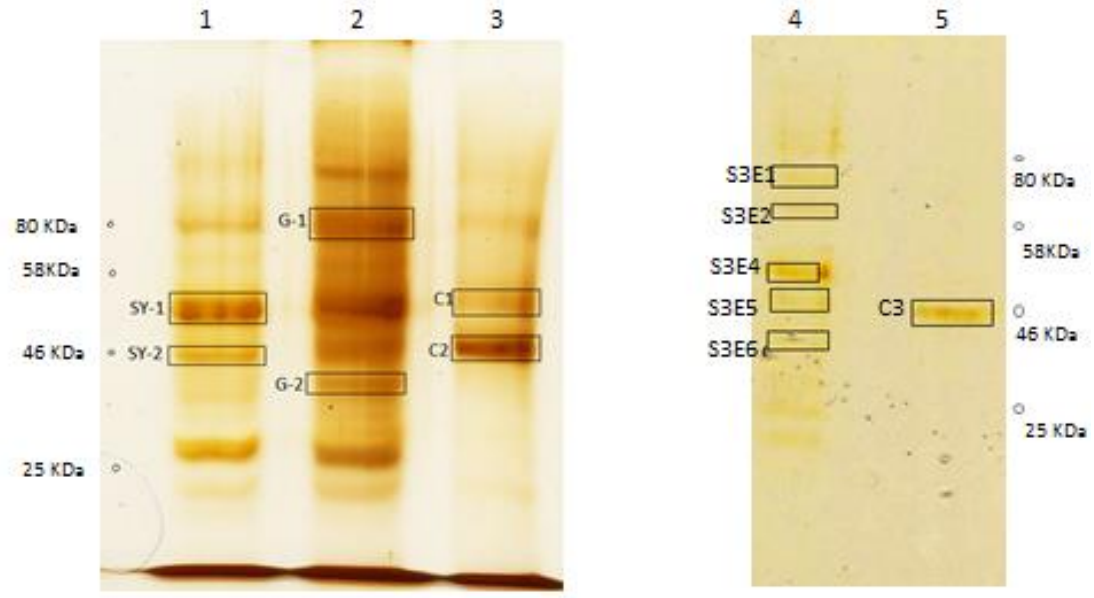


Figure 4.8 Identification of proteins binding to *Smcp* and $G^5G^CS^3$-mut2 3'UTRs with RNA affinity chromatography and mass spectrometry sequencing. Total testis cytoplasmic extract was reacted with 5'-biotinylated RNA probes with heparin, bound to streptavidin-agarose and eluted with SDS sample buffer. Proteins were resolved by SDS-PAGE and visualized with silver staining. Bands marked with boxes were excised and analyzed by Mass spectrometry sequencing. Lane 1, Smcp 3'UTR YRS, AAGGAUAGAAACAUCUUGUCUAGUGAUCCUG. Lane 2, 3' terminus of $G^5G^CS^3$-mut2 3'UTR, GACAUUUAGAUAGCAAAGAAAUAAAAGCAUUUUUUUUCACUGC. Lane 3, cytoplasmic extract incubated with streptavidin-agarose resin as a negative control. Lane 4, 3' terminus of *Smcp* 3'UTR, GUGAUCCUGACAUUUAGAUAGCAAAGAAAUAAAAGAGCAAAUAAAAAG. Lane 5, cytoplasmic extract incubated with streptavidin-agarose resin as a negative control.

PROTEIN	IDENTIFIED IN FRACTION[2]	NUMBER AA[3]	ACC. NUMBER[4]
RNA binding proteins			
Y-box protein 2(YBX2)	SY1**, S3E4**, PRM1-TCE**	359 AA	NP_058571.2
Y-box protein 3 long isoform (YBX3L)	SY1, G1, PRM1-TCE	361 AA	NP_620817.2
Y-box protein 1 (YBX1)	SY2	322 AA	NP_035862.2
Far upstream element-binding protein 1 (FUBP1)	G1*, S3E1**	642 AA	NP_476513.2
Heterogeneous nuclear RNP D0 isoform b (HNRNPD/AUF1)	G2**, S3E6*	336 AA	NP_001070734.1
Heterogeneous nuclear RNP D0 isoform b (HNRNPD/AUF1)	G2**, S3E6*	351 AA	NP_001070733.1
HNRNPL	S3E2	586 AA	NP_796275.3
HNRNPM	S3E2	729 AA	NP_084080.1
HNRNPM	S3E2	690 AA	NP_001103383.1
FUBP2	G1	748 AA	NP_034743.3
HNRNPA2/B1	G2		
HNRNPC	G2		
ELAV1/HuR	G2		
DDX17	S3E1		
HNRNPL	S3E1		
HNRNPM	G1, S3E1		
HNRNPK	PRM1-TCE		
PABPC1	G1, S3E2		
RNA BINDING MOTIF 46	S3E4		
RNA BINDING PROTEIN 14	S3E1**	669 AA	NP_063922.2
RNA BINDING MOTIF 17	S3E5		
RNA 3' TERMINAL CYCLASE	S3E6		
DAZAP1	S3E6		
CELF1	PRM1-TCE		
RNA BINDING MOTIF SS INTERACTING PROTEIN 2	S3E6		

Heat shock proteins			
Heat shock cognate 71 kDa protein (HS71A)	G1**	646 AA	NP_112442.2
HSPA2	G1*	633 AA	NP_001002012.1
HSPA5	G1, S3E1		
HSPA8	G1		
HEAT SHOCK PROTEIN 7C	S3E1		
HEAT SHOCK PROTEIN 5	S3E1		
Cytoskeletal proteins			
Tubulin-beta 4B	C1**	445 AA	NP_666228.1
TUBULIN B2	S3E2, S3E4, S3E5		
TUBULIN B2A	S3E4, S3E5		
TUBULIN B3	C1* ,S3E2, S3E4, PRM1-TCE		
TUBULIN A1	S3E4 , PRM1-TCE		
Tubulin-beta 2A	C1*, S3E4		
TUBULIN B1	PRM1-TCE		
ACTIN 1	G2		
ACTIN A2	S3E2		
ACTIN B	G2*		

Other proteins			
ALBUMIN	G1		
Eukaryote elongation factor 1A(EEF1A)	G2*, C2**, C3**	462 AA	P_034236.2
STRAP kinase receptor	G2		
CYTOCHROME 17A	C1, PRM1-TCE		
IMMUNITY GTPASE	C1		
INTERFERON GTPASE	C1, S3E4		
G-KINASE ANCHOR 1	C2		
RUV-LIKE AAA/ATPASE	C2, S3E5, C3		
N-GLYCANASE 1	S3E1		
CLATHRIN INTERACTOR 1	S3E1		
CALCIUM RESPONSE FACTOR	S3E2		
PARASCPECKLE COMPONENT 1	S3E2		
MYELIN EXPRESS FACTOR	S3E2		
CHAPERONIN TCP1	S3E2		
EH DOMAIN	S3E2		
CDK REGULATORY SUBUNIT	SE32		
ATP SYNTHETASE SUBUNIT B	S3E5		
ATP SYNTHETASE SUBUNIT A	S3E6		
POLYMERASE DELTA INTERACT	S3E5, C3		
CCR4-NOT TRANSCRIPTION	S3E5		
ADENYLATE KINASE 8	S3E5		

Table 4.2 Proteins identified with mass spectrometry sequencing of SDS-PAGE bands in affinity chromatography with streptavidin beads and biotinylated RNA probes.

[1] Name of protein derived from NCBI protein accession number. The proteins are divided into several groups according to function.

[2] Fractions are named according to RNA probe used to capture proteins and SDS-PAGE band that was excised for mass spectrometry sequencing. The asterisks denote the relative abundance of the identified protein in the designated band deduced from precursor intensity: **, highly abundant; *, moderately abundant; no asterisk, low abundance.

[3] Length of protein in number of amino acids derived from accession number.

[4] Reference sequence accession number.

4.5 Discussion

Although many mRNAs are thought to undergo delayed translational activation in spermatids, only two mRNAs have been studied with multiple point and deletion mutations in transgenic mice to identify the cis-elements that mediate the initial translational repression in round spermatids, the *Prm1* and *Smcp* mRNAs. However, the mechanisms of developmental regulation of the *Prm1* and *Smcp* mRNAs exhibit clear differences. Delayed translational expression of the *Prm1* mRNA from step 7 to step 10 spermatids has been reported to be mediated by a 17 nt 3'UTR translational control element (TCE) which has been described as "necessary and sufficient" for complete translational repression in round spermatids (Zhong et al., 2001). In contrast, as implied previously (Hawthorne et al., 2006; Bagarova et al., 2010) and confirmed here with the $S^5G^CS^3$ transgene, delayed translational expression of the *Smcp* mRNA from step 3 to step 11 spermatids requires both the 5' UTRs and 3' UTRs. Repression by the *Smcp* 5'UTR is short, steps 3 and 4 (1-2 days), and is mediated primarily by upstream reading frames (uORFs) which produce small polysomes instead of free-mRNPs (Bagarova et al., 2010). In contrast, repression by the *Smcp* 3'UTR is longer, steps 3 through 8 (~5 days), and is mediated by elements which produce free-mRNPs (Bagarova et al., 2010). However, the *Smcp* 3'UTR also contains positive control elements which neutralize repression by the uORFs in elongated spermatids (Bagarova et al., 2010), and we cannot address the possibility that interactions of the *Smcp* 3'UTR and 5'UTR affect other forms of positive and negative translational regulation. The *Smcp* mRNA model involving multiple mechanisms of translational repression may be more typical than the single element *Prm1* mRNA model, because post-transcriptional regulation of individual

mRNAs in somatic mammalian cells is thought to be regulated by multiple cis-elements, RNA binding proteins and miRNAs (Morris and Keene 2010; Kishore et al., 2010). In addition, the contention that the *Prm1* TCE is sufficient for translational repression in round spermatids may be incorrect because translational regulation by the *Prm1* 3'UTR has never been analyzed in the absence of the *Prm1* 5'UTR and the pathway by which the TCE represses *Prm1* mRNA translation may require other cis-elements (Braun et al., 1989; Braun 1990; Fajardo et al., 1997; Giorgini et al., 2001; Zhong et al., 2001).

The most significant question is the identity of the elements and factors in *Smcp* 3'UTR that repress translation in steps 3 through 8. This is a difficult problem to resolve in a system such as spermatogenesis in which transgenic mice offer the only experimental system for analyzing mutations in cis-elements. However, transgenic mice are an absolute necessity in view of evidence for spermatid-specific mechanisms of translational regulation mentioned in the Introduction, and evidence discussed below that translational repression by Y-box proteins requires proximity to the 3' end of the 3'UTR (Chapter 3; Zhong et al., 2001; Soundajaram et al., 2010). Since the analysis of even one transgene is an expensive, prolonged and laborious process, transgenes have to be designed carefully to avoid squandering precious resources. The transgenes analyzed to date were based on the evidence that *Prm1* negative control elements that repress translation in early spermatids only function when the elements are at the 3' terminus of the 3'UTR (Fajardo et al., 1997; Giorgini et al., 2001; Zhong et al., 2001, Soundararajan et al., 2010), and the distal end of the *Smcp* 3'UTR is highly conserved (Kleene and Bagarova, 2006; Chowdhury and Kleene, 2012).

Our first transgene, designed to identify elements in the *Smcp* 3' UTR that repress translation in early spermatids, contained a randomized 39 segment 28-61 nt upstream of the poly(A) site, ($G^5G^CS^3$-mut1), a position similar to those of the two elements in the *Prm1* 3'UTR that repress translation in early spermatids (Background). This mutation produced a partial loss of translational repression: GFP expression is detected in step 4 spermatids (Figure 4.2), and the level of polysomal $G^5G^CS^3$-mut1 mRNA in sucrose and Nycodenz gradients in 21 day testes, ~10%, is intermediate between those of the repressed *Smcp* mRNA, ~4.5%, and the translationally active $S^5G^CG^3$-no-uORF1&2 and $S^5G^CG^3$-no-uORF1 mRNAs, ~31%.

We next studied a transgene, $G^5G^CS^3$-mut2, in which the 16 nt downstream of the first AAUAAA poly(A) signal is replaced by the 17 nt downstream of the pEGFP poly(A) signal (Figure 4.1). The assumption that this segment of the *Smcp* 3'UTR contains regulatory elements is unorthodox, because the cis-elements in the short, 15-30 nt, segments of 3' UTRs between the poly(A) signals and the polyadenylation sites (Tian et al. 2005), rarely contain regulatory elements. However, the segment of the *Smcp* 3' UTR downstream of the upstream poly(A) signal contains two of the most conserved sequences in the *Smcp* 3'UTR, a second AAUAAA poly(A) signal and GAGC (Chowdhury and Kleene, 2012), and is consistent with evidence that translational repression of the *Prm1-hGH* transgenes by the YRS and TCE require position near the 3' terminus of the *Prm1* 3'UTR (Fajaro et al., 1997; Giorgini et al., 2001; Soundajaran et al., 2010). In addition, the requirement for a specific position strongly suggests that the *Prm1* TCE is not sufficient for translational repression and requires proximity to additional *cis*-elements.

GFP is first detected in step 3 spermatids in three independent lines suggesting that the $G^5G^CS^3$-mut2 completely abolishes the delay in GFP expression (Figure 4.2), and this conclusion is supported by high levels of polysomal mRNA in sucrose and Nycodenz gradient analysis of 21 dpp testes. This transgene was designed on the assumption that the 17 nt derived from the early SV40 polyadenylation signal and poly(A) site in the eGFP plasmid lacks regulatory elements. This assumption is supported by exhaustive literature searches that the early SV40 polyadenylation signal binds the 160 kDa subunit of the cleavage and polyadenylation stimulation factor, CPSF160, and no other proteins (Murthy and Manley, 1995; Chao et al., 1999), supporting our assumption that replacing the segment of the *Smcp* 3'UTR downstream of the first poly(A) signal with the early SV40 polyadenylation segment was only removing elements in the Smcp 3' UTR. Unexpectedly, we found that the early SV40 polyadenylation unit binds several proteins in testis extracts, some of which could function as translational activators. Thus, abrogation of translational repression of the $G^5G^CS^3$-mut2 mRNA in 21 dpp testis may be because the early SV40 polyadenylation segment binds translational activators as well as translational repressors. However, the possibility that the 16 nt at the distal end of the *Smcp* 3'UTR contain negative translational control elements cannot be dismissed.

CHAPTER 5

SUMMARY AND PERSPECTIVES

Many mRNAs, such as *protamine 1* (*Prm1*) and s*perm mitochondria associated cysteine-rich protein* (*Smcp*) mRNAs, are transcribed in round spermatids, and stored as translationally inactive messenger ribonucleoprotein particles (free-mRNPs) for several days to a week before the mRNA is translated in transcriptionally inactive elongating and elongated spermatids (reviewed in Kleene, 2003, 2013).The initial translational repression is critical for normal sperm development since premature activation of translation in transgenic mice, leads to deformed spermatozoa and male subfertility or infertility (Lee et al., 1995; Tseden et al., 2007).

The mechanisms that regulate mRNA translation in spermatids are poorly understood, although several observations suggest that the mechanisms are novel and therefore especially interesting. For example, spermatids are the only system where translationally inactive mRNAs have long poly(A) tail and active mRNAs have short poly(A) tails (Kleene 1989; Jackson et al., 2009). In addition, spermatids are the only system in which excess cytoplasmic poly(A) binding protein PABPC1 functions as a translational repressor instead of a translational activator (Yanagiya et al., 2010).

Generally, mRNA- specific translational regulation involves *cis*-elements which bind *trans*-factors, RNA binding proteins (RBPs) and/or small non-coding RNAs, which either promote or block translation (reviewed in Kleene, 2013). In addition, translation of individual mRNAs may be regulated cooperatively by multiple RNA binding proteins and/or small non-coding RNA (reviewed in Kleene 2013). Gene knock-outs have implicated more than 20 RNA binding proteins in negative and positive translational regulation in spermatogenic cells (Paronetto & Sette, 2010; Idler and Yen, 2010; Nguyen Chi and Morello 2011). However, the mechanisms by which these factors affect translation and cause defects in spermatogenesis remain unclear, since it has not been established whether any of these knockouts have direct or indirect effects on translation of specific mRNA (reviewed in Kleene, 2013).To elaborate, RNA binding proteins are expressed for long periods during spermatogenesis and potentially affect many downstream targets (Kleene 2013).Some of these primary targets likely encode regulatory factors which affect other secondary targets. Indeed, the RNA binding protein, ELAV1/HuR, has been described as a "regulator of regulators" (Pullman et al. 2007; Mukherjee et al., 2012). Consequently, gene knockouts are incapable of pinpointing whether a factor affects translation of a mRNA by binding it directly and the magnitude of the effects on the timing and level of translational activity. The importance and challenges of defining the functions of *cis*-elements that repress translation have been cogently documented by Farley and Ryder (2012).

Few *cis*-elements that control mRNA translation have been identified in mammalian spermatogenic cells. This is because the most straight forward method of identifying *cis*-elements is by studying mutations in transgenic mice, since rapid methods

using cell culture or cell-free translation systems have not been established for spermatogenic cells and the atypical regulatory mechanisms mentioned above invalid using systems based on somatic cells (Hunter et al., 2012; Kleene, 2013). Unfortunately, the analysis of mutations in *cis*-element in transgenic mice is considered (with good reason) to be too expensive, slow, and risky by many investigators. So far, the *Prm1* and *Smcp* mRNA are the only mRNAs to be studied through multiple mutations in transgenic mice. The goal of my thesis research is to identify factors and elements that repress the translation of *Prm1* and *Smcp* mRNAs in round spermatids and activate their translation in late spermatids.

The *Prm1* mRNA is transcribed in step 7 round spermatids and is repressed in free-mRNPs until translation is activated in step 10 elongating spermatids (Kleene 1989; Mali et al., 1989; Meistrich et al., 2003). Studies in transgenic mice have concluded that the 3'UTR is "necessary and sufficient" for the initial translational repression of the *Prm1* mRNA (Braun et al. 1989). Subsequent studies of deletion and point mutations in transgenic mice identify a 17-nucleotide long *Translational Control Element* (TCE), GAACAAUGCCACCUGUC, within the *Prm1* 3'UTR as the critical element for translational delay (Braun 1990; Zhong et al 2001; Giorgini et al 2001). An RNA-binding protein that binds the *Prm1* TCE has not been identified after many years of research (Fajardo et al., 1994), which not surprisingly has led to speculation that the *Prm1* mRNA is repressed by microRNAs (Papaioannou and Nef, 2010).

The idea that Y-box proteins repress *Prm1* mRNA translation emerged unexpectedly from experiments described in Chapter 2 (Chowdhury and Kleene 2012) which refine the *Y-box recognition sequence* (YRS) to which Y-box proteins bind. Two

Y-box proteins, YBX2 and YBX3L, have been proposed to play a prominent role in repression of the *Prm1* mRNA because the levels of YBX2 and YBX3L – high in round spermatids and undetectable in elongated spermatids – correlate with the periods of translational repression and active translation of the *Prm1*, *Smcp,* and many other mRNAs (Oko et al., 1994; Davies et al., 2000; Giorgini et al., 2001).

In order to understand the role of YBX2 in mRNA specific translational repression, it is crucial to be able to identify the elements (YRS) and mRNA targets to which YBX2 and YBX3L bind specifically. Giorgini et al. (2001) analyzed the YRS in the *Prm1* 3'UTR, UCCAUCA, which binds Y-box proteins, YBX2 and YBX3L, through single base mutations with RNA EMSAs and competition assays. Their analysis defines the YRS as a C-rich 7 nt element which lacks G, [ACU][CA]<u>CA</u>[UC]<u>C</u>[ACU]. The nucleotides in brackets are alternative bases that exhibit maximal binding and underlined nucleotides are critical for binding. Chapter 2 (Chowdhury and Kleene, 2012) analyzed the 29 YRSs predicted by Giorgini et al. (2001) that hadn't been analyzed before through UV crosslinking assay. This chapter specifically examined the YRSs that deviate at 2–4 bases from UCCAUCA. I find that most of these 29 YRSs bind Y-box proteins bind strongly, except those YRSs that begin and end with triplets of Cs. Thus, the YRS is a degenerate element which explains differences in YRS sequences in different studies (Bouvet et al., 1993; Giorgini et al., 2001; Wei et al., 2012, and references cited therein).

The degeneracy of YRS was a critical insight that led to identification of YBX2 as the RNA-binding protein that binds the TCE in Chapter 3 (Chowdhury and Kleene, in revision). Giorgini et al. (2001) also carried out parallel studies in yeast three hybrid assay which show that YBX2 and YBX3L bind G in the first position, which defines the

YRS as [ACGU][CA]CA[UC]C[ACU]. I hypothesized that the *Prm1* TCE contains a YRS, underlined, GAACAAUGCCACCUGUC, which is compatible with my conception of the degenerate YRS (Chapter 2, Chowdhury and Kleene 2012) and the yeast three hybrid studies of Giorgini et al. (2001). I also hypothesized that the failure to detect binding of Y-box proteins to the *Prm1* TCE arises from the use of RNaseT1 which digests after G-residues, because rabbit Y-box protein 1 (YB-1) greatly increases RNaseT1 degradation of rabbit β-globin mRNA (Evdokimova et al 1995). To test my hypotheses I used a probe for the *Prm1* TCE in UV crosslinking assays. Use of RNaseT1 before UV crosslinking abrogated complex formation, but strong complexes were noticed when the reactions were UV crosslinked before RNaseT1 digestion (Chapter 3). To confirm that the complexes obtained with the UV crosslinking assay are indeed Y-box proteins, I carried out RNA affinity chromatography followed by mass spectrometry sequencing. Mass spectrometry sequencing shows that both YBX2 and YBX3L bind the *Prm1* TCE, but unexpectedly YBX2 is ~13 fold more abundant than YBX3L (Chapter 3). Thus, YBX2 is the major protein that binds the TCE that has been described as being is "necessary and sufficient" for the repression of *Prm1* mRNA until step 10 spermatids (Zhong et al., 2001). The inference that YBX2, and not YBX3L, is critical for *Prm1* translational repression is consistent with reports that the knockout of the *Ybx2*-gene results in defects in the differentiation of elongating and elongated spermatids (Yang et al., 2007), while the knockout of the *Ybx3* gene results in defects in spermatogenic cell renewal, but not defects in spermatid differentiation (Lu et al., 2006). These findings also imply that *Prm1* mRNA repression is established by binding of YBX2 to a regulatory

element in the mature, cytoplasmic mRNA, and not by transcription in the nucleus (Yang et al., 2005).

The *Smcp* mRNA is transcribed in step 3 spermatids and SMCP protein is first detected in step 11 spermatids (Shih and Kleene 1991; Cataldo et al., 1996). However, unlike *Prm1* mRNA, translational regulation of the *Smcp* mRNA is controlled by both the 5'UTR and the 3'UTR. Studies in transgenic mice show that *Smcp* 5'UTR alone delays translation until step 5 and that the *Smcp* 3'UTR alone delays translation until step 9 (Hawthorne et al 2006; Bagarova et al 2010), suggesting both UTRs are necessary to delay activation of the natural *Smcp* mRNA translation until step 11. My studies of a chimeric transgenic construct containing the *Smcp* 5'UTR, *Gfp* coding region and *Smcp* 3'UTR ($S^5G^CS^3$), confirm that both the *Smcp* 5'UTR and *Smcp* 3'UTR are required for proper developmental repression until step 11 (Chapter 4, Chowdhury et al., in preparation). The $S^5G^CS^3$ transgene also shows that the *Smcp* 3'UTR counteracts repression by uORFs in the 5'UTR, but the mechanism is unknown. Perhaps, the *Smcp* 3'UTR promotes more efficient reinitiation of ribosomes at the *Smcp* initiation codon as seen in the case of *Her2* mRNA (Mehta et al., 2006). However, I have not rigorously proven that the uORFs are responsible for the delayed activation of translation of the $S^5G^CS^3$ mRNA to step 11. The alternative possibility exists that interactions of the *Smcp* 5'UTR and 3'UTR are required to repress translation of the *Smcp* mRNA in free-mRNPs until step 11. At this point, it should also be noted that the assertion that the *Prm1* 3'UTR is "necessary and sufficient" for translational repression in round spermatids is unfounded because all of the *Prm1* transgenes that have been studied contain both the *Prm1* 5' UTR and the *Prm1* 3' UTR (Braun et al. 1989; Braun 1990; Fajardo et al., 1997;

Zhong et al. 2001). Thus, the possibility cannot be ruled out that repression by the *Prm1* TCE requires *cis*-elements in the *Prm1* 5'UTR that bind factors that interact with YBX2 bound to the TCE YRS. Conceivably, this hypothetical factor is YBX2 because the *Prm1* 5'UTR contains a highly conserved YRS (Chapter 2, Chowdhury and Kleene).

Despite the differences mentioned above, the *Prm1* and *Smcp* mRNAs share similarities that both mRNAs are repressed initially in free-mRNPs and the repression is mediated *primarily* by the 3' UTR (Kleene 1989; Braun et al., 1989; Hawthorne et al., 2006; Bagarova et al., 2010). This raises an important question whether YBX2 and a YRS or another cis-element and trans-factor repress the *Smcp* mRNAs in steps 3-9.

At the time I began my thesis research there were few clues as to the identity of the *cis*-elements that repress translation in round spermatids. Chapter 2 (Chowdhury and Kleene 2012) describe the use of comparative genomics to identify *cis*-elements that control mRNA translation in spermatids. This approach is based on the assumption that regulatory elements that repress translation can be identified as short conserved sequences, because mutation in these elements will be eliminated from the gene pool by decreased male reproductive success. The analysis of 12 mRNA species that are repressed in round spermatids and active in late spermatids did not reveal obviously conserved elements that are shared by all of these mRNAs. This is either because different elements repress different mRNAs or because purportedly sequence-specific RNA binding proteins are currently understood to bind very short elements with degenerate bases that are difficult to recognize (Morris et al., 2010; Kishore et al., 2010). The degenerate YRS may be a perfect example, especially if Y-box proteins bind YRSs that cannot be identified from the degenerate YRS defined above. However, Chowdhury

and Kleene (2012) detected conserved sequences upstream of poly(A) signals, downstream of poly(A) signals and non-canonical and multiple poly(A) signals, which suggests unexpectedly that elements which repress translation in round spermatids may be present at the 3' ends of the 3'UTR of diverse mRNA species.

Two *Smcp-Gfp* transgenes with 3' UTR mutations have been studied to date; their sequences are depicted in Figure 4.1 in Chapter 4. The $G^5G^CS^3$-mut1 transgene, which contains a randomization of the sequence 6-39 nt upstream of the first poly(A) signal, resulted first detection of GFP fluorescence in steps 4-6 and a small increase in polysomal mRNA in 21 dpp testis, from ~4.5% to ~9%, indicative of partial release of translational repression (Bagarova et al., 2010). Clearly, this segment of the *Smcp* 3'UTR which is in the same position as the *Prm1* TCE does not contain elements which mediate strong repression of the *Smcp* translation.

The second mutation of the *Smcp* 3'UTR, $G^5G^CS^3$-mut2, replaced the 16 nt segment downstream of the first *Smcp* poly(A) signal with 17 nt downstream of the Gfp 3' UTR AAUAAA polyadenylation signal. The developmental expression of the $G^5G^CS^3$-mut2 mRNA, which was studied by Danielle Cullinane in Chapter 4, resulted in complete loss of translational repression: GFP fluorescence is first detected in step 3 spermatids, and ~35% of the $G^5G^CS^3$-mut1 mRNA is associated with polysomes in 21 dpp testis.

The $G^5G^CS^3$-mut2 transgene was designed on the assumption that the 17 nt downstream of the *Gfp* polyadenylation signal lacks regulatory *cis*-elements. This assumption is justified by many precedents in which replacement of a putative regulatory element in one mRNA with a segment from another mRNA results in loss of post-

transcriptional regulation (eg., McGrew and Richter 1989; Braun et al. 1989; Shyu et al. 1991). In addition, there are few reports of regulatory element in the short, 15-30 nt, segments between the poly(A) signal and the poly(A) addition site (Tian et al., 2005). This assumption is further justified by evidence that the 3' terminus of the *Gfp* 3'UTR, which is derived from the SV40 early polyadenylation signal (Kessler et al., 1986), is present in many expression vectors, and has never been reported to bind factors other than the 160 kDa subunit of the cleavage and polyadenylation specificity factor which binds the AAUAAA polyadenylation signal (Murthy and Manley 1995). Nevertheless, my UV-crosslinking RNA binding and RNA-affinity chromatography studies in Chapter 4 reveal that a short probe at the 3' terminus of the $G^5G^CS^3$-mut2 mRNA binds several proteins including the negative regulators of mRNA translation and stability, HNRNPD/AUF1 and far upstream binding protein 1 are particularly abundant (reviewed in White et al., 2013; Zhang and Chen 2013). It is conceivable that these RNA binding proteins promote translation of the $G^5G^CS^3$-mut2 mRNA, because individual RNA binding proteins can function as translational activators and repressors (Eliseeva et al., 2011; Mukherjee et al., 2011; Lebedeva et al., 2011). The functions of these proteins in post-transcriptional regulation of the $G^5G^CS^3$ and *Smcp* mRNAs will require identification of the elements to which these factors bind *in vivo*, mutations which abrogate binding, and analysis of the regulatory consequences of mutations in transgenic mice (Kleene, 2013). Unfortunately, the protein binding to the 3' terminus of the $G^5G^CS^3$-mut2 3'UTR leaves unanswered a basic question whether this segment removes negative regulatory elements that are present at the 3'terminus of the *Smcp* 3'UTR or introduces positive regulatory elements that promote translation.

The functions of YBX2 in translational repression in mouse spermatids are complex. Y-box proteins in general are thought to repress translation of all mRNAs by binding of the C-terminal domain with low specificity to all mRNAs, and mRNA specific translational repression by sequence specific binding of the cold shock domain to the degenerate YRS (Bouvet et al., 1995; Matsumoto et al., 1996; Eliseeva et al., 2011). The levels of YB2 in round spermatids, estimated at ~40 molecules of YBX2 for each molecule of mRNA (Davydova et al., 1997; Yu et al., 2001; Yang et al., 2005), are sufficient to repress translation of all mRNAs in round spermatids. However, high affinity YRSs in the *Prm1* and *Smcp* 5' UTRs, 33 nt upstream of the *Smcp* poly(A) signal, 110 and and 16 nt upstream of the *Prm1* poly(A) signal result, respectively in negligible, slight, negligible and partial repression in transgenic mice (Fajardo et al., 1997; Giorgini et al., 2001; Zhong et al., 2001; Bagarova et al., 2010). In addition, the *Prm1* TCE strongly represses translation in its natural position 4 nt upstream of the *Prm1* poly (A) signal, and no repression in the *Prm1* 5' UTR and 110 nt upstream of the *Prm1* poly(A) signal (Soundajaram et al., 2010).

All of these findings are consistent with the hypothesis that position of a YRS near the 3' terminus of the 3'UTR is necessary for strong translational repression in round spermatids. Table 5.1 compiles a list of mRNAs that are repressed in round spermatids, translationally active in elongating and elongating spermatids, all of which have degenerate YRSs in 1-8 nt upstream of canonical and non-canonical poly(A) signals. However, the importance of position also implies existence of another *cis*-element near the 3' terminus of the 3'UTR. This *cis*-element could be one of the well-known *cis*-elements that are present at the 3' terminus of all mRNAs, the poly(A) signal

and the poly(A) tail, or a short degenerate element that cannot be recognized at present. There is no evidence supporting any of these possibilities. YBX2 is not known to displace or inactivate the functions PABPC1 in translational activation, and this line of speculation is awkward because excess PABPC1 represses translation in round spermatids (Yanagiya et al., 2010). There is also no evidence that the poly(A) signal binds factors which repress translation in the cytoplasm. In addition, my preliminary findings in Chapter 4 indicate that more YBX2 binds the 3' terminus of the translationally active $G^5G^CS^3$-mut2 mRNA than the repressed *Smcp* mRNA.

Table 5.1. Proximity of Y-box recognition sequences to canonical and non-canonical poly(A) signals and poly(A) sites in translationally regulated mRNAs in *Mus musculus* spermatids

Name of mRNA (Acc Number)	Sequence
Diazepam binding inhibitor-like 5 (NM_021294.2)	AACACCAAUAAAUCAUUCAAACUGCA
Protamine 1 (X14003.1)	GCCACCUGUCAAUAAAUGUUGAAAACUCA
Protamine 2 (BC049612.1)	GCCACCUGCCAAAUAAAGCUUGACACGAGA
Transition protein 1 (BC048494.1)	ACAUCUCAAUAACAUUUUGAAAACAAAUAAAAUUGUGA
Transition protein 2 (NM_013694.4)	AACAUGGAUUAAAGCUUGUACCCUGGAAGACUAAAA
Y-box protein 2 (NM_016875.	GCCACCUGAGCCUCCAGUAAAAACAAAAGCAGGCUUUCA

The first five mRNAs are translationally repressed in round spermatids and translationally active in late elongating and elongated spermatids. The *Ybx2* mRNA is predicted to be autoregulated. High affinity YRSs are bold underlined, and a putative YRS is doted underlined. Canonical and non-canonical poly(A) signals are highlighted grey, and polyadenylation sites are highlighted red.

Regardless of the complications above, studies of YBX2 and the 3' terminus of *Prm1* and *Smcp* 3'UTRs provide intriguing intimations of a novel mechanism of mRNA translation. Several experiments to explore this possibility are suggested below.

First, novel RNA binding proteins that bind the A-rich poly(A) signal could be identified with UV-crosslinking assays with RNA probes labeled α[^{32}P]-ATP, RNA affinity chromatography and mass spectrometry sequencing.

Second, understanding the mechanisms and mRNA targets of YBX2 enforced repression would be increased by transcriptome-wide identification of the YRSs and mRNAs to which YBX2 binds *in vivo*. This could be accomplished with HITS-CLIP (Zhang & Darnell 2012), a highly sensitive procedure in which aromatic amino acids that are in direct contact with pyrimidines in living cells are cross-linked with 254 nm UV light, partially digested with RNase T1 to produce short pieces of RNA surrounding the binding site, immunoprecipitated with antibody to an RNA binding protein, resolved by SDS-PAGE, and the RNA fragments are reverse transcribed, cloned and deep sequenced. HITS-CLIP identifies binding sites precisely because amino acids that are covalently-bound to bases cause single nucleotide deletions during reverse transcription.

Third, the atypical features of mRNA repression in spermatids also indicate the absolute importance of validating the functions of YRSs by analyzing the effects of point mutations that abrogate YBX2-binding on the duration and strength of translational repression in transgenic mice (Kleene, 2013). Quantification is necessary to establish whether the mutated-YRS results in partial or complete release of repression, thereby indicating whether repression requires additional *cis*-elements, such as YRSs or binding sites of other RBPs and miRNAs (Bagarova et al., 2010).

Fourth, the question whether strong repression by a high affinity YRS at the distal end of the *Prm1* 3'UTR requires proximity to the poly(A) tail or the polyadenylation signal can be addressed with a transgene in which the TCE YRS and the polyadenylation signal are moved to an upstream 3'UTR position and the TCE in its natural position is inactivated by mutation. The absence of the downstream polyadenylation element in the 3' flanking region will inactivate the addition of a poly(A) tail directed by the poly(A) signal in its new position (reviewed in Lutz, 2008).

REFERENCES

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, et al. 2008 A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 31, 85-799.

Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T 2012 Identification of RNA-protein interaction networks using PAR-CLIP. Wiley Interdiscip Rev RNA. 3, 159-177.

Bagarova J, Chowdhury TA, Kimura M, Kleene KC 2010 Identification of elements in the Smcp 5' and 3' UTR that repress translation and promote the formation of heavy inactive mRNPs in spermatids by analysis of mutations in transgenic mice. Reproduction 140, 853-864.

Balhorn R 2007 The protamine family of sperm nuclear proteins. Genome Biol 8, 227.

Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M et al. 2012 The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46, 674-690.

Bartel DP 2009 MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233.

Bouvet P, Matsumoto K, Wolffe AP 1995 Sequence-specific RNA recognition by the Xenopus Y-box proteins. An essential role for the cold shock domain. J Biol Chem 270, 28297-28303.

Braun RE 1990 Temporal translational regulation of the protamine 1 gene during mouse spermatogenesis. Enzyme 44,120-128.

Braun RE, Peschon J, Behringer RR, Brinster RL, Palmiter RD 1989 Protamine 3'-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. Genes Dev 3, 793-802.

Brown PR, Miki K, Harper DB, Eddy EM 2003 A-kinase anchoring protein 4 binding proteins in the fibrous sheath of the sperm flagellum. Biol Reprod 68,2241-2248.

Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA 1998 Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis. Biol Reprod 58, 834-841.

Burmester S, Hoyer-Fender S. Transcription and translation of the outer dense fiber gene Odf1 during spermiogenesis in the rat: A study by in situ analyses and polysome fractionation. Mol Reprod Dev. 1996;45:10-20.

Calvo SE, Pagliarini DJ, Mootha VK 2009 Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci USA 106, 7507-7512.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM et al. 2012 Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149, 1393-1406.

Cataldo L, Baig K, Oko R, Mastrangelo MA, Kleene KC 1996 Developmental expression, intracellular localization, and selenium content of the cysteine-rich protein associated with the mitochondrial capsules of mouse sperm. Mol Reprod Dev 45, 320-331.

Chao LC, Jamil A, Kim SJ, Huang L, Martinson HG 1999 Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. Mol Cell Biol 19, 5588-5600.

Chi MN, Auriol J, Jégou B, Kontoyiannis DL, Turner JM, de Rooij DG, Morello D 2011 The RNA-binding protein ELAVL1/HuR is essential for mouse spermatogenesis, acting both at meiotic and postmeiotic stages. Mol Biol Cell 22, 2875-2885.

Chowdhury TA, Kleene KC The translational repressor, Y-box protein 2 (YBX2/MSY2), binds the cis-element (TCE) that inactivates mouse Prm1 mRNA translation in round spermatids. In review.

Chowdhury TA, Kleene KC 2012 Identification of potential regulatory elements in the 5' and 3' UTRs of 12 translationally regulated mRNAs in mammalian spermatids by comparative genomics. J Androl 33, 244-256.

Cléry A, Boudet J, Allain FH 2013 Single-stranded nucleic acid recognition: is there a code after all? Structure 21, 4-6.

Dadoune JP. Expression of mammalian spermatozoal nucleoproteins. Microsc Res Tech. 2003;61:56-75.

Danshina PV, Geyer CB, Dai Q, Goulding EH, Willis WD, Kitto GB, McCarrey JR, Eddy EM, O'Brien DA 2010 Phosphoglycerate kinase 2 (PGK2) is essential for sperm function and male fertility in mice. Biol Reprod 82,136-145.

Dass B, Tardif S, Park JY, Tian B, Weitlauf HM, Hess RA, Carnes K, Griswold MD, Small CL, Macdonald CC 2007 Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. Proc Natl Acad Sci USA 104, 20374-2039.

Davies HG, Giorgini F, Fajardo MA, Braun RE 2000 A sequence-specific RNA binding complex expressed in murine germ cells contains MSY2 and MSY4. Dev Biol. 221, 87-100.

Delmas V, Sassone-Corsi P. The key role of CREM in the cAMP signaling pathway in the testis. Mol Cell Endocrinol. 1994;100:121-124.

Dolfini D, Mantovani R 2013 YB-1 YBX1 does not bind to Y/CCAAT boxes in vivo. Oncogene 32, 4189-4190.

Eliseeva IA, Kim ER, Guryanov SG, Ovchinnikov LP, Lyabin DN 2011 Y-box-binding protein 1 (YB-1) and its functions. Biochemistry (Mosc) 76, 1402-1433.

Evdokimova VM, Wei CL, Sitikov AS, Simonenko PN, Lazarev OA, Vasilenko KS, Ustinov VA, Hershey JW, Ovchinnikov LP 1995 The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding

Fajardo MA, Haugen HS, Clegg CH, Braun RE. Separate elements in the 3' untranslated region of the mouse protamine 1 mRNA regulate translational repression and activation during murine spermatogenesis. Dev Biol. 1997;191:42-52.

Farley BM, Ryder SP 2012 POS-1 and GLD-1 repress glp-1 translation through a conserved binding-site cluster. Mol Biol Cell 23, 4473-4483.

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ et al. 2004 Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 428, 493-521.

Giorgini F, Davies HG, Braun RE 2001 MSY2 and MSY4 bind a conserved sequence in the 3' untranslated region of protamine 1 mRNA in vitro and in vivo. Mol Cell Biol 21, 7010-7019.

Giorgini F, Davies HG, Braun RE 2002 Translational repression by MSY4 inhibits spermatid differentiation in mice. Development 129, 3669-3379.

Goossen B, Hentze MW 1992 Position is the critical determinant for function of iron-responsive elements as translational regulators. Mol Cell Biol 12, 1959-1966.

Groppo R, Richter JD 2009 Translational control from head to tail. Curr Opin Cell Biol 21, 444-451.

Grzmil P, Boinska D, Kleene KC, Adham I, Schlüter G et al. 2008 Prm3, the fourth gene in the mouse protamine gene cluster, encodes a conserved acidic protein that affects sperm motility. Biol Reprod 78, 958-967.

Gu W, Tekur S, Reinbold R, Eppig JJ, Choi YC et al. 1998 Mammalian male and female germ cells express a germ cell-specific Y-box protein, MSY2. Biol Reprod 59, 1266-1274.

Gupta YK, Nair DT, Wharton RP, Aggarwal AK 2008 Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity. Structure 16, 549-657.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M et al. 2010 Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129-141.

Hawthorne SK, Goodarzi G, Bagarova J, Gallant KE, Busanelli RR, Olend WJ, Kleene KC 2006A Comparative genomics of the sperm mitochondria-associated cysteine-rich protein gene. Genomics 87, 382-391.

Hawthorne SK, Busanelli RR, Kleene KC 2006B The 5'UTR and 3'UTR of the sperm mitochondria-associated cysteine-rich protein mRNA regulate translation in spermatids by multiple mechanisms in transgenic mice. Dev Biol 297, 118-126.

Herbert TP, Hecht NB 1999 The mouse Y-box protein, MSY2, is associated with a kinase on non-polysomal mouse testicular mRNAs. Nucleic Acids Res 27, 1747-1753.

Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16, 7351-7367.

Hinnebusch AG 2005 Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol 59, 407-450.

Howard TE, Shai SY, Langford KG, Martin BM, Bernstein KE 1990 Transcription of testicular angiotensin-converting enzyme (ACE) is initiated within the 12th intron of the somatic ACE gene. Mol Cell Biol 10, 4294-4302.

Hunter D, Anand-Ivell R, Danner S, Ivell R 2012 Models of in vitro spermatogenesis. Spermatogenesis 2, 32-43.

Idler RK, Yan W 2012 Control of messenger RNA fate by RNA binding proteins, an emphasis on mammalian spermatogenesis. J Androl 33, 309-337.

Iida H, Honda Y, Matsuyama T, Shibata Y, Inai T. Spetex-1 a new component in the middle piece of flagellum in rodent spermatozoa. Mol Reprod Dev. 2006;73:342-349.

Iida H, Ichinose J, Kaneko T, Mōri T, Shibata Y. Complementary DNA cloning of rat spetex-1, a spermatid-expressing gene-1, encoding a 63 kDa cytoplasmic protein of elongate spermatids. Mol Reprod Dev. 2004;68:385-393.

Jackson RJ, Hellen CU, Pestova TV 2010 The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 11, 113-127.

Jenkins HT, Malkova B, Edwards TA 2011 Kinked β-strands mediate high-affinity recognition of mRNA targets by the germ-cell regulator DAZL. Proc Natl Acad Sci USA 108, 18266-18271.

Kashiwabara S, Noguchi J, Zhuang T, Ohmura K, Honda A, Sugiura S, Miyamoto K, Takahashi S, Inoue K, Ogura A, Baba T 2002 Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP. Science 298, 1999-2002.

Kessler MM, Beckendorf RC, Westhafer MA, Nordstrom JL 1986 Requirement of A-A-U-A-A-A and adjacent downstream sequences for SV40 early polyadenylation.

Kierszenbaum AL, Tres LL 1975 Structural and transcriptional features of the mouse spermatid genome. J Cell Biol 65, 258-270.

Kishore S, Luber S, Zavolan M 2010 Deciphering the role of RNA-binding proteins in post-transcriptional control of gene expression. Brief Funct Genomics 9, 391-404.

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M 2011 A A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 8, 559-564.

Kotaja N, Kimmins S, Brancorsini S, Hentsch D, Vonesch JL, Davidson I, Parvinen M, Sassone-Corsi P 2004 Preparation, isolation and characterization of stage-specific spermatogenic cells for cellular and molecular analysis. Nat Methods 1, 249-254.

Kleene KC 1989 Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. Development 106, 367-373.

Kleene KC 2003 Patterns, mechanisms and functions of translational regulation in mammalian spermatogenic cells. Cytogenet Genome Res 103, 217-224.

Kleene KC 2013 Connecting cis-elements and trans-factors with mechanisms of developmental regulation of mRNA translation in meiotic and haploid mammalian spermatogenic cells. Reproduction 146, R1-R19.

Kleene KC, Bagarova J 2008 Comparative genomics reveals gene-specific and shared regulatory sequences in the spermatid-expressed mammalian Odf1, Prm1, Prm2, Tnp1, and Tnp2 genes. Genomics 92:101-106.

Kleene KC, Bagarova J, Hawthorne SK, Catado LM 2010 Quantitative analysis of mRNA translation in mammalian spermatogenic cells with sucrose and Nycodenz gradients. Reprod Biol Endocrinol 25, 155.

Kleene KC, Smith J 1994 Translational activity of mouse protamine 1 messenger ribonucleoprotein particles in the reticulocyte and wheat germ cell-free

Kolmer M, Pelto-Huikko M, Parvinen M, Höög C, Alho H 1997 The transcriptional and translational control of diazepam binding inhibitor expression in rat male germ-line cells. DNA Cell Biol 16, 59-72.

Koonin EV 2005 Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet Annu Rev Gene 39, 309-338.

Korhonen HM, Meikar O, Yadav RP, Papaioannou MD, Romero Y, Da Ros M, Herrera PL, Toppari J, Nef S, Kotaja N 2011 Dicer1 is required for haploid male germ cell differentiation in mice. PLoS One 6, e24821.

Kozak M 1991 Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 266,19867-19870.

Kramer JM 1981 Immunofluorescent localization of PGK-1 and PGK-2 isozymes within specific cells of the mouse testis. Dev Biol 87, 30–36.

Kwon YK, Hecht NB 1993 Binding of a phosphoprotein to the 3' untranslated region of the mouse protamine 2 mRNA temporally represses its translation. Mol Cell Biol 13, 6547-6557.

Kwon YK, Murray MT, Hecht NB 1993 Proteins homologous to the Xenopus germ cell-specific RNA-binding proteins p54/p56 are temporally expressed in mouse male germ cells. Dev Biol 158, 99-100.

Langford KG, Zhou T, Russell LD, Wilcox JN, Bernstein KE. Regulated expression of testis angiotensin-converting enzyme during spermatogenesis in mice. Biol Reprod. 1993;48:1210-1218.

Lebedeva S, Jens M, Theil K, Schwanhäusser B, Selbach M, Landthaler M, Rajewsky N 2011 Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol Cell 43, 340-352.

Lee K, Haugen HS, Clegg CH, Braun RE 1995 Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. Proc Natl Acad Sci USA 92, 12451-12455.

Levy S, Avni D, Hariharan N, Perry RP, Meyuhas O. Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. Proc Natl Acad Sci USA. 1991;88:3319-3323.

Liu D, Brockman JM, Dass BG, Hutchins LN, Singh P et al. Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. Nuc Acids Res. 2006;35:234-246.

Lu ZH, Books JT, Ley TJ 2006 Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. Mol Cell Biol 26, 8410-8417.

Lutz CS 2008 Alternative polyadenylation: a twist on mRNA 3' end formation. ACS Chem Biol 17, 609-617.

Lyabin DN, Eliseeva IA, Skabkina OV, Ovchinnikov LP 2011 Interplay between Y-box-binding protein 1 (YB-1) and poly(A) binding protein (PABP) in specific regulation of YB-1 mRNA translation. RNA Biol. 8, 883-892.

Mali P, Kaipia A, Kangasniemi M, Toppari J, Sandberg M, Hecht NB, Parvinen M 1989 Stage-specific expression of nucleoprotein mRNAs during rat and mouse spermiogenesis. Reprod Fertil Dev 1, 369-382.

Mastrangelo MA, Kleene KC 2000 Developmental expression of Y-box protein 1 mRNA and alternatively spliced Y-box protein 3 mRNAs in spermatogenic cells in mice. Mol Hum Reprod 6, 779-788.

Mathews MB, Sonenberg N, Hershey JWB 2007 Origins and Principles of translational control. In Translational Control in Biology and Medicine, pp 1-40. Eds.

Matsumoto K, Meric F, Wolffe AP 1996 Translational repression dependent on the interaction of the Xenopus Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. J Biol Chem 271, 22706-22712.

Mayr F, Schütz A, Döge N, Heinemann U 2012 The Lin28 cold-shock domain remodels pre-let-7 microRNA. Nucleic Acids Res 40, 7492-7506.

Mathews MB, Sonenberg N, Hershey JWB 2007 Origins and Principles of translational control. In Translational Control in Biology and Medicine, pp 1-40. Eds.
MB Mathews, N Sonenberg N and JWB Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Press.

McGrew LL, Dworkin-Rastl E, Dworkin MB, Richter JD 1989 Poly(A) elongation during Xenopus oocyte maturation is required for translational recruitment and is mediated by a short sequence element. Genes Dev 3, 803-815.

Mehta A, Trotta CR, Peltz SW 2006 Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. Genes Dev 20, 939-953.

Meistrich ML, Mohapatra B, Shirley CR, Zhao M 2003 Roles of transition nuclear proteins in spermiogenesis. Chromosoma 111,483-488.

Morales CR, Oko R, Clermont Y 1994 Molecular cloning and developmental expression of an mRNA encoding the 27 kDa outer dense fiber protein of rat spermatozoa. Mol Reprod Dev 3, 229-240.

Morris AR, Mukherjee N, Keene JD 2010 Systematic analysis of posttranscriptional gene expression. Wiley Interdiscip Rev Syst Biol Med 2, 162-180.

Mueller PP, Hinnebusch AG 1986 Multiple upstream AUG codons mediate translational control of GCN4. Cell 45, 201-207.

Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M Jr, Tuschl T, Ohler U, Keene JD 2011 Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Mol Cell. 43, 327-339.

Murthy KG, Manley JL 1995 The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev 9, 2672-2683.

Nguyen-Chi M, Morello D 2011 RNA-binding proteins, RNA granules, and gametes: is unity strength? Reproduction 142, 803-817.

Nikolov M, Schmidt C, Urlaub H 2012 Quantitative mass spectrometry-based proteomics: an overview. Methods Mol Biol 893, 85-100.

Oko R, Korley R, Murray MT, Hecht NB, Hermo L 1996 Germ cell-specific DNA and RNA binding proteins p48/52 are expressed at specific stages of male germ cell development and are present in the chromatoid body. Mol Reprod Dev 44, 1-13.

Papaioannou MD, Nef S 2010 microRNAs in the testis: building up male fertility. J Androl 31, 26-33.

Paronetto MP, Sette C 2010 Role of RNA-binding proteins in mammalian spermatogenesis. Int J Androl 33, 2-12.

Soundararajan R, Dearth A, Sadate-Ngatchou PI, Sharma M, Braun RE (2010) Positional dependence of the translational control element and role of MSY2 and MSY4 in PRM1 translational control during murine spermatogenesis. Translational Control Meeting, Cold Spring Harbor New York, NY, Sept 13-17, Abstract 264.

Pusch W, Balvers M, Weinbauer GF, Ivell R 2000 The rat endozepine-like peptide gene is highly expressed in late haploid stages of male germ cell development. Biol Reprod 63, 763-768.

Radford HE, Meijer HA, de Moor CH. Translational control by cytoplasmic polyadenylation in Xenopus oocytes. Biochim Biophys Acta 1779, 217-29.

Romero Y, Meikar O, Papaioannou MD, Conne B, Grey C, Weier M, Pralong F, De Massy B, Kaessmann H, Vassalli JD et al. 2011 Dicer11 depletion in male germ cells leads to infertility due to cumulative meiotic and spermiogenic defects. PLoS One 6, e25241.

Russell LD, Ettlin RA, Sinha Hikim AP, Clegg ED 1990 Histopathological Evaluation of the Testis. Clearwater FL, Cache River Press.

Sachs MS, Geballe AP 2006 Downstream control of upstream open reading frames. Genes Dev 20, 915-921.

Sachs R, Max KE, Heinemann U, Balbach J 2012 RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. RNA 18, 65-76.

Sartini BL, Wang H, Wang W, Millette CF, Kilpatrick DL 2008 Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. Biol Reprod 78, 472-482.

Schmidt EE, Hanson ES, Capecchi MR (1999) Sequence-independent assembly of spermatid mRNAs into messenger ribonucleoprotein particles. Mol Cell Biol 19, 3904-3915.

Shih DM, Kleene KC. A study by in situ hybridization of the stage of appearance and disappearance of the transition protein 2 and the mitochondrial capsule seleno-protein mRNAs during spermatogenesis in the mouse. Mol Reprod Dev. 1992;33:222-227.

Shyu AB, Greenberg ME, Belasco JG 1989 The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev 3, 60-72.

Skabkin MA, Liabin DN, Ovchinnikov LP 2006 Nonspecific and specific interaction of Y-box binding protein 1 (YB-1) with mRNA and posttranscriptional regulation of protein synthesis in animal cells. Mol Biol (Mosk) 40, 620-633.

Sonenberg N, Hinnebusch AG 2009 Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 20, 136, 731-745.

Steel LF, Telly DL, Leonard J, Rice BA, Monks B, Sawicki JA 1996 Elements in the murine c-mos messenger RNA 5'-untranslated region repress translation of downstream coding sequences. Cell Growth Differ 7, 1415-1424.

Svitkin YV, Evdokimova VM, Brasey A, Pestova TV, Fantus D, Yanagiya A, Imataka H, Skabkin MA, Ovchinnikov LP, Merrick WC et al. 2009 General RNA-binding proteins have a function in poly(A)-binding protein-dependent translation. EMBO J 28, 58-68.

Tafuri SR, Familari M, Wolffe AP 1993 A mouse Y box protein, MSY1, is associated with paternal mRNA in spermatocytes. J Biol Chem 268, 12213-122120.

Tian B, Hu J, Zhang H, Lutz CS 2005 A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33, 201-212.

Tseden K, Topaloglu O, Bohm D, Wolf S, Muller C et al. 2007 Premature translation of transition protein 2 mRNA causes sperm abnormalities and male infertility. Mol Reprod Dev 74, 273-279.

Ule J, Jensen K, Mele A, Darnell RB 2005 CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37, 376-386.

Ursini F, Heim S, Kiess M, Maiorino M, Roveri A, Wissing J, Flohé L 1999 Dual function of the selenoprotein PHGPx during sperm maturation. Science 285, 1393-1396.

Walker J, de Melo Neto O, Standart N 1998 Gel retardation and UV-crosslinking assays to detect specific RNA-protein interactions in the 5' or 3' UTRs of translationalluy regulated mRNAs. Methods Mol Biol 77, 365-378.

Wang H, Sartini BL, Millette CF, Kilpatrick DL 2006 A developmental switch in transcription factor isoforms during spermatogenesis controlled by alternative messenger RNA 3'-end formation. Biol Reprod 75, 318-323.

Wei WJ, Mu SR, Heiner M, Fu X, Cao LJ, Gong XF, Bindereif A, Hui J 2012 YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. Nucleic Acids Res 40, 8622-8636.

Weiner AM, Deininger PL, Efstradiatis A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Ann Rev Biochem. 1986;55:631-661.

Welch JE, Brown PR, O'Brien DA, Eddy EM (1995) Genomic organization of mouse glyceraldehyde 3-phosphate dehydrogenase gene (Gapd-s) expressed in post-meiotic spermatogenic cells. Dev Genet 16, 179-189.

Welch JE, Sakai Y, Eddy EM 1992 In situ localization of spermatogenic cell-specific glyceraldehyde 3-phosphate dehydrogenase Gapd-s messenger ribonucleic acid in mice. Biol Reprod 46, 859-868.

White EJ, Brewer G, Wilson GM. Post-transcriptional control of gene expression by AUF1: mechanisms, physiological targets, and regulation. Biochim Biophys Acta. 2013 Jun-Jul;1829(6-7):680-8. doi: 10.1016/j.bbagrm.2012.12.002. Epub 2012 Dec 14. Review. PubMed PMID: 23246978; PubMed Central PMCID: PMC3664190.

Wilusz J, Shenk T 1988 A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell 52, 221-228.

Yan W, Ma L, Burns KH Matzuk MM 2003 HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. Proc Natl Acad Sci USA 100, 10546-10551.

Yanagiya A, Delbes G, Svitkin YV, Robaire B, Sonenberg N 2010 The poly(A)-binding protein partner Paip2a controls translation during late spermiogenesis in mice. J Clin Invest 120, 3389-3400.

Yang J, Medvedev S, Reddi PP, Schultz RM, Hecht NB 2005 The DNA/RNA-binding protein MSY2 marks specific transcripts for cytoplasmic storage in mouse male germ cells. Proc Natl Acad Sci USA 102, 1513-1518.

Yang J, Morales CR, Medvedev S, Schultz RM, Hecht NB 2007 In the absence of the mouse DNA/RNA-binding protein MSY2, messenger RNA instability leads to spermatogenic arrest. Biol Reprod 76, 48-54.

Zhang J, Chen QM. Far upstream element binding protein 1: a commander of transcription, translation and beyond. Oncogene. 2013 Jun 13;32(24):2907-16. doi:10.1038/onc.2012.350. Epub 2012 Aug 27. Review. PubMed PMID: 22926519.

Zhang C & Darnell R 2012 Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 29, 607-614.

Zhong J, Peters AHFM, Kafer K, Braun RE 2001 A highly conserved sequence essential for translation repression of the protamine 1 messenger RNA in murine spermatids. Biol Reprod 64, 1784-1789.

Zhong X, Kleene KC 1999 cDNA copies of the testis-specific lactate dehydrogenase (LDH-C mRNA are present in spermatogenic cells in mice, but processed pseudogenes are not derived from mRNAs that are expressed in haploid and late meiotic spermatogenic cells. Mamm Genome 10, 6-12.