Systems for studying the non-ubiquitous functions of the TATA-binding protein
by Tammy Alice Tucker

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Veterinary Molecular Biology
Montana State University

Abstract:
As the core protein for transcription initiation by all three nuclear RNA polymerases, the TATA-binding protein, TBP, is one of the most ubiquitously expressed transcription factors in somatic tissues. The C-terminal amino acid sequence and structure are identical from Archaebacteria to humans; perhaps making TBP one of nature's most highly conserved proteins as well. Despite these properties, work from our laboratory and others has revealed some non-ubiquitous and non-conserved characteristics of TBP. First, rodent TBP mRNA is expressed in spermatids at levels nearly 1,000-fold higher than levels in somatic cells. This is due three major isoforms of TBP mRNA - one ubiquitous, one spermatid-specific and one spermatid- and brain-specific, all of which exhibit different translational regulation in spermiogenesis. Second, gene targeting has generated a line of mice with a modified form of tbp. Although most functions, including spermiogenesis, are normal for mice homozygous for the mutation, mutant placentas often fail to correctly interact with the maternal immune system. Lastly, vertebrate TBPs contain a homo-glutamine (Q) repeat that can differ in length between individuals in a population. Gross expansion of this type of repeat in humans, similar to expansions of the Q-repeats in other proteins (e.g., Huntingtin) can lead to neurological disorders including trembling and ataxia. This Master of Science thesis comprises three parts that either address the peculiarities of TBP or uses the tbp gene to investigate basic biological functions. Based on differential translational regulation of TBP mRNA isoforms in spermatids, the first section investigated the mechanism that times the translation of stored mRNAs during spermiogenesis. The second section explored a potential tissue-specific function of TBP by focusing on a naturally occurring neuropathic aspermic mutant mouse line, quaking (Qkv), and a potential link between TBP, Qkv, and the immune system. The final section of this thesis utilized the tbp gene to study the parameters affecting homologous recombination-mediated gene targeting in bovine cells. The system devised here may be useful gene targeting in other mammalian species as well.

SYSTEMS FOR STUDYING THE NON-UBIQUITOUS FUNCTIONS OF THE TATA-BINDING PROTEIN

by

Tammy Alice Tucker

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Veterinary Molecular Biology

MONTANA STATE UNIVERSITY-BOZEMAN
Bozeman, Montana

October 2003

APPROVAL

of a thesis submitted by

Tammy Alice Tucker

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Edward E. Schmidt, Ph.D.
Chair
(Signature) Ed Schmidt
Date 10/21/3

Approved for the Department of Veterinary Molecular Biology

Allen G. Harmsen, Ph.D.
Department Head
(Signature)
Date 10/24/03

Approved for the College of Graduate Studies

Bruce R. McLeod, Ph.D.
Graduate Dean
(Signature) Bruce R. McLeod
Date 10-28-03

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under the rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature Tamy Tucker

Date 10/27/03

ACKNOWLEDGEMENTS

I would like to extend special thanks Jim Thompson and Keri Rask for their assistance in purchasing and caring for our heifers, and the support of Dr. J. Page Lindsey, Dr. Mick Guerini, the members of the Schmidt lab, my friends and family.

I dedicate this work to my daughters, Rea LeeAnn Walker, and Corey Page Tucker.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES (CONTINUED)

ABSTRACT

As the core protein for transcription initiation by all three nuclear RNA polymerases, the TATA-binding protein, TBP, is one of the most ubiquitously expressed transcription factors in somatic tissues. The C-terminal amino acid sequence and structure are identical from Archaebacteria to humans; perhaps making TBP one of nature's most highly conserved proteins as well. Despite these properties, work from our laboratory and others has revealed some non-ubiquitous and non-conserved characteristics of TBP. First, rodent TBP mRNA is expressed in spermatids at levels nearly 1,000-fold higher than levels in somatic cells. This is due three major isoforms of TBP mRNA - one ubiquitous, one spermatid-specific and one spermatid- and brain-specific, all of which exhibit different translational regulation in spermiogenesis. Second, gene targeting has generated a line of mice with a modified form of *tbp*. Although most functions, including spermiogenesis, are normal for mice homozygous for the mutation, mutant placentas often fail to correctly interact with the maternal immune system. Lastly, vertebrate TBPs contain a homo-glutamine (Q) repeat that can differ in length between individuals in a population. Gross expansion of this type of repeat in humans, similar to expansions of the Q-repeats in other proteins (e.g., Huntingtin) can lead to neurological disorders including trembling and ataxia. This Master of Science thesis comprises three parts that either address the peculiarities of TBP or uses the *tbp* gene to investigate basic biological functions. Based on differential translational regulation of TBP mRNA isoforms in spermatids, the first section investigated the mechanism that times the translation of stored mRNAs during spermiogenesis. The second section explored a potential tissue-specific function of TBP by focusing on a naturally occurring neuropathic aspermic mutant mouse line, quaking (Qk^v), and a potential link between TBP, Qk^v, and the immune system. The final section of this thesis utilized the *tbp* gene to study the parameters affecting homologous recombination-mediated gene targeting in bovine cells. The system devised here may be useful gene targeting in other mammalian species as well.

CHAPTER ONE

STUDY INTRODUCTION

TATA-binding protein and transcription initiation

Some of the key features of eukaryotic protein-coding genes include the protein-coding sequences (1), a promoter (2), various regulatory elements (3), and a transcriptional start codon (2). Protein-coding genes are transcribed by RNA pol (ribonucleic acid polymerase) II to produce messenger RNA (mRNA), which directs the production of every polypeptide required by a cell (4). RNA pol II alone cannot initiate transcription of a gene (2). Regulatory elements and the promoter moderate the production of mRNAs, and in turn, provide the vast diversity of development, differentiation, function, processes, and maintenance encoded within the genes of a cell (1). Proteins known as basal and inducible transcription factors (TF) bind respectively to the promoter and regulatory elements in a specific manner (3). RNA pol II is recruited to this complex of proteins at the promoter and transcription is initiated (5). The production of ribosomal RNA through RNA pol I and transfer RNA through RNA pol III is similar; only requiring different TF arrangements at the promoter and regulatory elements (2). The only TF that is known to be common to all RNA pol transcription initiation complexes is the TATA-binding protein (TBP) (4). TBP is a DNA (deoxyribonucleic acid)-binding TF that is constitutively expressed throughout development and by all cells (6). The amino acid sequence in the carboxy-terminal region (C-terminus) of TBP is greatly conserved between Archaebacteria and humans, and the amino-terminal region

(N-terminus) is conserved among vertebrates (4). Consequently, the global importance of this TF in transcription initiation is obvious. In addition, recent research has described characteristics of TBP that extend beyond these basal or ubiquitous functions.

TBP has long been considered the universal transcription factor because of its presence in the initiation complexes for all nuclear transcription (4). Recently, though, evidence for the existence of TBP-related factors (TRFs) which replace TBP function in particular cell types undermines the concept of global necessity for TBP. For example, TRF 2 is a transcription factor found to replace TBP at certain somatic cell loci in *Caenorhabditis elegans*, *Xenopus laevis*, and *Danio rerio* (6). Although TRF 2 is necessary for the differentiation of the male germ cell in the testis of *Mus musculus* (7), TBP expression in spermatids is approximately 1000 fold over that found in somatic tissues of the same species (8). This is of particular interest because transcription arrests at the time of spermatid differentiation (spermiogenesis). The presence of TRF 2 and TBP in spermiogenic cells that lack transcription of any kind brings into question the true function of TFs and TBP in the testis (9).

The importance of TBP in developmental processes has been demonstrated by gene targeting experiments that remove all or part of the *tbp* gene. Embryos that cannot produce their own TBP do not progress past the blastocyst stage (10). On the other hand, modification of only the N-terminus of TBP does not adversely affect basic cellular processes or development, and germ cell production in animals lacking this region is normal as well (11). The TBP N-terminus is specific to and conserved in vertebrates (4). This region is thought to facilitate DNA-binding by inducing conformational changes

once additional activating transcription factors have bound to this region (6). Processes that are affected by deletions of the N-terminus are those that are highly specialized, like the interaction between the placenta and the maternal immune system (12). It is possible that the N-terminus of TBP regulates expression of a very restricted subset of genes in mammals, and that this activity is dependent upon interactions with tissue-specific proteins.

Despite the conservation of the core amino acids within TBP among vertebrates, the N-terminus contains a region of homo-glutamine (Q) repeats whose length can vary among individuals in a population (4). Elongated stretches of Q repeats have been shown to cause protein aggregation that can lead to neurodegenerative disorders in humans (13). These diseases include myotonic dystrophy, Kennedy's disease, fragile X syndrome, various ataxias, and Huntington's disease (14, 15, 16). Insoluble TBP was shown to accumulate in the brain of patients with Huntington's disease (17). Whether the accumulation of TBP is a cause or a consequence of the disease is undetermined (15, 17); however, since the Q repeats of TBP can be expanded between individuals without these extreme pathogenic consequences as well, the conservation of amino acid sequence in this protein is variable.

Clearly, the unconventional, less ubiquitous functional characteristics of TBP are just beginning to be explored. Either this protein is less universal than once thought, or it is more flexible than previously perceived. We hypothesize the latter, in that TBP functions in specific biological processes as well as global transcription initiation. Additional studies in unique systems are needed to address this hypothesis. I investigated

TBP in three different biological settings: murine spermatogenesis, the male-sterile quaking mutant mouse, and gene targeting in bovine embryonic fibroblasts. The particulars of these systems and how they facilitate the study of TBP are defined in the sections that follow.

Biological systems that may involve TBP unique functions

Spermatogenesis

The transformation of a male primary germ cell into dozens of highly specialized spermatozoa is achieved through three sequential phases: mitosis, meiosis and spermiogenesis, and is termed spermatogenesis. Intricate mechanisms and thousands of genes under stringent control guide the germ cells through each phase with precision (18). Although the time it takes for each phase of development varies for many species, the fundamental processes and outcomes are remarkably similar (19). The three phases of spermatogenesis are outlined in Figure 1.

Mitosis. Multiple mitotic divisions of the primary germ cell result in the spermatogonia. Each spermatogonium maintains a diploid set of chromosomes (40 for the mouse) and is visibly distinguishable from the primary germ only by size (20). In order to preserve the stem cell population, one of the daughter cells from the first few divisions remains dormant until the next cycle of spermatogenesis (21). The dividing cells are linked to the others in the same lineage through cytoplasmic bridges and remain attached throughout meiosis and spermiogenesis (22). An apparent failure of cytokinesis to go to completion following division is a valuable feature because these bridges are essential for

synchronous development at all phases of the cycle and accommodate the inherent haploid state of post-meiotic spermatocytes by allowing cytoplasmic contents to be shared (23).

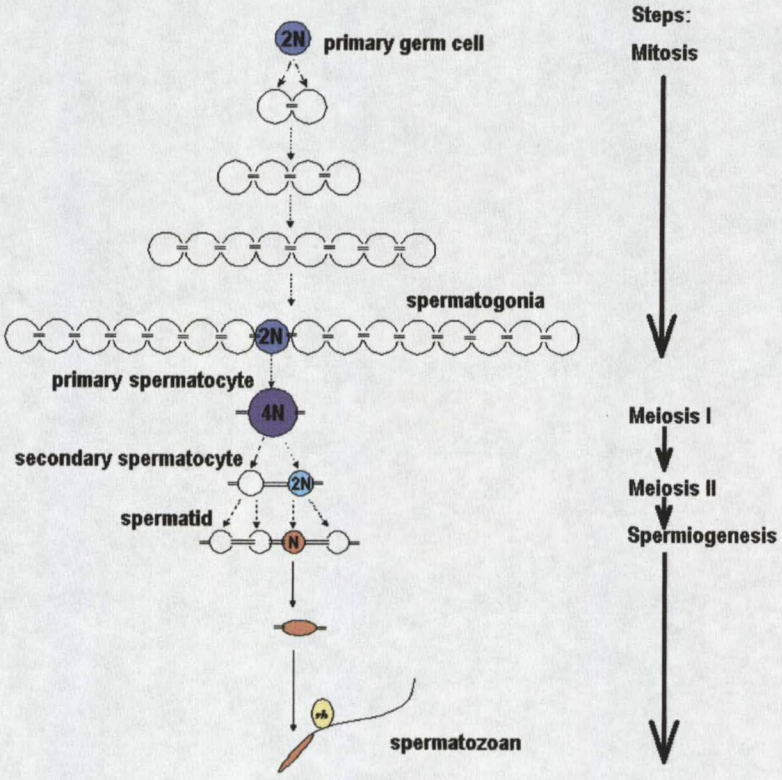

Figure 1. The three phases of spermatogenesis in the mouse. Mitosis generates 16 interconnected, diploid (2N) spermatogonia in the mouse and takes approximately 7 days. Primary spermatocytes replicate their DNA (4N) and divide once to yield 32 interconnected secondary spermatocytes after meiosis I. The reduction division of meiosis II results in 64 haploid (N) spermatids still connected by cytoplasmic bridges. The meiosis phase takes approximately 13 days. Spermiogenesis, the differentiation of the haploid spermatid, occurs in 16 distinguishable stages (only 3 are represented here: round spermatids, elongating spermatids and spermatozoan). The nucleus is condensed, the cells elongate, the flagella and acrosome forms, and the residual bodies (*rb*) are sloughed-off during this phase. The *rb* contain excess cytoplasm, RNA, organelles and proteins, and they maintain connection to other *rb* through cytoplasmic bridges until they are phagocytosed by the surrounding Sertoli cells. The final product, the mature spermatozoan, is released into the lumen of the testis after nearly 15 days of differentiation. The complete cycle takes approximately 34.5 days. Spermatogenesis phases are reviewed in reference 2. Picture adapted from Braun et al. (23).

Mitosis sustains the number of undifferentiated cells, contributes significantly to the expansion of the population of cells that will eventually be released into the lumen of the testis, and ensures that each spermatogonium has the capacity and contents to reach the next phase of differentiation.

Meiosis. Once the mitotic phase is complete, the cells undergo meiosis, reducing the number of chromosomes by half (20 in the mouse) and expanding the cell number by a factor of four. Obviously, a considerable amount of modified transcription must occur for this specialized division. Valerio Monesi (24) found that both RNA and protein synthesis increased in primary spermatocytes as evidenced by the uptake of tritium-labeled uridine and arginine by cell extracts. According to the work of Yu et al. (18), substantial up- and down-regulation of genes occurs during the transition from spermatogonia to primary spermatocytes. Transcription and translation are rapidly activated in primary spermatocytes, however, once the divisions begin, the nucleus becomes silent, and polysomal loading ceases until meiosis is completed (25). In essence, spermatocytes undergo the two divisions of meiosis and, at the end of this phase, the round spermatids are haploid and begin to differentiate.

Spermiogenesis. The features of spermatid differentiation, including unique structures such as the flagella, acrosome and residual body, as well as the complete restructuring of the nucleus, are of particular importance to successful delivery of the genomic message. While round spermatids that have completed meiosis are capable of producing viable and fertile offspring (26), they cannot do so naturally without these

sophisticated modifications. For instance, as the axoneme, mitochondrial sheath, and fibrous sheath extend from the proximal and distal centrioles the flagella forms (27), the synthesis of which occupies a significant portion of spermiogenesis (28, 29). Likewise, the acrosome is produced by the Golgi apparatus, and although its formation is complete by the time spermatids elongate, Yashinaga and Toshimori (30) present a model of maturation that continues through fertilization. Excess cytoplasm, RNA, and proteins that would otherwise make the arduous journey toward fertilization less efficient are shed in the residual body at the end of spermatid maturation (24). Protamine replaces histones on the DNA in order to condense the nucleus and facilitate travel, and perhaps, for important interactions with the female genome upon fertilization (31). Each of these changes is absolutely required to produce an efficient genome-transport vector. Yet, as in meiosis, the events of spermiogenesis are preceded by, and dependent on, the large scale transcription of required messenger ribonucleic acids (mRNA)s at the onset of spermiogenesis (32). Furthermore, the mRNAs produced at this stage come entirely from a haploid genome (33). The remaining steps of development are no longer regulated by the nucleus, as it has been compacted in such a way as to make transcription impossible (34). Still, a system exists that allows translation of certain messenger ribonucleic acids mRNAs into protein only when they are needed. Protamine, for example, is transcribed in round spermatids but is not translated until spermatids elongate (31, 34, 35). Translation of protamine at an earlier stage results in defective spermatid development (36). The mechanism by which the timing of expression is achieved may be programmed into the

mRNA itself and/or into trans-acting proteins that convert round spermatids to spermatozoa.

For decades researchers have dissected known spermiogenic genes in order to determine how they are regulated and thereby infer a mechanism for protein expression in spermiogenesis. Unfortunately, sufficient data does not exist to make any conclusive arguments. Nevertheless, as with somatic translational regulation, messenger ribonucleoprotein particles (mRNP)s have been implicated in regulation of spermatogenic translation (33, 37). Spermatogenic transcripts are processed in the nucleus and are transported to the cytoplasm where most, if not all, are bound into mRNP complexes (38). How, when, and if the mRNA escapes the mRNPs is the subject of much speculation. Data exists to connect the untranslated regions (UTR)s of the mRNA via sequence or secondary structure to translational expression or suppression (33, 37, 39, 40). From the viewpoint of N. B. Hecht, the 3' UTR establishes whether or not the mRNA will be stored or loaded with polysomes and translated immediately (33). R. E. Braun proposes that various elements in the 5' and 3' UTRs of different mRNAs determine the sequestration or activation at particular stages of differentiation (37, 44). Zhong et al. (39), assert that elements in the 3' UTR are required to repress translation, but are not also necessary for incorporation into mRNPs. Schmidt et al. (38) argue that all transcripts are bound into mRNPs regardless of the 3' or 5' UTR sequence, and that regulation involves the timing of translation of these repressed mRNAs. Evidence reviewed by K. Steger for the involvement of repressor genes within mRNPs such as Poly A-binding protein (PABP) and Y-Box proteins has intriguing implications for

outside mechanisms recognizing patterning within mRNPs at particular stages (40). Steger also proposes that the presence of chromatoid bodies within the nucleus contain repressed RNAs, and as such, may also be a mechanism for control of translation of unnecessary mRNAs (40). In one review K.C. Kleene (25) points out that proteins from inefficiently translated transcripts are abundant in spermatogenic cells due to the promiscuous transcription following meiosis. Kleene believes that mRNPs as well as other mechanisms, such as the alteration of the length of the poly-A tail, exist to offset the surplus of mRNA (41, 42, 43). Hecht disagrees that such promiscuous transcription would ensue, given the importance of this process in maintaining the species, and insists that alternative gene products from alternative promoters play a role in how regulation is managed in the testis (45). Using TBP as an example, Schmidt et al. (46) discuss the possibility that promiscuous splicing may be the cause of repression of translation in the testis. TBP, as described earlier, is a transcription factor required by all cells to initiate production of mRNA (4). Two alternate promoters utilized only in testis (1D) or only in testis and brain (1E), are differentially spliced and regulated during spermiogenic stages, although each generates an identical gene product (46). The specific mechanism for translational regulation of these isoforms is unresolved, however promoter 1D TBP is most often polysomal, and promoter 1E TBP is almost entirely packaged into mRNPs (9). While neither isoform, as defined by the traditional role of TBP, can be effective in a transcriptionally silent cell, they may have alternative functions necessary for continued development of spermatids (9). Definition of commonalities between mRNAs that are

either sequestered or translated at any given moment in spermiogenesis is imperative to the discovery of the true timing mechanism.

Timing of translation in spermiogenesis

To date, the structural DNA-binding protein, protamine 1 (*prm1*), is the best characterized of the translationally regulated genes in spermiogenesis. It has been well established that *prm 1* mRNAs are transcribed in round spermatids, but are not translated until nuclear condensation begins in elongating spermatids, and that the suppression of the *prm 1* mRNA is likely achieved through assembly into mRNPs (44). Transgenic experiments have isolated particular sequences within the 3' UTR that establish *prm 1* translational repression (39). Other genes have been identified that are under translational control, however, none have shown similar repression by way of sequence or structure to *prm 1* (37). The mechanisms of release of *prm 1* from mRNPs and its transition to polysomes are also unclear. R.E. Braun (37) refers to the application of activators on the mRNPs which release the mRNA, however, the governance of the activator must then be considered. Positional analysis of the mRNA within the cytosolic compartments has revealed no localized storage of prm 1 (47). Braun, however, proposes that the arrangement of microtubules and the proteins bound to them as the nucleus condenses brings the activator in contact with mRNPs at particular stages of development to initiate translational activation (37). All things considered, the mechanism that defines when an mRNA is associated with mRNPs or is actively translated is undefined.

One of the inherent difficulties in defining the timing mechanism is that isolation and characterization of numerous transcripts at a single stage has not been accomplished.