UNIVERSITY OF CALGARY

The Transcriptional Regulator BmGATAß of the Silkmoth *Bombyx mori*: Isolation of Putative Interacting Factors from Ovarian Follicular Cells

by

Martin Glushek

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Calgary, Alberta

January, 2001

Abstract

The expression of the chorion genes during oogenesis in *Bombyx mori* is regulated by the zinc finger transcription factor BmGATAß, which undergoes extensive post transcriptional and posttranslational modifications. The yeast two hybrid system has been used to isolate seven potential interacting cDNA clones, four of which have closely related sequence paralogues and may be involved in BmGATAß processing and function. The remaining three appear to be anonymous in current sequence databases.

5' RACE and RT PCR based techniques were used to expand the sequence information available for these clones. Northern hybridizations were used to establish their tissue specific and temporal patterns of expression. *In situ* hybridizations performed against two of the clones confirmed the Northern analyses, and restricted expression of their transcripts to the follicular epithelium. Finally, GST-BmGATAß fragment fusions were utilized in an attempt to demonstrate *in vitro* interactions through the use of pull down experiments.

Acknowledgements

I would like to offer my sincere gratitude to my supervisor, Dr. Kostas Iatrou, for giving me the opportunity to participate in an interesting and rewarding program of research. I also extend my thanks to my committee members, Dr. Derrick Rancourt and Dr. Frans van der Hoorn, for their suggestions and ideas throughout this program. I am also grateful to all of my examining committee members, including my external examiner Dr. Steven Zimmerly, for their efforts in the evaluation of this thesis.

Much of the work represented here was significantly enhanced through the efforts and detailed protocols of Dr. Ken Ito; his friendliness and willingness to assist and endure endless technical questions with a smile are greatly appreciated.

Finally, I would like to thank my parents, Robert and Connie Glushek, for their encouragement and unquestioning support throughout the years.

Table of Contents

List of Figures

SECTION I : General Introduction

The processes of cellular differentiation and development are driven by the action of transcriptional activators and repressors that ultimately control the appropriate spatial and temporal expression of genes required for successful completion of the processes. In turn, the activities of these drivers are controlled and modulated through three primary mechanisms: direct transcriptional regulation, post transcriptional processing generating isoforms with altered activities, and post translational modifications.

The ability of modern molecular biology to sort out the details of these processes has been greatly hampered by their complexity, with much information arising from reconstitution type experiments that usually fail to properly address the nature of the original inductive signals or the subtleties of the transcriptional response. Conceivably, the most productive system for such studies would encapsulate all of these features, including induction and immediate response factors, upregulation and modification of the necessary transcriptional modulators, and a well defined, developmentally significant series of target genes. All of these features are encapsulated within the chorion gene cluster of *Bombyx mori*, where a hormonally driven developmental cascade concludes with the coordinated regulation of more than 100 well characterized target genes required for the development and assembly of the oocyte chorion. The purpose of this thesis is to detail the recovery and initial characterization of developmentally regulated factors that may be involved in the post translational processing and modulation of activity of the chorion gene specific transcription factor BmGATAß.

Ovarian Development in *Bombyx mori*

The initiation of ovarian development in *B. mori* is under the control of the Lepidopteran moulting hormone 20 - hydroxyecdysone (20E) (Tsuchida et al., 1987). This hormone is released from the prothoracic gland during pupal development and

continues to accumulate in the haemolymph and follicular cells well into maturation. The developmental cascade that ensues differentiates both the ovary itself, and the ovarioles and follicles within. It has previously been shown that 20 E is both necessary and sufficient for this process to occur, with ovaries and oocytes being capable of completing terminal differentiation *in vitro* beyond a designated point (Swevers and Iatrou, 1995). The use of ligation / injection experiments and ecdysone analogues have suggested that the rise and fall of 20E in the haemolymph is required for most follicles to terminally differentiate (Swevers and Iatrou, 1992; Swevers and Iatrou, 1995). These experiments not only demonstrated that the ability to undergo terminal differentiation is acquired through the expression of a response factor, but they also suggested that this factor must be produced within the follicular cells because all stages of developing follicles are in contact with the same haemolymph.

Within the ovary, oocytes are formed by the asymmetric distribution of the initial germ cell cleavage products into an oocyte and overlay of seven nurse cells (Figure 1b). The follicular epithelium is of somatic origin, however, and is presumed to be peeled off of the surrounding epithelial lining of the germarium. These cells divide as the oocyte expands until they number approximately 5000, after which mitosis stops and the remaining expansion occurs via changes in cellular volume and responses to polyploidization (Kafatos et al., 1977). This epithelium is ultimately responsible for the expression and construction of two primary structures, the vitelline membrane and the chorion.

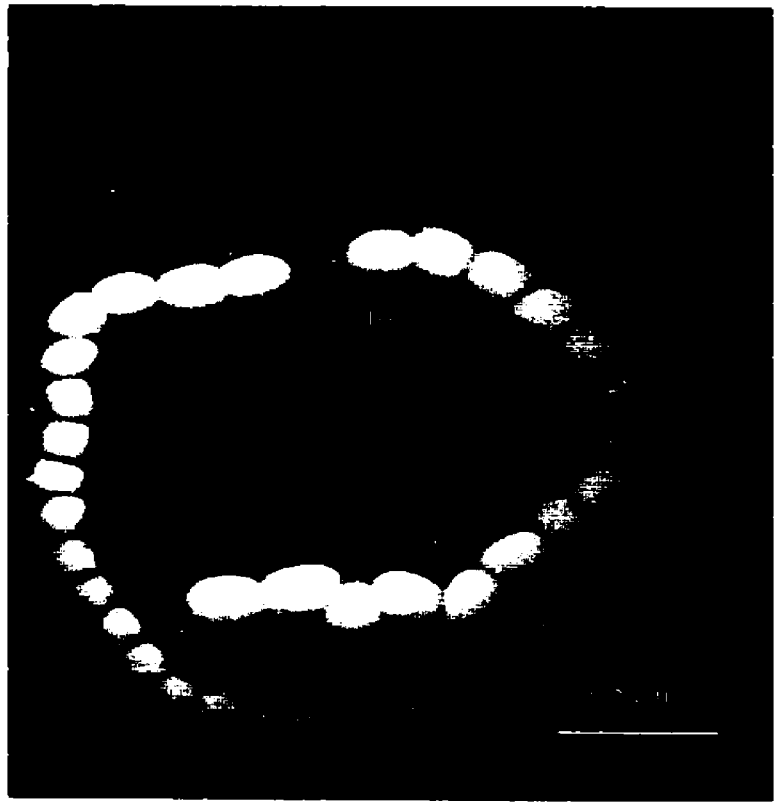
One of the most powerful and interesting features of follicle development in *Bombyx* is its asynchronous nature; follicles within a developing ovariole differentiate in a linear array, with each follicle differing from its neighbor by approximately 2 to 2.5 developmental hours (Swevers and Iatrou, 1986; Figure 1a). Thus, factors responsible for the commitment to or process through all developmental stages can be recovered from a single animal. This has proven extremely useful for the identification of the factors

Figure 1: Follicle Development and Oocyte Morphology

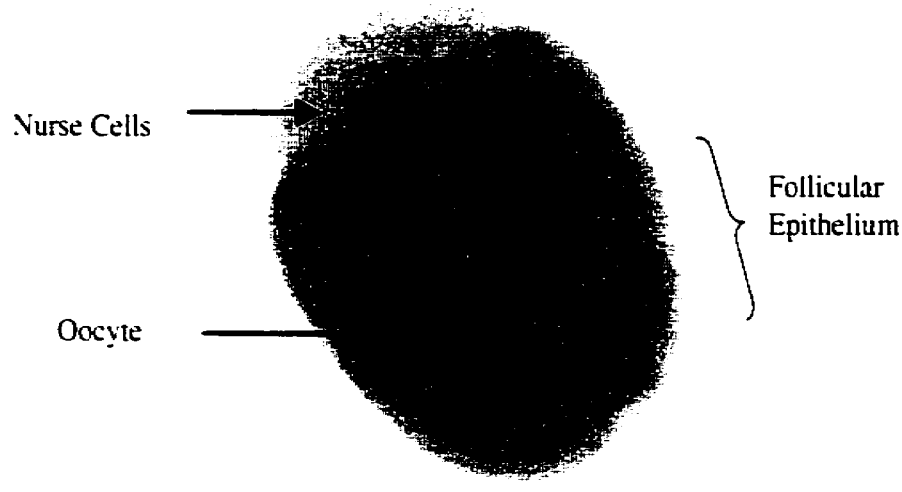
Panel A : Ovariole dissected from a day 6 pupae. The first follicle expressing chorion genes is designated as +1. Relative size is denoted by the scale bar at the bottom of the figure.

Panel B : Single vitellogenic follicle prepared and fixed for *in situ* hybridization showing the three major cell types found in the vitellogenic Silkmoth follicle. The picture was taken with a Zeiss III RS Light Microscope.

(A)



(B)



involved in the process, and has provided a simple system for the characterization and analysis of the chorion gene expression cascades (discussed below).

In *Bombyx*, it is convenient to divide the oogenic process into three functional stages: previtellogenesis, which establishes the cellular polarity and form of the oocyte; vitellogenesis, when nurse cells become active in the synthesis and distribution of mRNA and yolk proteins into the awaiting oocyte, and the cells of the follicular epithelium synthesize and secrete the components of the vitelline membrane that surrounds the oocyte; and choriogenesis, where cells of the follicular epithelium are induced to express and secrete the proteins that form the chorion (Bock et al., 1986). The rigorously maintained pattern of chorion gene expression has permitted the creation of a developmental timeline for the oogenic process (Figure 1a); in this scheme, the first follicle expressing messages for the earliest chorion proteins is designated as +1, with increasing negative numbers assigned to follicles of decreasing age, and increasing positive numbers to those follicles undergoing choriogenesis.

The Chorion Locus

The silkmoth chorion gene superfamily stretches over several hundred kilobases of chromosome II, with the vast majority of genes segregating into 2 closely linked classes (alpha and beta). Chorion genes are organized as divergently oriented pairs sharing a short common promoter, ranging in size from 190 to 340 bp (Eickbush and Izzo, 1995) with one member of a pair being a member of the alpha class of genes and the other of the beta. This arrangement suggests that the chorion gene families arose through the creation of these tandem units, followed by the duplication and segregation of successive gene pairs (Spoerel et al., 1989). These clusters are, in turn, segregated into multigene families (ErA, ErB, A, B, HcA, and HcB) whose physical organization appears to be echoed by both relatedness of protein products and temporal patterns of gene expression. To date, over 100 distinct proteins have been identified (Swevers and Iatrou, 1992)

arising from as many as 200 structural genes (Eickbush and Kafatos,1982). While members of the same gene family may share sequence, promoter and expression characteristics, the overall organization of the chorion gene pairs does not immediately suggest a linear correlation between physical location and pattern of expression; several genes in each class do not share common promoters, gene pairs with disparate expression patterns may reside in close proximity, and there is little similarity in sequence motifs between the early chorion clusters and those of the middle and late classes (Hibner et al., 1988; Spoerel et al., 1989).

The expression patterns for several of the chorion genes have been elucidated, and they appear in striking and elegant overlapping gradients (Figure 2b). While only some of the trans acting factors controlling their expression have been identified (discussed below), it is assumed that all of the necessary cis information required is contained within the shared promoter regions, as relevant expression constructs introduced into *Drosophila* demonstrate the appropriate tissue specific and developmentally regulated patterns of expression (Mitsialis et al., 1989).

The Chorion Promoters

As mentioned above, comparisons of the chorion promoter families have shown that a great deal of divergence exists between early, middle, and late classes both at the primary level and with respect to the sequence motifs contained therein. While there appear to be a number of conserved motifs between the middle and late classes, this conservation is not apparent in the early chorion promoters (Hibner et al, 1988). This situation is further complicated by the structure of the bidirectional promoters themselves, as many of these presumptive regulatory elements demonstrate directionality and are found only at one end of the two promoter halves. These features suggest that the regulatory elements themselves may not be as important as the secondary and tertiary structures they can create upon interaction with the appropriate trans acting factors. This idea is supported by the finding that the mutation of single base pairs within these elements prevents the

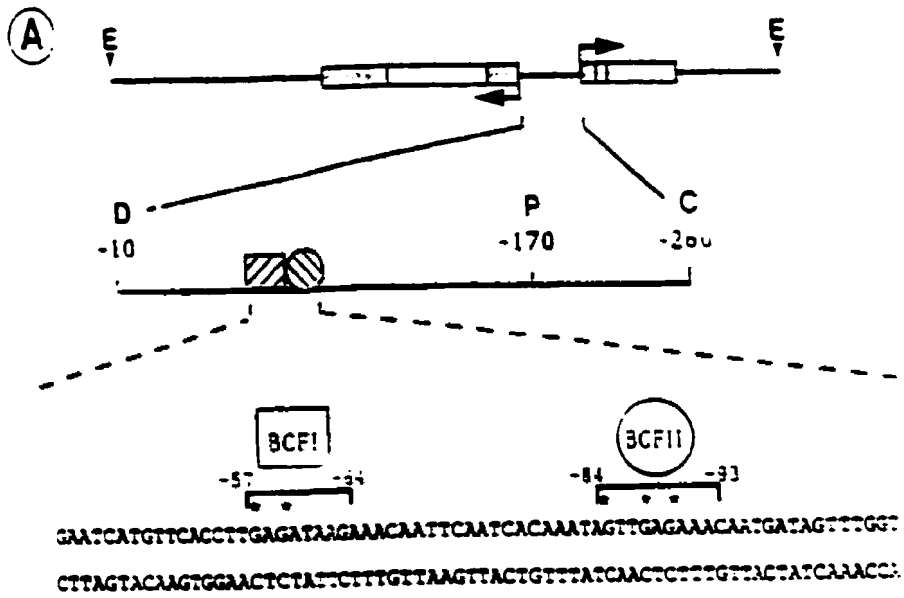
Figure 2 : Chorion Promoter and Expression Cascade

Panel A: Bidirectional chorion promoter structure and sequence motifs found in the *B. mori* late chorion gene pair HcA/B.12. The binding sites for two major transcriptional activators BCFI (BmGATAß) and BCFII (uncloned) are shown at the bottom of the diagram. Figure taken from Iatrou et al., JBC (1994).

Panel B : Developmentally regulated expression characteristics of three major chorion locus types: m2G12 (early), ABL.12, ABL.11 (middle), and m2574 (late). The numbers at the top of the figure indicate choriogenic stages. Figure taken from Kafatos et al, JMB (1986).

Chorion Promoter : Structure

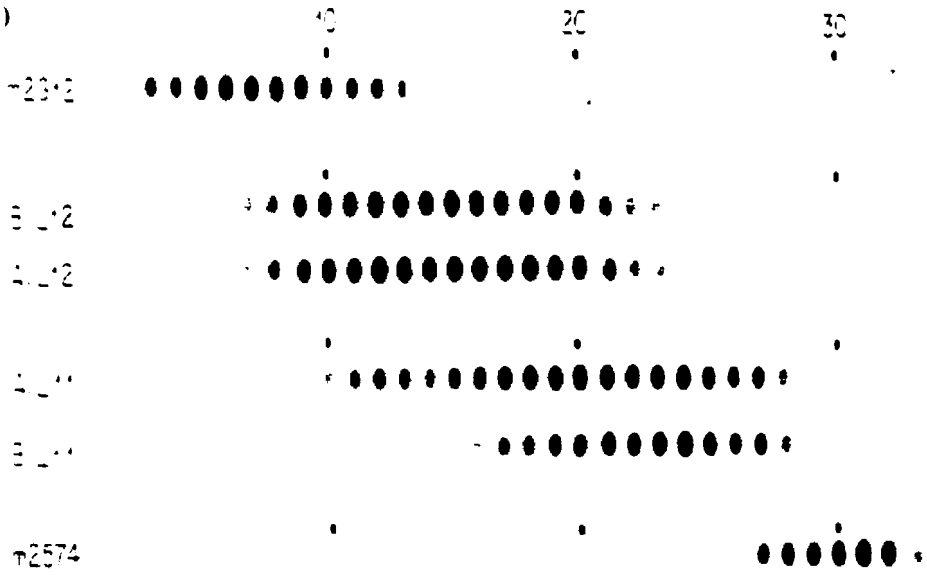
(A)



From : Iatrou et al. JBC (1994)

Chorion Expression Cascade

(B)



From : Kafatos et al. JMB(1986)

appropriate expression of both members of a gene pair (Spoerel et al, 1993), as well as the observation that some of the late (high cysteine) chorion gene promoters do contain regions of dyad symmetry that are capable of forming cruciform structures *in vitro* (Iatrou and Tsitilou, 1983).

Consensus and maintained motifs within the middle and late chorion promoters (Figure 2a) include a highly conserved TCACGT hexamer identified as the binding site for the *Bombyx* homologue of the *Drosophila* Ultraspiracle gene product (Tzertzinis et al., 1994), which encodes a nuclear hormone receptor. Several canonical A/TGATAA/G hexamers are found within the middle and late chorion gene promoters, but also appear in the early chorion gene promoters in non representative positions and orientations (Hibner et al., 1988). Finally, most of the middle and late promoters also contain a conserved TTG(T/A)GAAA octamer within close proximity to the GATA binding site (Skeiky and Iatrou, 1991); the identities of factors associating with this element have yet to be elucidated.

Isolation and Characterization of BmGATAß

Previous analyses using the late chorion HcA/B.12 promoter and follicle cell extracts identified all of the consensus motifs noted above, and suggested the formation of higher order complexes involving multiple trans acting factors (Skeiky and Iatrou, 1991). DNAse I footprinting and methylation interference experiments were able to identify at least two primary binding sites for transcriptional modulators (Figure 2a), with the higher order complexes presumed to arise through interactions between these factors and the shared promoter DNA. Factor BCFI was shown to bind to the A/TGATAA/G motif noted above, and BCFII bound to the octamer immediately downstream. Gel retardation experiments also suggested that the stable binding of BCFII was dependent on the preformation of a promoter/BCFI complex. While the characterization of the factors responsible for the formation of these complexes remains incomplete, these results

demonstrate that expression from the late chorion promoters is accomplished (at least in part) through transcriptional coactivation involving a number of distinct modulators.

Following the identification and demonstrated usage of canonical GATA binding motifs within these late chorion gene promoters, screening of a follicular cell derived genomic library with primers recognizing the highly conserved zinc finger region provided at least two candidates (Drevet et al., 1994). Both of these genes were found to encode zinc finger transcription factors (BmGATAα and BmGATAß), but only BmGATAß was demonstrated to be expressed in ovarian tissues over the developmental periods examined (Drevet et al, 1994). The confirmation of BmGATAß as factor BCFI was demonstrated by bandshift of a GATA site containing oligonucleotide with a bacterially expressed ß-gal:BmGATAß fusion protein, and by supershift experiments using follicular extracts and a polyclonal antibody directed against the zinc finger domain. The exonal arrangement and primary isoform structure of the BmGATAß factor is diagramatically illustrated in Figure 3a. Subsequent experiments demonstrated that BmGATAß is not expressed until immediately prior to the onset of choriogenesis (developmental stage -6), is downregulated during early choriogenesis, but continues to be expressed in low abundance throughout the subsequent stages (Drevet et al., 1995). Tissue specific analyses indicated that BmGATAß is expressed in most tissues during pupal development, but appears to demonstrate isoform restriction in the gonads (discussed below).

GATA Targets and Modulation in Alternate Systems

Over the past 10 years, GATA factors have emerged as hallmarks for the commitment of cells to developmental or differentiation pathways. The expression of these transcriptional activators has been implicated in a number of disparate systems including GATA-2 in *Xenopus* oogenesis (Brewer et al., 1995), GATA4/5/6 in vertebrate hematopoiesis (Elefanty et al., 1996), elt-1/end-1 in *C. elegans* endoderm/mesoderm

Figure 3 : Structure of BmGATAß

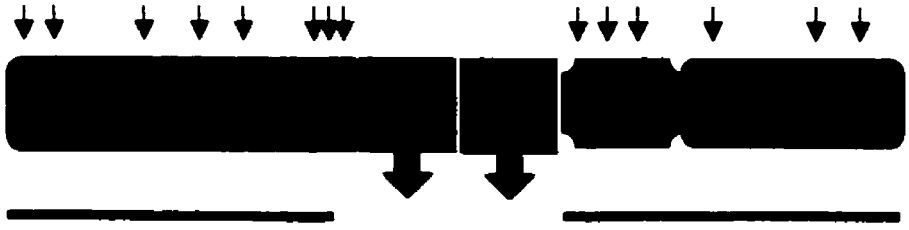
Panel A : Exon arrangement for BmGATAß gene. 5 primary N-terminal exons, 2 zinc finger exons, and 2 C-terminal exons are spliced together to give the general structure shown in B.

Panel B : Structure of the BmGATAß2 isoform used in this thesis. Arrows above indicate sites of potential phosphorylation. The bars below indicate the approximate size and locations of the two deletions used as bait in the two hybrid screens, and are not to scale.

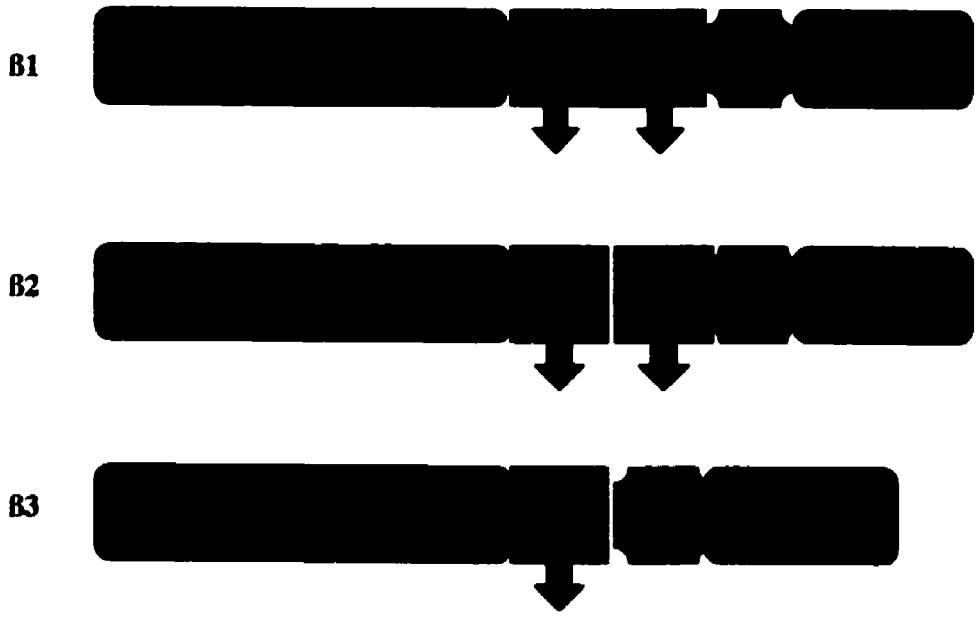
Panel C : Structures of the three major isoforms of the BmGATAß transcript.

(A)

(B)



(C)



differentiation (Zhu et al., 1998), and flower development in *Arabidopsis thaliana* (Putterill et al., 1995). Although these examples represent very distant developmental systems, the high degree of sequence and structural maintenance demonstrated by their associated GATA factors, as well as their conserved promoter target sequences, argue that similar regulatory mechanisms may be conserved as well.

Two potentially conserved regulatory mechanisms that may be pertinent to the proposed model for BmGATAß activity during choriogenesis are cellular localization and transcriptional coactivation. Regulation of GATA-2 during *Xenopus* oogenesis has been shown to involve cellular restriction; maternal GATA-2 appears to remain nuclear during the course of early oogenesis, and sequestered into non-transcriptionally active nuclear bodies (Patient et al., 1997), while the zygotically derived products remain restricted to the cytoplasm until interaction with the transcriptional modulator CBTF (CCAAT Box Transcription Factor) forces nuclear translocation of the complex (Orford et al., 1998). Several other examples of GATA-specific transcriptional modulators exist in the literature, including positive associations such as GATA-1:FOG and GATA-1:CBP (Orkin et al., 1998), and negative interactions as with GATA-4:FOG2 (Svensson et al., 1999) and Pannier:Ush in *Drosophila* (Haenlin et al., 1997). It is interesting to note that in many of these systems, the GATA factors appear unable to initiate transcription of their target genes in the absence of the cited transcriptional partners.

Protein Phosphorylation and Isoform Diversity of BmGATAß

Analysis of the BmGATAß gene product over the choriogenic period indicated that the GATA binding activity and gel migration characteristics of this factor change as choriogenesis proceeds (Skeiky et al., 1995). Cytoplasmic extracts appeared to accumulate a modified form of BCFI (cBCFI) during early choriogenesis, with an apparent shift back to the native BCFI during the period of middle/late chorion gene expression. Furthermore, treatment of cytoplasmic extracts with CIP equalized the

binding affinities of cBCFI and BCFI, indicating that the difference between the two forms was the result of phosphorylation. These results suggested a model of transcriptional regulation whereby phosphorylation leads to the cytoplasmic retention of a latent (transcriptionally inactive) BmGATAß until middle/late chorion gene expression is required.

Phosphorylation of transcription factors as a means of altering cellular localization and/or transcriptional activity has been noted in a number of systems, including SWI5 in yeast (Jans and Hubner, 1996), the yeast GATA factor GLN3 (Beck and Hall, 1999), and GATA1 in human erthyroleukaemia cells (Parthington and Patient, 1999). Indeed, phosphorylation of GLN3 and the yeast zinc finger protein MsN2p in yeast by the TOR signaling pathway (GLN3) or PKA (MsN2p) appears to result in the cytoplasmic localization and retention of these factors in manners very much akin to that suggested for BmGATAß (Beck and Hall,1999; Gorner et al., 1998).

There are, however, inconsistencies with respect to BmGATAß localization and activity suggested by the above model. First, in the examples noted above, phosphorylation and cytoplasmic retention are used in response to extracellular stimuli and result in down -regulation of factors already present in the nucleus; there does not appear to be any significant impetus for the *Bombyx* follicular epithelium to downregulate the activity of a developmentally regulated transcription factor (ie: what is the purpose of synthesizing a factor and sequestering it prior to its utility?). More notably, however, recent work which focused on characterizing BmGATAß subcellular localization failed to demonstrate significant cytoplasmic levels of BmGATAß at any point following its synthesis (Lunke, 1999). This work also clearly showed the presence of BmGATAß in the nucleus throughout choriogenesis, indicating that it may play a greater role in the transcription of earlier chorion genes than was previously thought. Although BmGATAß appears confined to the nucleus, it does not appear to be randomly distributed throughout the chromatin and, in some instances, the staining appears speckled and peri-nuclear (Lunke, 1999). This had also been observed for the murine GATA-1 transcription factor

(Elefanty et al., 1996), but no functional purpose has been found. Finally, extensions to the mobility shift and promoter affinity assays noted earlier for BmGATAß (Skeiky and Iatrou, 1991; Drevet et al., 1994) testing the abilities of both native and modified forms to bind early and early middle chorion gene promoters have not been reported. Until a better understanding of the role of BmGATAß and its phosphorylation in the transcription of all of the chorion genes is achieved, a resolution to these conflicting results cannot be properly addressed.

The final point to be introduced concerns the nature and function of BmGATAß isoforms (figure 3c). Screening of follicle specific cDNA libraries indicated the presence of at least three primary isoforms (ß1, ß2, ß3) which differ from each other only with respect to the DNA binding zinc finger domain (Figure 3c; Drevet et al., 1995). The ß2 isoform contains an additional 14 amino acid stretch within the spacer regions between its zinc fingers, and ß3 completely lacks the N-terminal zinc finger domain. The presence of the additional stretch of hydrophobic amino acids in the ß2 isoform would very likely result in changes to its DNA and/or promoter affinities. The loss of the N-terminal zinc finger would probably result in changes to possible protein:protein interactions since this finger is not required for binding of most bifingered GATA proteins to their cognate promoter sequences (Yang and Evans, 1992) and has been implicated in protein interactions in alternate systems. Tissue specific Northern analysis for these isoforms indicates that the primary ß2 isoform is expressed in several tissues during pupal development, whereas ß1 appears to be gonad specific, and ß3 expression shows restriction to the gonads and fat body (Drevet et al., 1995). All of these isoforms are present in the ovary over the expression period of BmGATAß, but there appears to be a shift in isoform choice to ß1 during late choriogenesis (perhaps in parallel with the expression of the late chorion genes).

Recent examinations of the N-terminal end of BmGATAß using 5' RACE identified three new potential isoforms that appear to maintain the ß1/ ß2/ ß3 structures, but differ

in that the N2 exon appears to be removed (Figure 2a; Lunke, 1999). The significance of these N-terminal deletions has not been addressed, but would be expected to result in changes to the protein conformation of the products; as this region is not directly involved in promoter binding, we assume that these changes alter BmGATAßs ability to interact with protein factors modulating its transcriptional activity and/or localization. N-terminal isoform specific differences in C_2H_2 type zinc finger transcription factors have been reported elsewhere (Andrews et al., 2000) involving the *Drosophila* ovo˜ gene products necessary for germline and epidermis differentiation; these OVO isoforms maintain their DNA binding domain identities, but appear to act antagonistically at their designated promoters. An analysis of the developmental regulation and tissue specificities of the BmGATAß N-terminal isoforms might indicate that similar mechanisms are used in the *Bombyx* follicular epithelium to regulate chorion gene expression.

A Model for Control of BmGATAß Activity

As noted above, encapsulating all of the relevant data concerning BmGATAß expression, localization, and activity has proven to be a challenge. While the purpose of the proposed isoforms and posttranslational modifications have not been elucidated, most of the prior work on this factor are presumed to be sound. Thus, any pathway or model generated must account for all of the following observations:

1) BmGATAß expression is induced in response to "choriogenic commitment", and may autoregulate its own synthesis.
2) Several isoforms are generated in a developmentally regulated fashion differing primarily in the DNA binding domain, and thus may affect promoter affinities or interactions.

3) The BmGATAß products are post translationally modified by phosphorylation; the GATA-specific kinases and residues involved have not yet been identified.
4) This phosphorylation results in BmGATAß interaction with auxiliary factors; the identification (or function) of these factors has not been determined.
5) Activation of the middle/late chorion genes requires the presence of a secondary transcriptional modulator. BmGATAß involvement in the transcription of earlier chorion genes (or other targets) has not been determined.
6) The overlapping gradients of expression noted for the chorion genes implies both the ability to activate and repress transcription in a highly regulated and coordinated way. This may in turn imply additional promoter-specific factors that are not addressed by transcriptional coactivation.

Given these observations it is immediately apparent that a large number of factors will be involved in the genesis and control of this complex transcriptional system. These would include (but are not restricted to):

1) Transcriptional cofactors acting at the BmGATAß promoter.
2) Splicing factor associations modifying isoform specificity
3) Translation-associated proteins and chaperones.
4) Cytoplasmic anchoring / nuclear shuttle proteins.
5) BmGATAß specific kinase(s).
6) BmGATAß phosphatase(s).
7) Chorion promoter specific transcriptional coactivators/repressors.
8) Factors controlling as yet undefined BmGATAß activities.

Objectives of this Thesis

The characterization of the BmGATAß gene, transcript, and product have been a continuing focus of this laboratory for almost 10 years. In that time, several observations

(detailed above) suggest the existence of a large number of BmGATAß specific and non-specific factors that result in the aforementioned changes. As of yet, none of these factors has been identified or confirmed, and the realization of a consistent and experimentally testable model for BmGATAß activity depends on their isolation and characterization. The purpose of this thesis is the isolation of some of these associated factors through the use of the yeast two hybrid system, and confirmation of the potential of these factors to interact with BmGATAß through the analysis of their patterns of expression. This work will also seek to begin the initial stages of characterization for these isolated factors, and attempt to demonstrate similar types of interaction *in vitro*.

The initial stages of this work have focused on the development of BmGATAß related baits for use in screening an available yeast two hybrid cDNA library. This was followed by the isolation and confirmatory testing of the recovered preys within the yeast system. Preys surviving this selection were then analyzed for their expression characteristics through the use of developmental and tissue specific Northern analyses and *in situ* hybridization. Screening for and recovery of full length cDNAs from lambda gt11 and 5' RACE libraries was pursued to complement these analyses, and attempts were made to verify the observed interactions in yeast through *in vitro* studies using GST pull down experiments.

SECTION II : MATERIALS AND METHODS

Moths

The animals used in this study were a *B. mori* hybrid strain that was obtained from the Forest Pest Management Institute (Sault Ste-Marie, Ontario). The animals were reared at 25 ^{o} C and fed with mulberry leaf artificial diet (Yakuroto Company, Japan).

Preparation of Electrocompetent *E. coli*

A 5 mL overnight culture of *Escherichia coli* was grown at 37 ^{o} C in LB medium (10 g/L tryptone; 5 g/L yeast extract; 5 g/L NaCl), and used to inoculate 300 mL LB that was then allowed to grow until the OD 600 reached 0.5 (approximately 5 hours). Cells were chilled on ice, centrifuged in a GS-3 rotor at 2700 rpm for 5 minutes at 4 ^{o} C, resuspended in 300 mL ice cold 10% glycerol (in water) and kept on ice for 20 minutes. The cells were pelleted at 4000 rpm in a GSA-3 rotor for 10 minutes, and resuspended in 30 mL ice cold 10% glycerol. Following a 20 minute incubation on ice, the cells were again pelleted, and finally resuspended in 1200 ìL of 10% glycerol. 100 ìL aliquots were distributed into eppendorf tubes on dry ice and the suspension was stored at - 70 ^{o} C until use.

Plasmid DNA Preparations

Plasmid Micropreps : Supercoil Screening

Screening of Exonuclease III deletions was accomplished by direct analysis of plasmid products prior to purification. Two hundred ìL of overnight cultures were spun at 3500 rpm for 2 minutes and resuspended in 40 µL STE (100mM NaCl; 20 mM Tris-HCl, pH 7.5; 10 mM EDTA). Forty µL of phenol/choloroform/isoamyl

alcohol (25:24:1) was added, the mixture was vortexed, and spun at 14,000 rpm for 1 minute in a microcentrifuge. The supernatant was transferred into a separate tube containing 1 µL of RNase A (10 mg/mL) and incubated at room temperature for 5 minutes. Twenty µL of this was analyzed by electrophoresis on a 1% agarose gel, and the migration of the supercoiled products was compared to that of controls.

Minipreps: Alkaline Lysis

Three to five mL of bacterial culture was grown overnight at 37 ^{o} C and pelleted by centrifugation at 3500 rpm (Eppendorf microcentrifuge) for 2 minutes. The supernatant was aspirated away, and the pellet was resuspended in 100 µL ice cold solution I (50mM glucose, 25 mM Tris HCl, 10 mM EDTA). Two hundred µl solution II (0.2 M NaOH, 1% SDS) was added and the suspension was mixed by inversion. One hundred and fifty µL ice cold solution III (5 M potassium acetate, 11.5% glacial acetic acid) was added, mixed by repeated inversion, and the suspension was centrifuged for 5 minutes at 14,000 rpm (Eppendorf microcentrifuge). The supernatant was transferred to a separate tube, extracted once with phenol/choloroform/isoamyl alcohol, and the DNA was ethanol precipitated.

Maxipreps: CsCl Ultracentrifugation

A 4 mL overnight bacterial culture was used to inoculate 400 ml of selective medium (LB-ampicillin), and was grown at 37 ^{o} C to an O.D. of 0.6 to 0.7 overnight. Bacteria were collected by centrifugation (4000 rpm in a Sorvall GS3), and subjected to alkaline lysis (Sambrook et al, 1989) in the presence of 2 mL lysozyme (10 mg/mL). Following filtration through 2 layers of Kimwipe, the crude DNA was precipitated with 0.6 volumes of isopropanol, collected by centrifugation at 5000 rpm (Sorvall RC-5B centrifuge), washed with 70% ethanol, and redissolved in 3.2 mL TE (1M Tris-EDTA, pH 8.0). CsCl was added to 1 g/mL, and 80 µL of

10 mg/mL EtBr was added to every mL of this DNA/CsCl solution. Following centrifugation in a clinical centrifuge to remove precipitates, the cleared solution was transferred to 3.9 mL Quick-Seal tubes, and centrifuged overnight at 100,000 rpm in an ultracentrifuge (TL-100 Tabletop ultracentrifuge; Beckman near vertical rotor). The purified band was removed with a needle, and repeatedly extracted with water saturated n-butanol to remove the EtBr. The plasmid was then ethanol precipitated, washed in 70% ethanol, and resuspended in TE.

Gel Purification of DNA Fragments

Freeze and Squeeze Method

Small restriction fragments (<1kb) were resolved by 1% agarose gel electrophoresis and the band of interest was cut out using a sterile razor blade. The gel slice was placed into a 0.5 mL punctured eppendorf tube plugged with glass wool, and placed on dry ice for 10 min. The tube was spun at 12000 rpm for 2 minutes (Eppendorf microcentrifuge), and the recovered solution was phenol/choloroform extracted and precipitated in ethanol.

DEAE Cellulose

Following separation on a 1% agarose gel, a small incision was made ahead of the desired fragment and a small square of prewet DEAE cellulose (NA45 membrane; Mandel Scientific) was fed into the gap. Electrophoresis was continued until the majority of the DNA was stuck to the membrane. The fragment was washed with low salt buffer (150 mM NaCl; 0.1 mM EDTA; 20 mM Tris-HCl, pH 8) and eluted with high salt buffer (1.0 M NaCl; 0.1 mM EDTA; 20 mM Tris-HCl, pH 8) at 60 $^{\circ}$C for 15 minutes. The eluate was extracted with an equal volume of

phenol:chloroform:isoamyl alcohol (25:24:1) and the DNA was ethanol precipitated.

Electroelution

Following separation on a 0.8 to 1.2% agarose gel in TAE (4.84 g/L Tris Base; 0.744 g/L Na_2,EDTA $2H_20$; 1.14 mL/L glacial acetic acid), a thin slice of the gel containing the desired fragment was excised and placed into 400 µL TAE in a Spectra/Por molecularporous membrane (Spectrum Laboratories; 6000-8000 Molecular Weight Cut Off). Electrophoresis was continued until the band had migrated out of the gel slice. The buffer was removed from the membrane, extracted with an equal volume of phenol:choloroform:isoamyl alcohol (25:24:1) and the eluted DNA fragment was ethanol precipitated.

Exonuclease III Deletions

Template Preparation

Full length BmGATAß cDNA was obtained from clone 2-3a (produced by M. Lunke, 1999) by *Not I* digestion, and end filled with T4 DNA polymerase as per manufacturers protocols. This blunt ended fragment was cloned into a blunt ended *SpeI* site in pBluescript SK+ (Stratagene), and ligation products in both orientations were identified and recovered by restriction analysis and sequencing. Exonuclease III activity is specific for 5' overhangs and will not digest DNA ends with blunt ends or 3' overhangs. Therefore in order to produce a template that would be specifically digested from the inserted BmGATAß side and not from the vector

sequences, 5' overhangs at the 3' ends of the BmGATAß inserts were created using *BamHI* and 3' overhangs protecting the vector were created by digestion with *PstI*.

Exonuclease III Reaction

Fifty µg linearized template was suspended in 200 µL reaction solution containing 20 µL 10X buffer (660 mM Tris-HCl; 66 mM $MgCl_2$; 50 mM DTT; 500 µg/mL BSA), 20 µL 2-mercaptoethanol, and 100 U Exonuclease III. The reaction was incubated at 27 ^{o} C and a total of ten 25 µL aliquots were removed at 1 minute intervals and placed into 175 µL of 1X Mung Bean Buffer (50 mM sodium acetate; 30 mM NaCl; 1mM $ZnSO_4$, pH 5.0) on dry ice. Following the removal of the last sample, the reactions were heated to 68 ^{o} C for 15 minutes, 15 U of Mung Bean Nuclease was added to each tube, and incubation was continued for 30 minutes at 30 ^{o} C to create blunt ends. To each tube 4 µL 20% SDS, 10 µL 1M Tris-HCl (pH 9.5), 20 µL 8M LiCl, and 250 µL phenol:chloroform:isoamyl alcohol (25:24:1) were added sequentially. Following extraction, the products were precipitated on dry ice with 1/10 volume sodium acetate and 2.5 volumes of 95% ethanol.

Ligation and Recovery of Exonuclease III products

The deletion products were resolved on a 1% agarose gel, and the high molecular weight bands were excised and recovered by electroelution. Following extraction and precipitation, the blunt ended pBS-BmGATAß products were ligated at 4 ^{o} C overnight in a 20 µL reaction containing 1 mM ATP and 2U of T4 DNA ligase. Two µL aliquots of these ligations were then used to electroporate 20 µL of electrocompetent *E.coli* HB101.

Northern Analysis

Preparation of RNA

Preparation of total RNA from follicular cells was accomplished as follows: Staged follicles (in groups of 5) were dissected in Hunsley's medium (4 mM NaCl; 26 mM KCl; 18 mM $MgCl_2$; 3 mM $CaCl_2$; 250 mM sucrose; 10 mM potassium phosphate buffer, pH 6.6), and snap frozen on dry ice. One half of one mL of Trizol reagent (Gibco, BRL) was added and the follicles were allowed to thaw for 2 minutes at room temperature with occasional inversion. RNA was subsequently purified from this extract following the manufacturers protocols.

Total RNA from selected tissues was prepared as above, except that the tissues were subjected to manual disruption in Trizol reagent using a plastic pestle.

Gel Analysis and Blotting

For both developmental and tissue specific Northerns, 5 µg of total RNA was loaded per lane onto a 1.5% agarose/formaldehyde gel. Following electrophoresis (Bio-Rad Mini Sub Gel System), the gel was washed for 10 minutes in double distilled H_2O, and the RNA was transferred onto Hybond N membranes by capillary transfer with 3 mM NaOH. The filters were washed in 2X SSC (0.3 M NaCl; 0.03M Na_3-citrate) and baked for 2 hours at 80 ^{o} C prior to hybridization.

Probe Preparation

cDNA probes for developmental and tissue specific Northern analysis were created as follows. Twenty five ng of template cDNA (in 10 µL water) was denatured by boiling for 5 minutes, and added to the following mixture: 4 µL 5X OLB (see below), 0.5 µL Bovine Serum Albumin, 2.0 µL α-^{32} P dCTP, 1.0 µL

Klenow polymerase at 1 unit/μL, 2.5 μL H_2O. This was incubated at 37 ^{o} C for 1 hour, purified through a spin column (Sephadex G50), and precipitated overnight in ethanol with 10 % sodium acetate and 10 μg of single stranded salmon sperm DNA as carrier.

The 5X OLB buffer contains 2 volumes of Solution A, 5 volumes solution B and 3 volumes solution C: Solution A (1.0 mL of 1.25 M Tris-HCl, 18 μL 2-mercaptoethanol, 5 μL each 100 mM dNTPs (dATP, dGTP, dTTP)). Solution B (2 M Hepes, pH 6.6). Solution C (Hexanucleotides in 3 mM Tris-HCl, 0.2 mM EDTA at 90 OD units per millilitre.)

Hybridizations

The membrane containing immobilized total RNA was sealed into a plastic hybridization bag, and 10 mL of hybridization buffer was added (0.3 M NaCl, 50 mM Na-phosphate, pH 7.0, 12.5 X Denhardt's solution, 10% dextran sulfate, 1% SDS, 5 mM EDTA, 5 mg/mL yeast RNA). This was allowed to prehybridize overnight at 65 ^{o} C, and the labeled probe was boiled to denature and added to an activity of 1 x 10^{6} cpm/mL (Cerenkov). Following hybridization overnight at 65 ^{o} C, the filter was removed and double washed each time for 10 minutes in 2X SSC, 1X SSC, and 0.2X SSC. The filter was mounted onto 3 MM paper, and autoradiographed overnight at - 70 ^{o} C.

Yeast Two Hybrid System

Except as noted below, all reagents, materials, and yeast techniques are as per manufacturers protocols (CLONTECH MATCHMAKER™ Two Hybrid System).

Host strain, vectors, cDNA libraries.

The host strain used for all cloning and library screening was *Saccharomyces cerivisiae* HF7c (MATa, ura3-52, his3-200, lys2-801, ade2-101, trp1-901, leu2-3, 112, gal4-542, gal80-538, LYS2::GAL1-HIS3, URA3::(GAL4 17mers)3-CYC1-lacZ).

Bombyx mori follicular cell derived cDNA libraries were assembled directionally into the Two hybrid pGAD424 activation domain vector (CLONTECH) by Dr. Kenichi Ito; both vitellogenic and choriogenic specific libraries were used for screening. The libraries were constructed using 10 µg each vitellogenic specific or choriogenic specific poly (A)+ RNA that had been size fractionated by gel purification to remove transcripts less than 1.0 kb in length. This is particularly important for the choriogenic specific library as the saturating chorion transcripts are usually less than 1.0 kb, and this procedure enriches for non-chorion transcripts.

Following second strand cDNA synthesis, the products were digested with SalI or BamHI to aid in directional cloning into the two hybrid activation domain vector pGAD424 (Clonetech). The complexity of the vitellogenic library was found to be in the order of 16 X 10^6 clones, with insert sizes greater than 1.0 kb. The double digested choriogenic cDNA library was found to have a complexity of approximately 5 X 10^6 clones, with insert sizes also greater than 1.0 kb.

Deletions of BmGATAß used as bait in screening the aforementioned library were cloned as EcoRI/SalI fragments into a pGBT9 DNA binding domain vector (CLONTECH).

Yeast Two Hybrid Library Screens

A combined vitellogenic/choriogenic cDNA library was screened with the N-terminal portion of BmGATAß (encompassing the majority of the N-terminal domain; Figure 5a). The bait was transformed into HF7c, and following overnight growth to a density of 0.4 (A600), these transformants were again transformed with 200 µg of vitellogenic-specific and 300 µg of choriogenic-specific cDNA libraries. After a 2 hour recovery incubation, these cells were washed and resuspended in 10 mL TE buffer. 0.4 mL of this was plated onto each of 20 150 mm SD agar plates (negative selection minus -his-trp-leu) containing 0.02 M 3-aminotriazole (to repress histidine biosynthesis). These plates were then incubated for 6 days at 30 ^{o} C until the colonies were sufficiently large for nitrocellulose lifts.

Library screens for the isolated C-terminal portion of BmGATAß were conducted essentially as above, except that 500 µg of the choriogenic specific cDNA library was used for screening, and the plates were left to grow at 30 ^{o} C for a total of 8 days following transformation.

Colour Reactions

Yeast transformants were grown on the appropriate selection medium until large, and then scraped and painted onto sterile 100 mm Whatman filter paper. Following recovery overnight on YP agar, the filters were lifted and submerged in liquid nitrogen for 15 seconds. These frozen filters were then allowed to thaw before being placed onto another 100 mm Whatman filter presoaked in Z buffer/X-gal solution (Na-Phosphate, 8.53 g/L; Na-hydrogen-phosphate, 4.78 g/L; KCl, 0.75 g/L; Magnesium sulfate, 0.246 g/L; X-gal 0.3 mg/mL; β-mercaptoethanol, 0.027 mL/L).

X-gal is : 5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside. These filters were allowed to develop overnight in the fume hood.

In Situ Hybridization

Follicles for *in situ* hybridization experiments with NTP 1-4 and CTP 15-1 were obtained from day 8 pupae; ovarioles were dissected in Hunsley's medium, and pools of follicles were desheathed and washed twice in PBST (1X Phosphate Buffered Saline, 0.3% Triton X-100) prior to fixing in 3.7% formaldehyde in PBST for 20 minutes at room temperature. Fixed follicles were washed three times in PBST and stored at 4 $^{\circ}$ C until hybridization.

Probe Preparation

EcoRI/BamHI fragments of NTP 1-4 and CTP 15-1 were obtained from the original yeast two hybrid prey vectors and directionally cloned into pBluescript SK+. These plasmids were then linearized with *XbaI* and *XhoI* to produce templates for T3 and T7 RNA polymerase in *vitro* transcription reactions respectively. *In vitro* transcriptions were done at 37 $^{\circ}$ C for 6 hours in 50 µL reaction volumes containing 1X polymerase buffer (40 mM Tris, pH 8; 10 mM $MgCl_2$; 10 mM DTT), 100 µM NTPs (65 µM ATP/CTP/GTP, 35 µM UTP mix (1:3 mixture of dig-11 UTP/UTP respectively), 5 U RNAguard (Amersham/Pharmacia). Template DNA was removed by addition of 2U RQ1 RNase free DNAse and incubation for 15 minutes. Products were first examined on a 2% agarose gel, and aliquots were fragmented in 0.2 M carbonate buffer (80 mM $NaHCO_3$, 120 mM Na_2CO_3) for 12 minutes. These digests were checked on a 2% agarose gel to assess the degree of fragmentation, size, and quantity, and then ethanol precipitated.

Tissue Preparation and Hybridization

The basal lamina of the follicles was disrupted with the addition of 50 µL of 50 mg/mL collagenase dissolved into 5 mL PBST and incubation at room temperature for 20 minutes. Following three washes with PBST, and one wash with 5 mL proteinase K buffer (100 mM Tris, pH8; 50 mM EDTA), proteinase K was added to a concentration of 1.5 µg/mL and the follicles were incubated at 37 $^{\circ}$ C for 30 minutes. After two double washes with PBS, the follicles were refixed in 3.7% formaldehyde in PBS for 20 minutes, then washed twice in PBS. Charge neutralization was done for 10 minutes in 5 mL trithethanolamine buffer (1.3% vol/vol) containing 12.5 µL acetic anhydride. The follicles were then washed 4 times in PBS prior to prehybridization.

The follicles were transferred to 3 mL hybridization tubes, each of which contained 1 mL prehybridization solution (50% formamide; 1X Haase mix (10 mMTris; 1 mM EDTA; 1X Denhardt's); 1 mg/mL sheared salmon sperm DNA; 100 µg/mL polyribo-adenylic acid; 20 mM DTT; 0.25 M NaCl). Prehybridization was carried out at room temperature for 1 hour. The prehybridization solution was removed and replaced with 200 µL hybridization buffer (80% prehybridization buffer; 20% of: 50% dextran sulfate in 50% formamide, 0.6 M NaCl), containing 100 ng of the dig-11 RNA probe that had been added and heated at 85 $^{\circ}$ C for 5 minutes prior to replacement. Hybridization was done overnight at 50 $^{\circ}$ C.

Washes and Antibody Detection

The hybridization solution was removed and the follicles were double washed in wash buffer {1X salts buffer (0.3 M NaCl; 5 mM NaH_2PO_4 ;100 mM Tris-HCl, pH 6.8; 50 mM EDTA, pH 8.0), 50% formamide}. The non-hybridized portions of the

probes were removed with the addition of 500 µL NTE buffer (0.5 M NaCl; 10 mM Tris, pH 8; 1 mM EDTA) containing 1U RNase T1 and 10 µg RNase A. Following incubation at 37 °C for 30 minutes, the follicles were double washed in increasing stringency SSC (2X; 1X; 0.1X) for 15 minutes at 50 °C. They were then double washed twice in PBST, 100 µL of 2% horse serum was added and the solution was preincubated at room temperature for 45 minutes. One hundred µL of 1/100 anti-DIG antibodies coupled to HRP (Boehringer) was finally added and incubation continued overnight at 4 °C.

Following incubation with the antibodies, the follicles were washed 6 times for 10 minutes with PBST, and transferred to 0.5 mL DAB solution (10 µL of 200 mg/mL β-D-glucose, 10 µL 40 mg/mL ammonium chloride; 50 µL of 10 mg/mL diaminobenzidine). After 15 minutes of incubation in the dark, 3 µL of glucose oxidase was added and the colour reaction was allowed to proceed for 5 and 15 minutes for NTP 1-4 and CTP 15-1 respectively. Reactions were halted with 5 changes of PBS. The follicles were then dehydrated in an ethanol series (30%, 50%, 70%, 95%, 100%) for 5 minutes each, followed by two washes in xylene for 10 minutes. They were then mounted in Permount prior to microscopy. Microscopes used for these analyses were the Zeiss Stemi 2000-C dissecting microscope, and a Zeiss III RS Light Microscope.

Screening of 5' RACE cDNA Libraries

In order to isolate full length cDNAs containing the 5' ends of the 3' fragments of clones isolated from the yeast two hybrid library screens, 5' RACE was performed for both NTP 1-4 and CTP 15-1. Gene specific primers used for the amplifications were:

NTP 1-4 GSP 1109: 5' - ACT CCT CAA ATT TGG ATT GTG CTG
NTP 1-4 GSP 1110: 5' - TGA CCC ACT GAC TGA CTA TTG TGA
CTP 15-1 GSP 1583: 5' - CGG CGA TTT GTT GTG CGG TCA TC

CTP 15-1 GSP 1584: 5' - CAA CTG ACT GCT ATC GAC TGG TG
RACE Primer AP1 : 5' - CCA TCC TAA TAC GAC TCA CTA TAG GGC

These primers were used to screen premade 5' RACE libraries (see below) specific for vitellogenesis (F. Kendirgi; 2000) and choriogenesis (M. Lunke; 1999) and their positions are shown in the sequences of the relevant clones (Appendix I). Fifty fold dilutions of these original RACE libraries were subjected to touchdown PCR (Marathon™ cDNA Amplification Kit; Clonetech Laboratories) using cycling parameters of 1 cycle 94 ^{o} C for 2 minutes; 5 cycles 94 ^{o} C (1 minute), 75 ^{o} C (30 seconds), 72 ^{o} C (2 minutes 30 seconds); 25 cycles 94 ^{o} C (1 minute), 72 ^{o} C (2 minutes 30 seconds). Reaction volumes and concentrations were as per manufacturers protocols. Generally, 50 µL reactions were done containing 36 µL H_2O, 5 µL of fifty fold diluted cDNA library, 5 µL 10X PCR reaction buffer (400 mM Tricine-KOH; 150 mM KOAc; 35 mM $Mg(OAc)_2$; 37.5 µg/mL Bovine Serum Albumin), 1 µL dNTP mix (10 mM each dATP, dTTP, dGTP, dCTP), 1 µL Advantage™ cDNA Polymerase mix (5 U/µL), and 1 µL each of 20 µM RACE AP1 and gene specific primers. Products were gel purified and cloned into the pBluescript SK+ vector for sequencing.

The 5' RACE libraries had been prepared from follicular/nurse cell poly(A)$^{+}$ RNA (F. Kendirgi; unpublished results) and follicle epithelium poly(A)$^{+}$ RNA (Dr. Ken Ito; unpublished results). The RNA had been reverse transcribed using an oligo$(dT)_{12-18}$ primer and made double stranded using manufacturers protocols (Clonetech). Adapters specific for the AP1 RACE primers were ligated onto these double stranded cDNAs to allow for PCR amplifications of the clones of interest.

5' Extension Strategy

In order to isolate sequences corresponding to the 5' ends of NTP 1-4 and CTP 15-1, a modified form of RT PCR was employed. In brief, 3 µg of choriogenic or vitellogenic specific poly $(A)^+$ RNA was subjected to first strand synthesis using the external 5' RACE primers 1110 and 1583 noted above. The RNA was added to a 20 µL reaction containing 4 µL 5X Superscript Buffer, 0.5 mM dNTPs, 4 mM sodium pyrophosphate, 200 ng gene specific primer, 10 U RNase inhibitor and 200 U Superscript RT polymerase (Gibco, BRL). The resulting cDNA:RNA hybrids were blunt ended with the addition of 2 µL RNAse A (10mg/mL) and ethanol precipitated following phenol/chloroform extraction. After resuspension in 10 µL H_20, 50 ng of EcoRV cut pBluescript SK+ (in 4 µL) was added along with 2U of T4 DNA ligase (Gibco, BRL) and 4 µL 5X DNA ligase buffer. Two µL aliquots of this ligation mixture were then PCR amplified using the internal RACE primers 1109 or 1584 and T3 primers.

GST Pull Down Assays

Cloning and Expression of GST-BmGATAß Fusion cDNAs

All BmGATAß clones used for yeast two hybrid screens were directionally cloned in frame as *EcoRI/BamHI* fragments into pGEX 5X-1 vectors. Following transformation into the *E. coli* host (DH10b), and overnight amplification in LB medium at 37 $^{\circ}$ C, minipreps of the fusion vector were made using the alkaline lysis protocol. Confirmation of inserts and orientations was accomplished by restriction analysis and sequencing. After verification of the inserts, these transformed hosts were assessed for fusion protein expression as per manufacturers protocols (GST Gene Fusion System; Pharmacia Biotech) with the noted exceptions: 3 mL cultures

were grown overnight, and 100 µL aliquots of each saturated culture was added to 3 mL fresh medium and allowed to grow for 4-5 hours at 25 °C. These cultures were induced with 100 mM IPTG and allowed to grow at 25 °C for an additional 60 minutes. For the full length GST-BmGATAß fusion, the incubation temperature was reduced to 22 °C, and induction was carried out for 90 minutes. In order to prevent insolubility created by manual denaturation, the expression cultures were spun down at 500 rpm (microcentrifuge) and resuspended in 0.5 mL of STE buffer (10 mM Tris-HCl, pH 7.5; 10 mM NaCl; 1 mM EDTA, pH 8) containing 1% Triton X-100 and 100 µg/mL lysozyme. The suspensions were allowed to sit at room temperature for 20 minutes prior to being sonicated for 2-3 seconds. The resulting supernatants containing the proteins were analyzed by SDS PAGE prior to scale up and purification.

Production of Large Scale Bacterial Sonicates

Bacterial transformants from the small scale cultures (above) were used to inoculate 300 mL of LB-ampicillin media. These cultures were grown at 23 - 27 °C until log phase, approximately 5 hours. Protein expression was induced with the addition of 1.5 mL of 100 mM IPTG (final concentration was 0.5 mM), and the culture was allowed to grow for an additional 2 hours. The cells were collected by centrifugation for 10 minutes at 8000 rpm in a Sorvall GS-3 rotor, resuspended in 1.5 mL 1X PBS, and sonicated on ice. Following the addition of Triton X-100 to a final concentration of 1%, the lysates were centrifuged at 12000 rpm in a microcentrifuge, and the resulting supernatants were collected into sterile eppendorf tubes. At each stage of this procedure, 20 µL aliquots were removed and placed on ice for subsequent analysis by SDS-PAGE.

Purification of GST Fusion Proteins

Fusion proteins from the large scale bacterial sonicates were purified using the Manufacturers Batch Protocols (Pharmacia Biotech). In brief, 1.0 mL aliquots of the sonicates were added to a 50% slurry of Glutathione Sepharose 4B. Following incubation at room temperature for 30 minutes, the slurry was centrifuged at 1000 rpm in a microcentrifuge and the supernatant was aspirated away. The matrix was then washed three times in 0.5 mL PBS, and the purified fusion protein was eluted with 100 µL glutathione elution buffer.

SDS PAGE

Five to ten µL of protein containing supernatants produced in *E coli* or by *in vitro* methodologies were heated to 95 ^{o} C following the addition of 2-3 µL SDS PAGE buffer (0.35 M Tris-HCl, pH 6.8; 10% SDS; 35% glycerol; 5% ß-mercaptoethanol; 0.01% Bromophenol Blue). These samples were then loaded onto 10 to 12% polyacrylamide gels, and following electrophoresis, fixed in 5% methanol/acetic acid and stained using Coomassie Blue G-250.

In Vitro Transcription

RNA templates used for *in vitro* translations were produced using the NTP 15-1 prey cloned into pBS; sense and antisense transcripts were produced from the T3 and T7 promoters respectively. Briefly, 1 µg of purified vector was cut with the appropriate restriction enzyme (*XbaI* for sense, *XhoI* for antisense) to completion. These templates were added to a 50 µL reaction containing: 5 µL 10X transcription

buffer (Promega: 400mM Tris-HCl; 60 mM $MgCl_2$; 20 mM spermidine), 100 µM NTPs, 2 µL BSA (10mg/mL), 5 µL 1mM DTT, 1 µL RNase Inhibitor and 2 Units of T3 or T7 polymerase. Following synthesis at 37 °C for 5 to 6 hours, 1 µL RQ1 RNAse free DNAse (1 unit per microlitre) was added to each reaction and incubation was continued for 30 minutes. The resulting product was phenol/chloroform extracted and ethanol precipitated.

In Vitro Translations

In vitro translations for use in GST pull down experiments were accomplished using the Promega Rabbit Reticulocyte Lysate Translation System™. Two hundred ng of the *in vitro* produced RNA was added to a 50 µL reaction containing 35 µL lysate, 1 µL 1mM amino acid mixture minus methionine, 1 µL RNAsin RNase inhibitor, and 2 µL ^{35}S -methionine (Amersham Pharmacia Biotech; 2.42 kBq/µL). Translations were allowed to proceed for 2 hours at 37 °C. Five µL aliquots were subsequently analyzed by SDS PAGE.

Pull Down Assays

Pull down assays were performed essentially as previously described (Ausubel, 1999). Twenty µL aliquots of the purified fusion protein bound to the Sepharose 4B matrix (above) were washed briefly in 1X PBS, and resuspended in a final volume of 100 µL. Five µL of the ^{35}S labeled *in vitro* translated preys were added, and the solution was incubated at 4 °C for 2 hours. The matrix was double washed six times with 500 µL 1X PBS, resuspended in 25 µL 1X PBS, and 5 µL of 5X SDS-PAGE loading buffer was added. The samples were boiled for 5 minutes, and directly loaded onto a 10% SDS-PAGE gel for analysis.

SECTION III : RESULTS

Analysis of Bait Potential Of BmGATAß

In order to assess whether or not the full length ORF of the BmGATAß2 cDNA (Figure 3b) could be used as bait for the screening of a follicular yeast two hybrid library, the full length product (1390 bp) was amplified using 5' and 3' specific primers that permitted the inclusion of 5' *EcoRI* and 3' *SalI* sites. Following sequence confirmation, this product was double digested and directionally cloned into the yeast two hybrid bait vector pGBT9. Following transformation into the yeast host strain (Hf7c), the transformants were tested both under histidine selection and in a standard colour reaction. It was determined that the full length fusion product could survive the selection, and was also capable of activating the β-galactosidase reporter in the absence of an activating partner (Figure 6b; controls Lane3). This promiscuous activity precluded its use as a potential bait for library screening.

To develop fusions that would remove this activity, two approaches were taken. First, following the lead of Tsang et al (1997) who used the zinc finger region alone for the isolation of the transcriptional coactivator FOG (Friend Of GATA), PCR primers were developed that permitted the recovery of the zinc finger domain of BmGATAß. This bait was shown to be non-activating (Figure 6b; controls Lane 1) and was used to screen a choriogenic specific yeast two hybrid cDNA library but, because a parallel series of experiments were already underway by another researcher in our laboratory, this line of investigation was not pursued further. The second approach involved the construction of BmGATAß deletions, with the goal of maintaining as much of the structure as possible, while removing the constitutive activating domains.

Exonuclease III Deletions

The full length cDNA for BmGATAß was blunt ended and cloned into pBluescript at a blunt ended SpeI site as described in Materials and Methods. Ligation products in both orientations were recovered, screened by restriction analysis, and confirmed by sequencing. These templates were linearized with *BamHI/PstI* double digests, subjected to Exonuclease III digestion (Figure 4a) in one minute intervals, blunt ended with Mung Bean nuclease, and religated. Screening of the deletion products was done using supercoil analysis in order to recover and assemble an overlapping set of deletion products from both ends. The result of this series is shown in Figure 4; verification of insert size was accomplished by *EcoRI/SalI* double digestion (Figure 4c) that resulted in insert excision. These inserts were then *EcoRI/SalI* directionally cloned in frame into the pGBT9 yeast two hybrid bait vector. Construction of the templates used for the Exonuclease III deletions permitted the direct, in frame, cloning of the 3' to 5' fragments. The 5' to 3' deletions had to be sequenced to confirm that they were in frame. Figure 5 diagramatically shows the deletions that were isolated from the Exonuclease III digestion series done in both directions.

Following retransformation into the yeast host, the deletions were tested for their ability to promiscuously activate the ß-galactosidase reporter gene. From the available panel, it was found that deletions removing the zinc finger domain in either orientation were sufficient in reducing ß-galactosidase activity to background levels (data not shown). Based on these results, two constructs were chosen for use as bait in the two hybrid screens. The end points of the N-terminal and C-terminal deletions are indicated in Figure 5.

The C-terminal deletion chosen for further analysis (N-terminal bait in Figure 5) contains most of the N-terminal region, and ends at amino acid position 266 within the

Figure 4 : Exonuclease III Digestion Series

Panel A : Exonuclease III digestions for 3' to 5' series generating C-terminal deletions. Marker lanes (lane 1) in all gels were run with 2 μg *HindIII/HincII* digested phage lambda. Lanes 2 to 11 : Five μg linearized template was loaded per lane, with aliquots removed at one minute intervals as detailed in the Materials and Methods.

Panel B : Supercoil screening of transformed and ligated deletion products. Lanes 1 and 10: lambda marker; Lane 2: undigested template; Lanes 3-10 and 11-17: supercoil plasmid DNA preparations.

Panel C : Completed set of overlapping C-terminal deletion products in pBluescript SK^+ excised using *EcoRI/ BamHI* double digestions. The digest products were directionally cloned into the PGBT9 two hybrid vector to give the products shown in Figure 5.

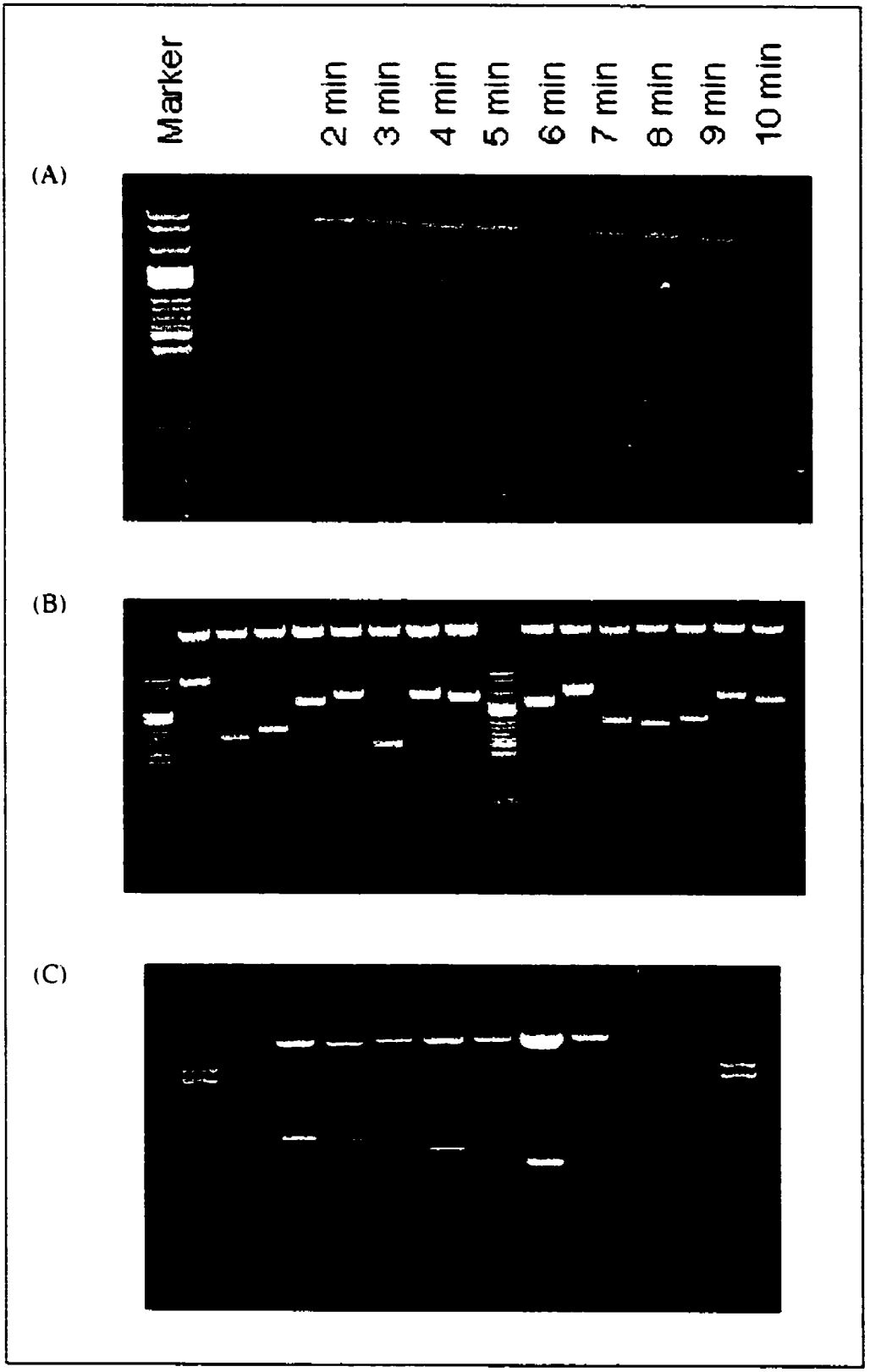
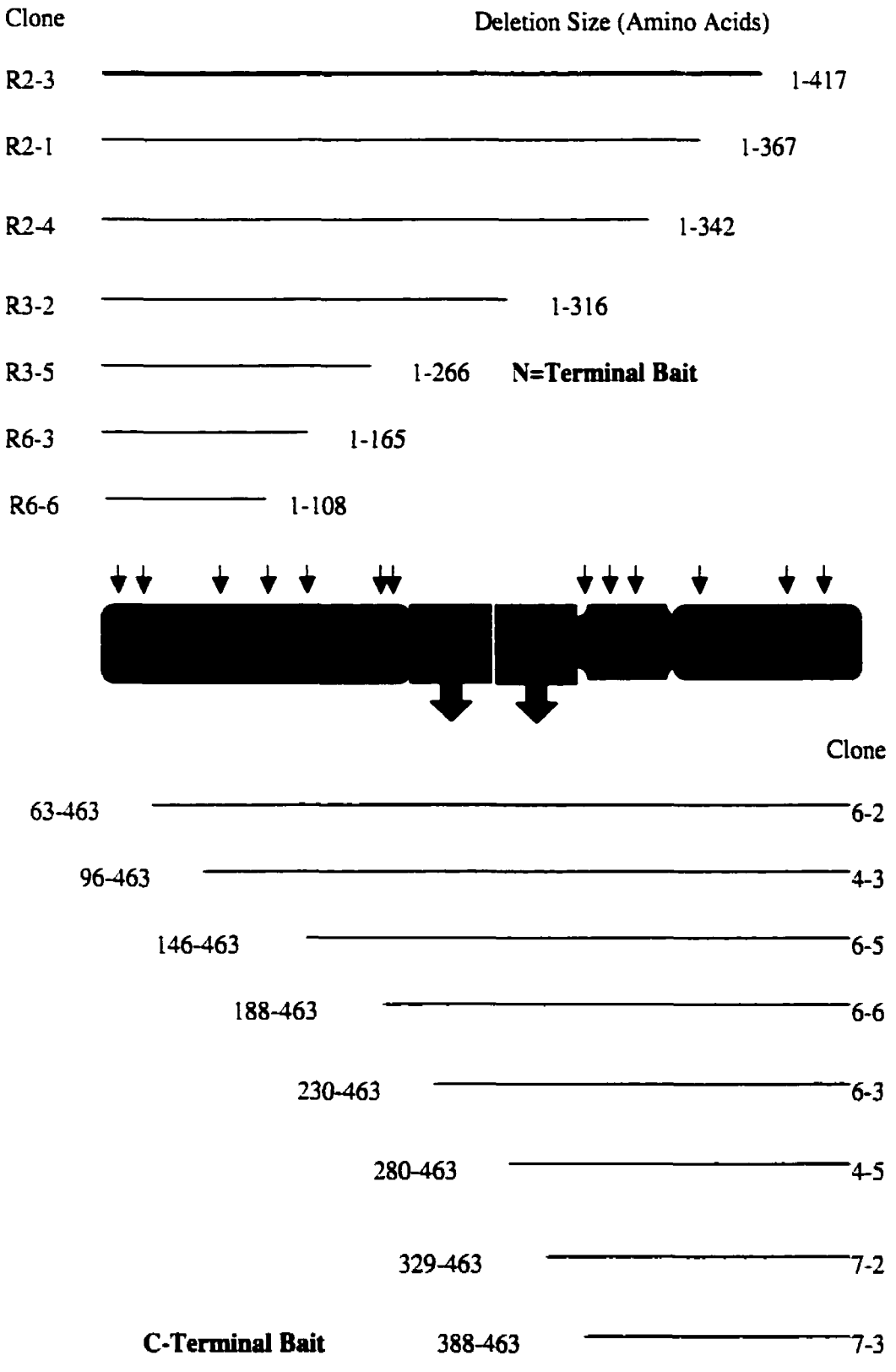

Marker
2 min
3 min
4 min
5 min
6 min
7 min
8 min
9 min
10 min
(A)
(B)
(C)

Figure 5 : Forward and Reverse Exonuclease III Deletions

Diagramatic representation of the individual clones isolated by supercoil analysis for both the 3' to 5' (top of figure) and 5' to 3' (bottom of figure) Exonuclease III deletion series. Clone designations are given adjacent to each deletion bar, with the size and termination point of each clone provided at the opposite end. Bar lengths are not to scale, but approximate the relative positions within the BmGATAß product where the deletions end.

The two deletions used as bait in the subsequent yeast two hybrid library screens are indicated by name beside clones R 3-5 (N-terminal bait) and 7-3 (C-terminal bait).


Clone
Deletion Size (Amino Acids)
R2-3
1-417
R2-1
1-367
R2-4
1-342
R3-2
1-316
R3-5
1-266
N=Terminal Bait
R6-3
1-165
R6-6
1-108
Clone
63-463
6-2
96-463
4-3
146-463
6-5
188-463
6-6
230-463
6-3
280-463
4-5
329-463
7-2
C-Terminal Bait
388-463
7-3

coding region from exon 5. The selected N-terminal deletion (C-terminal bait) removes all of the N-terminus and most of the zinc finger region, and begins coding at position 388 within the end of the second zinc finger. It is significant that this domain maintains both the nuclear localization signal and most of the surrounding phosphorylation sites that are suspected to modulate BmGATAß localization and activity.

Library Screens and Elimination of False Positives

Three library screens were conducted in total, two using the N-terminal bait and one using the C-terminal bait; as the expression pattern of BmGATAß overlaps both vitellogenesis and choriogenesis, the libraries used were also representative of both periods. (300 µg vitellogenic library cDNA at 4.2 µg/µL; 200 µg choriogenic library cDNA at 5.3 µg/µL).

The N-terminal screens produced a total of thirty potential interacting clones (NTPs: N-Terminal Preys), while the C-terminal screen yielded twelve (CTPs: C-terminal Preys). Positive responses included both the ability to survive histidine selection (in the presence of 20 mM 3-aminotriazole) and the ability to activate the ß-galactosidase reporter gene during colony lifts and colourimetric assays. The large number of clones recovered precluded direct sequencing approaches, and for this reason, a number of alternative strategies were employed to remove promiscuous activators and begin detailed characterizations. Promiscuous activators (and non-activators) were removed through control transformations in yeast of the prey alone or in combination with an empty bait vector or the bait used in screening (Figure 6a; 6b). Preys recovered from the library screens that did not appropriately activate the B-gal reporter (for example NTP 4-8, CTP 19-1; Figure 6a,b) were removed from the pool of potential positives. Only those preys that activated the reporter when in combination with the bait used in their isolation were

Figure 6 : Colour Reactions and Control Transformations

Colour reactions for selected clones isolated from N-terminal specific and C-terminal specific yeast two hybrid library screens. Examples of colour reaction series for confirmation of interaction and elimination of false positives. Each series runs left to right and reflects the following conditions : Left transformant, Prey plasmid alone ; Center transformant, Prey plasmid + empty bait vector (Gal4 DNA binding domain); Right transformant, Prey plasmid + BmGATAß bait.

Figure 6a: Control transformations for 3 N-terminal specific (top) and 3 C-terminal specific (bottom) preys. The weak positive response seen for NTP 1-2 is not readily evident in this figure. The positive control in the top filter is the Gal4DBD-Full Length BmGATAß2 fusion alone, and the negative control is a Gal4DBD-BmGATAß Zinc finger fusion alone.

Figure 6b: Control transformations for 5 N-terminal specific preys and an internal control series. The last series shown at the bottom of the figure demonstrates BmGATAß activation potential. The negative control at the left is a Gal4DBD-BmGATAß zinc finger fusion, the center control is the N-terminal bait used for screening, and the right control is the Gal4DBD-full length BmGATAß fusion.

Figure 6a

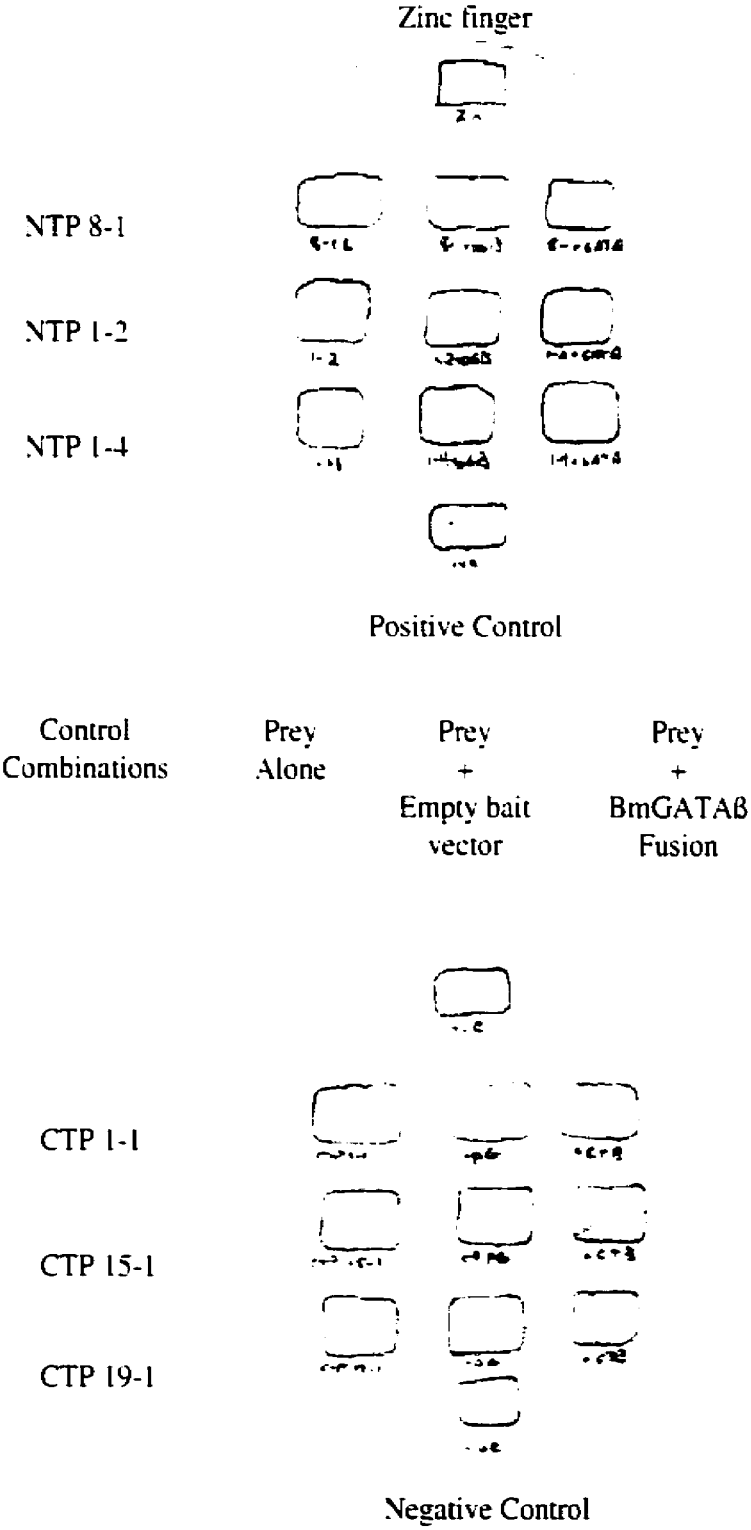
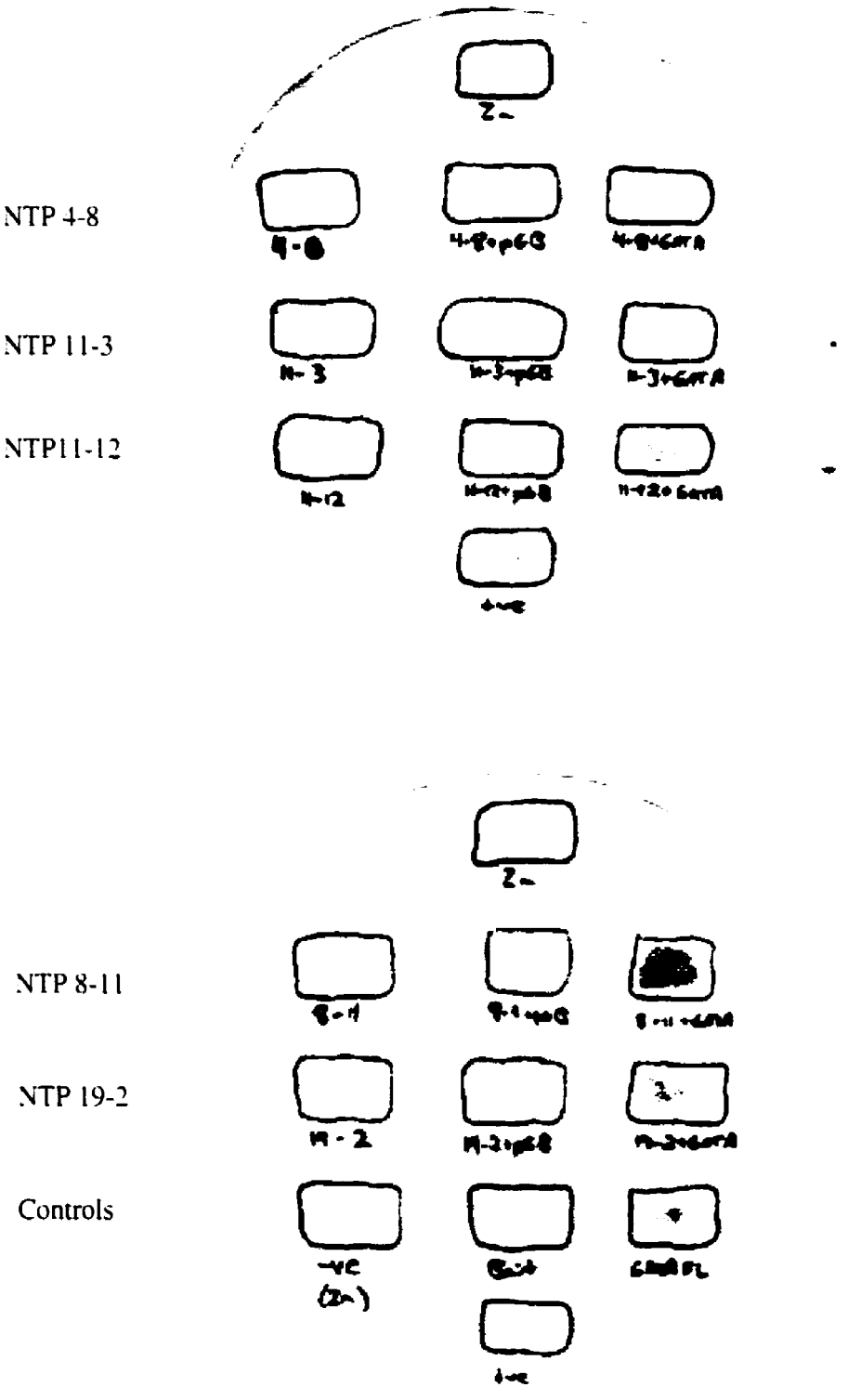

Zinc finger
NTP 8-1
NTP 1-2
NTP 1-4
Positive Control
Control Combinations
Prey Alone
Prey + Empty bait vector
Prey + BmGATAß Fusion
CTP 1-1
CTP 15-1
CTP 19-1
Negative Control

Figure 6b



maintained for subsequent characterizations (NTP 1-4, CTP 15-1: Figure 6a; NTP 11-3, NTP 8-11: Figure 6b).

DNA macroarrays and developmental dot blots (Figure 7a) were also used to reveal clones demonstrating the appropriate stage restricted expression characteristics. The dot blots were particularly useful in identifying preys with developmentally regulated transcripts (Figure 7a). Internal controls utilizing the predetermined expression patterns for BmHR3 and BmGATAß allowed for the establishment of detection limits and specificities. While some preys did not respond well to this protocol (for example NTP 1-2), others demonstrated striking patterns of expression (for example NTP 1-4) that led to their choice as preys for subsequent characterization.

Initial Characterization of Selected Preys

The methodologies noted above reduced the number of potential positives to seven which appeared to activate the His3 and ß-galactosidase reporters only in combination with the appropriate bait fusion. To begin the characterization of these clones, three approaches were taken: recovery and sequencing of the prey plasmids, stage specific developmental Northern analysis (examples shown in Figure 7b), and tissue specific Northern analyses (examples shown in Figure 8). The results of these initial analyses are summarized in Figure 9 (Table I). As can be seen from this table, five of the clones demonstrate developmental and/or tissue specific regulation, and all of these are expressed within the period of BmGATAß upregulation or activity.

Sequence Analyses

Currently available and verified sequence information for all of the characterized clones is provided in Appendix I (see RACE section below). Three of the clones (NTPs:

Figure 7 : Developmental Dot Blot and Northern Analyses for N-terminal Preys

Panel A: An example of developmental dot blots that were initially used to determine stage specific expression patterns. 5 µg staged follicle total RNA was spotted per dot. Internal controls of known developmental specificity were used including BmHR3 and BmGATAß.

Panel B : Developmental Northern Analyses for N-terminal preys NTP 1-2, NTP 8-11, and NTP 11-3. Each lane was loaded with five µg of total RNA obtained from staged pools of five follicles (staging is indicated at the top of the figure). Actin was used as a loading control for most of these determinations. The approximate sizes of the transcripts revealed are indicated to the right of the Northerns.

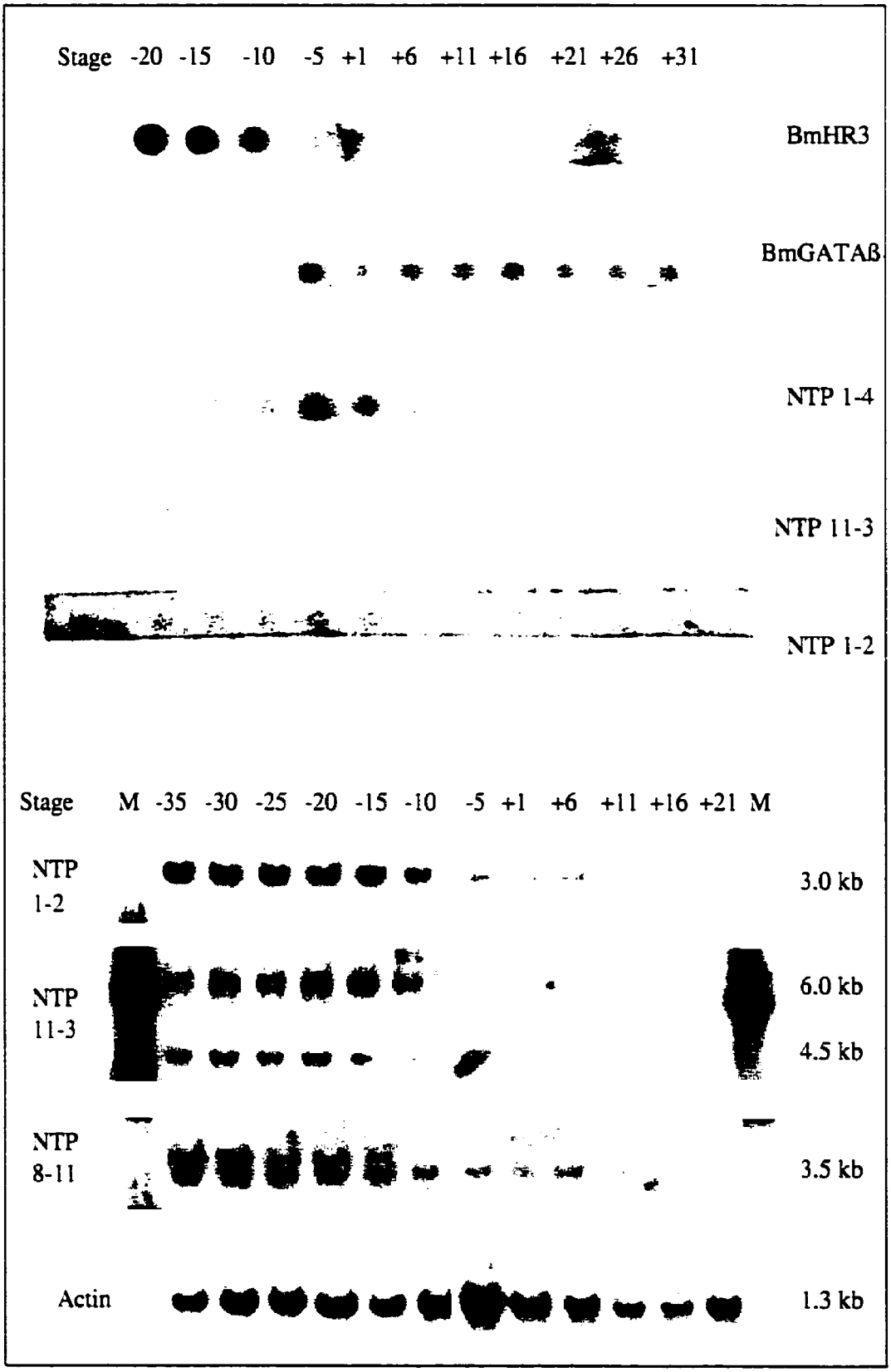

Stage -20 -15 -10 -5 +1 +6 +11 +16 +21 +26 +31
BmHR3
BmGATAß
NTP 1-4
NTP 11-3
NTP 1-2
Stage M -35 -30 -25 -20 -15 -10 -5 +1 +6 +11 +16 +21 M
NTP 1-2
3.0 kb
NTP 11-3
6.0 kb
4.5 kb
NTP 8-11
3.5 kb
Actin
1.3 kb

Figure 8 : Tissue Specific Northern Analyses for N-terminal Preys NTP 1-2, NTP 8-11 and NTP 11-3.

Total RNA was extracted from body wall, brain, fat body, malphigian tubule, wing, testes and ovary tissues of day 6 pupae. Five μg of total RNA was loaded onto each lane of a 1.2% agarose/formaldehyde gel. Exposure times for autoradiography were 48 hours for NTP 11-3, and 12 hours for NTP 1-2, NTP 8-11 and the actin control.

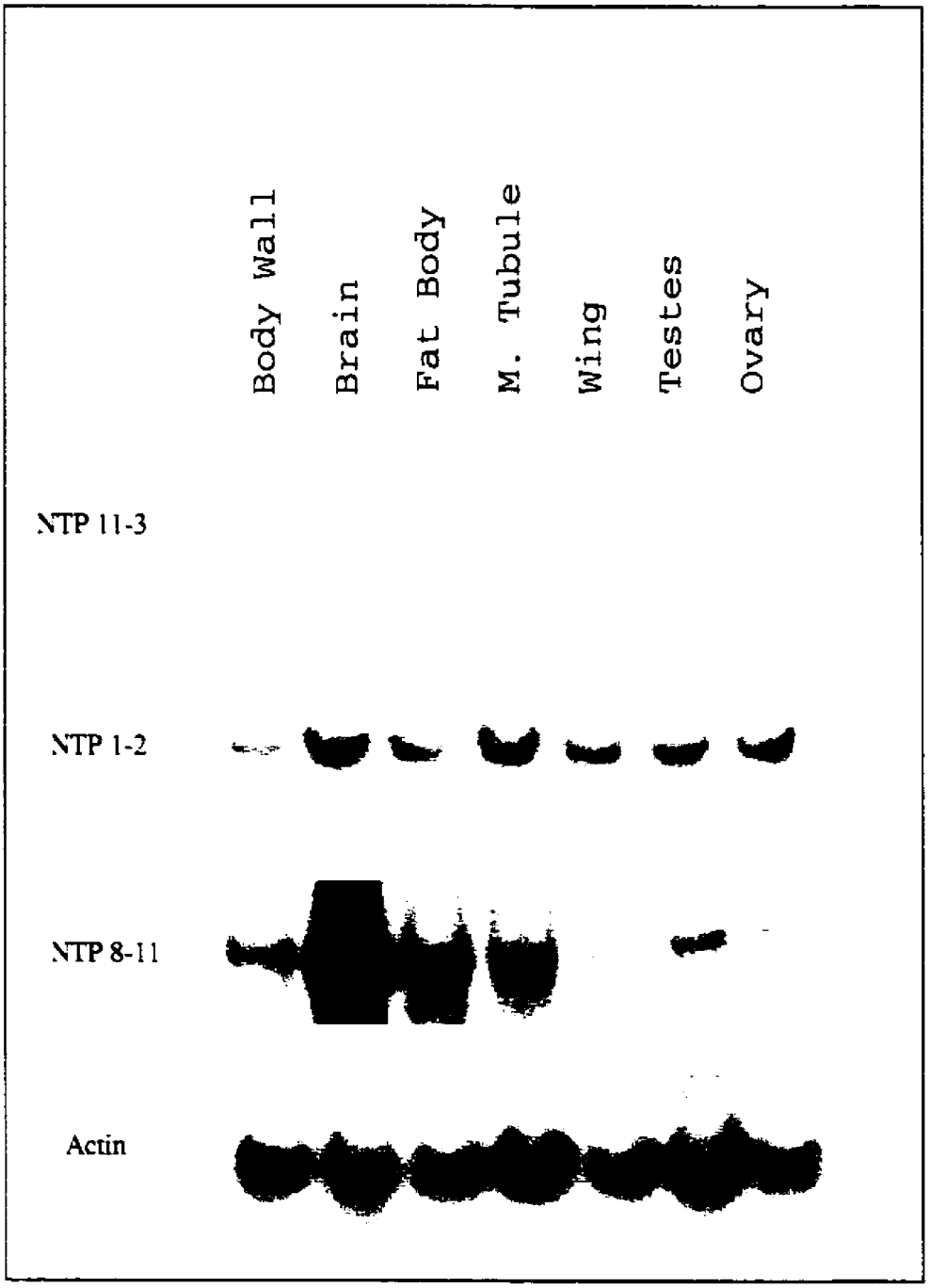

Body Wall
Brain
Fat Body
M. Tubule
Wing
Testes
Ovary
NTP 11-3
NTP 1-2
NTP 8-11
Actin

Figure 9 : Summary Table for Initial Two Hybrid Clone Characterizations

This table summarizes relevant information obtained for the original seven preys isolated using the N-terminal and C-terminal BmGATAß deletions. This data includes the original two hybrid clone size, current cloning status (vector), transcript size(s), and developmental and tissue specific Northern analyses.

Clones presenting closely related sequence homologues are indicated in bold (see Appendix II), while those clones for which identification is tentative are in brackets.

Table I : Summary of Two Hybrid Clones: Status and Expression Patterns

Clone	Insert Size	Cloning Status	Transcript Size	Homologue	Northern Analysis: Developmental	Northern Analysis: Tissue
NTP 1-2	1.3 kb Eco/Bam	In pGadd and pBS	3.0 kb	**Ribosomal Protein L34**	-45 to -10. Down - regulated at - 5, absent by +16	All Tissues
NTP 1-4	900 bp Eco/Bam	In pGadd and pBS	4.5 kb	None	-5 to + 5. Absent at all other stages	Ovary Specific
NTP 8-11	700 bp Eco/Bam	In pGadd and pBS	3.5 kb	**t-complex polypeptide 20**	-45 to -15. Down-regulated at - 10	All Tissues
NTP 11-3	650 bp Eco/Bam	In pGadd and pBS	6.0and 4.5 kb	(SWI/SNF)	-45 to -10. Down-regulated at - 5	Ovary Specific
CTP 1-1	1.1 kb Eco/Bam	In pGadd	N/A	**B. mori alpha - tubulin**	Not Available	Not Available
CTP 5-1	600 bp Eco/Bam	In pGadd	N/A	None	Not Available	Not Available
CTP 15-1	1.0kb Eco/Sal	In pGadd and pBS	3.0 kb	(Polycomb)	+ 1 to + 30. Absent during vitellogenesis	Ovary and Brain Specific

1-2, 8-11 ; CTP 1-1) have highly homologous sequence orthologues (Figure 9; Appendix II); the remaining three clones demonstrate little homology within currently available sequence databases, either at the nucleotide or protein level.. Possible sequence homologues (based on overall structure and small conserved motifs) for the unidentified clones are presented in the Discussion and results from BLAST searches are given in Appendix II.

Developmental and Tissue specific Northern Analyses

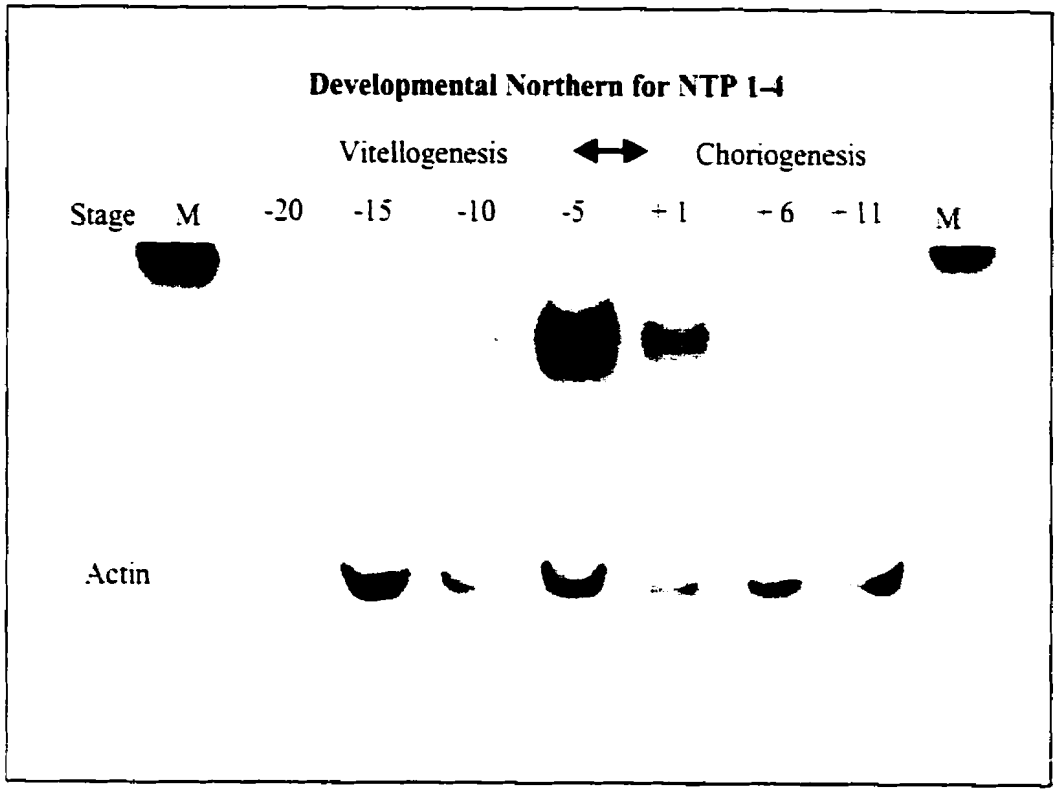
Following isolation and sequencing, it was necessary to determine whether the clones were expressed in a developmentally regulated and tissue specific fashion that would allow for potential interaction with BmGATAß. Developmental Northern analyses were accomplished using staged follicles (5 follicles per stage representing approximately 10-12 hours of developmental periods) spanning the range from approximately -30 to +30. This allows for characterization of factors from mid vitellogenesis to late choriogenesis. Two of the clones (CTP1-1 and CTP5-1) are not represented in the Northern hybridizations shown in Figure 7. Based on sequence analysis, CTP 1-1 encodes *Bombyx* alpha-tubulin, and would be expected to be expressed ubiquitously. No clear signal could be found for CTP 5-1 even after significant exposure. As can be seen from Table I, all of the remaining clones are expressed over the time period of BmGATAß upregulation or activity. Interestingly three of these, NTPs 1-2, 8-11, and 11-3 appear to be downregulated at or immediately before the upregulation of the BmGATAß transcript (Figure 7b). The possible significance of these products and the manner of their expression are detailed in the Discussion.

The remaining two clones (NTP 1-4 and CTP 15-1) demonstrate striking patterns of expression. NTP 1-4 is not expressed until stage -6 (Figure 10), is concurrent with the upregulation of BmGATAß, and as with BmGATAß, it is also downregulated upon entry into choriogenesis. Unlike its potential partner, however, NTP 1-4 does not persist past

Figure 10 : Developmental and Tissue Specific Northern Analyses for NTP 1-4

Panel A: Developmental Northern Blot. Each lane contains 5 μg of total RNA isolated from desheathed and staged follicles. A 450 bp EcoRI/BamHI fragment of NTP 1-4 was used as a template for probe synthesis. Both exposures in Panel A were for 8 hours.

Panel B: Tissue Specific Northern Blot. Total RNA was extracted from body wall, brain, fat body, gut, malphigian tubule, wing, testes, and ovary tissues of day 6 pupae. Exposures for autoradiography were 12 hours for both NTP 1-4 and actin.


Developmental Northern for NTP 1-4
Vitellogenesis
Choriogenesis
Stage
M
-20
-15
-10
-5
+1
+6
+11
M
Actin

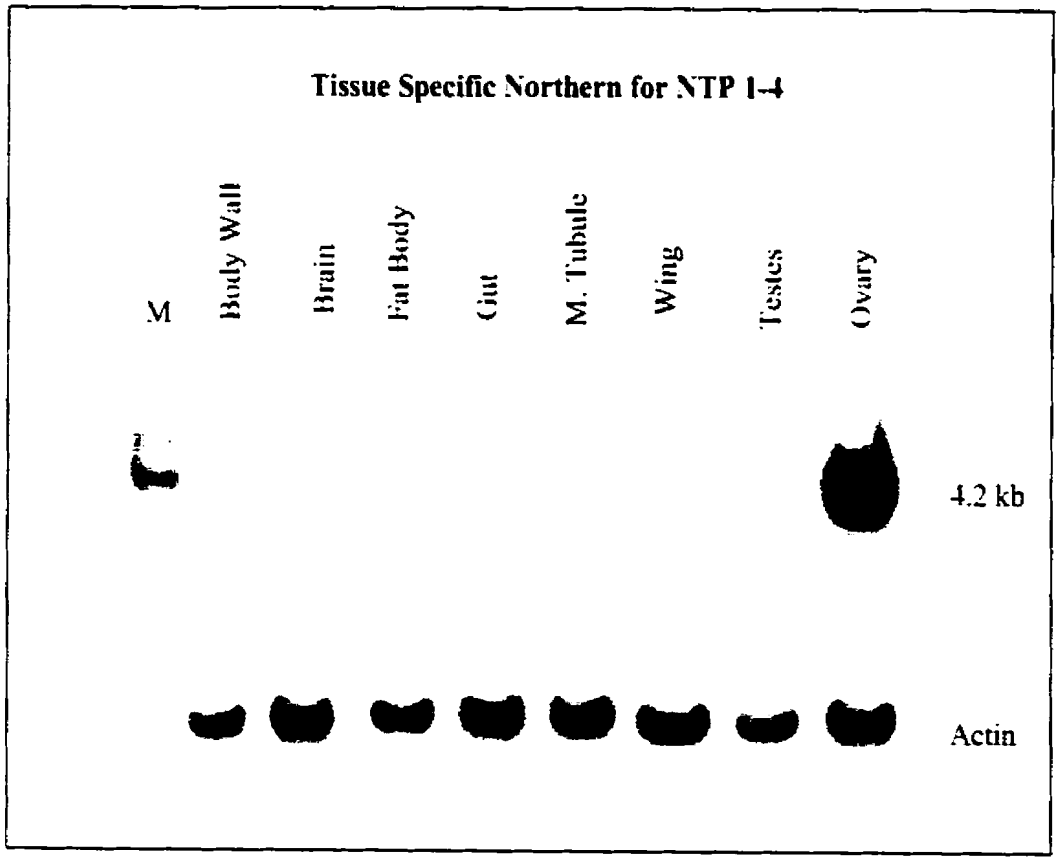
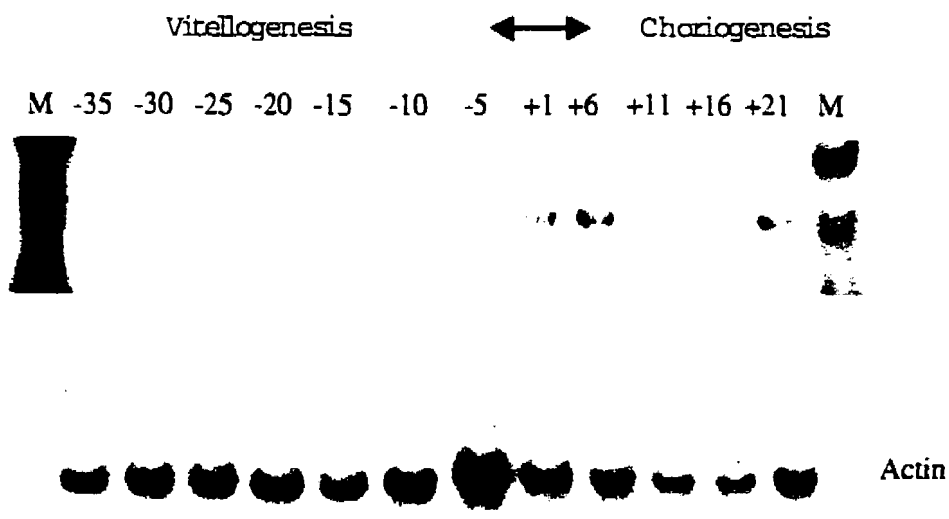

Tissue Specific Northern for NTP 1-4
M
Body Wall
Brain
Fat Body
Gut
M. Tubule
Wing
Testes
Ovary
4.2 kb
Actin

Figure 11 : Developmental and Tissue Specific Northern Analysis for CTP 15-1

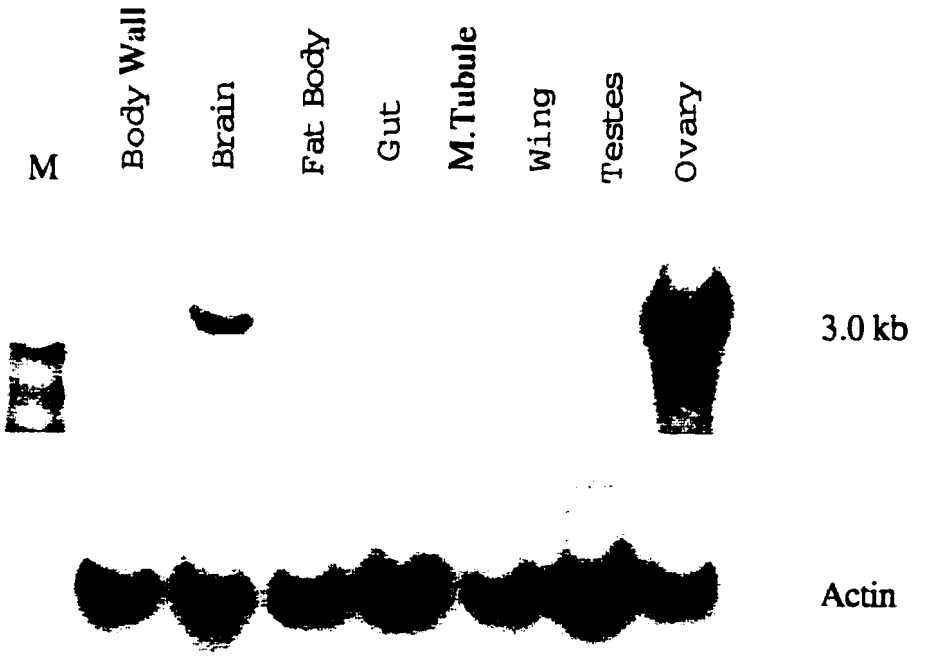
Panel A : Developmental Northern analysis using staged follicles. Total RNA was extracted from staged follicles in groups of five from a day 8 pupae. Five µg of total RNA was loaded per lane. Exposures were 72 hours for CTP 15-1 and 12 hours for actin control.

Panel B : Tissue specific Northern analysis using total RNA extracted from body wall, brain, fat body, gut, malphigian tubule, wing, testes, and ovary tissues. Exposures were done for 48 hours for CTP 15-1, and 12 hours for actin. The autorad was slightly overexposed in order to reveal the light signal arising from brain RNA.

Developmental Northern Analysis for CTP 15-1



Tissue Specific Northern Analysis for CTP 15-1



developmental stage +5. CTP 15-1 on the other hand, is completely choriogenic-specific and is not expressed until almost exactly +1 (Figure 11) and continues to be expressed at low levels throughout the periods examined, at least until stage +30.

BmGATAß demonstrates limited tissue-specific expression and appears to be expressed in several cell types, with isoform specification appearing in a more restricted manner. Since expression of the chorion genes is unique and specific to the follicular epithelium, one might expect that factors modulating BmGATAß activity would be specific to this tissue and cell type.

In order to characterize tissue specific expression profiles for the clones, total RNA from seven representative tissues (body wall, fat body, gut, malphigian tubule, wing, ovary, and testes) was separated by formaldehyde/agarose gel electrophoresis and examined for the presence of mRNA for the cloned sequences by Northern hybridization. CTPs 1-1 and 5-1 were not included in the tissue specific Northerns for the reasons noted above. NTPs 1-2 and 8-11 (Figure 8) are ubiquitously expressed at comparable levels in all tissues examined. The remaining three unknowns all demonstrate a high degree of tissue specificity, and are restricted to the ovary (Figure 8, NTP 11-3; Figure 10 NTP 1-4) with the exception of CTP 15-1 which is expressed at low levels in brain tissue (Figure 11) during pupal development. Thus all of the clones examined, for which signals could be generated, are expressed in the appropriate time and tissues that one would expect for potential BmGATAß interactors.

The identification of the cell type in which these sequences are expressed, and confirmation and refinement of the developmental Northern results, were accomplished by *in situ* hybridization (below). Based on the restricted tissue and developmental expression periods for NTP 1-4 and CTP 15-1, these two clones were chosen for subsequent analyses. This does not, however, limit the interest or integrity of the work accomplished for the other clones, particularly for NTP 11-3 (expanded on in the

Discussion), and at this point there is no reason to immediately preclude these products from playing a role in the generation or activity of BmGATAß.

In situ hybridizations for NTP 1-4 and CTP 15-1.

As mentioned in the Introduction, the *Bombyx* follicle contains three cell types, the oocyte proper and nurse cells which arise from the primordial germ cell, and the follicular epithelium which is mesodermally derived. In order to examine the distribution of the mRNA sequences in the different cell types of the follicles, whole mount *in situ* hybridizations were done against panels encompassing the entire range of follicle development. Digoxigenin labeled *in vitro* transcribed probes were prepared and used to delineate the expression patterns for NTP 1-4 and CTP 15-1. The results of these analyses are shown in Figure 12 (NTP 1-4) and Figure 13 (CTP15-1).

From the results of the developmental Northern analysis (Figure 10), NTP 1-4 would be expected to be expressed abundantly in follicles ranging from -5 to +5. As can be seen in Figure 12a, the transcript is expressed exclusively during the vitellogenic/choriogenic boundary. Figure 12b also shows that this expression is limited to the period defined above, and refines the expression period to -6 to +4/5; comparisons made with surrounding unstained follicles suggest that the -7 and the +5/6 follicles do not express this transcript. Cell type specificity is limited to the follicular epithelium as shown by Figure 12c and d. The transcript is expressed in the follicular epithelium as shown by the characteristic "honeycomb" staining pattern observed previously in this system for epithelium-specific targets. This expression is limited to the epithelium as shown in Figure 12d, as the overriding cap of nurse cells clearly does not stain with this probe. Figure 12e and 12f show hybridizations done in parallel with the sense probe; while there does appear to be some background staining, this is minimal, and non-specific.

Figure 12 : NTP 1-4 transcript localization by *in situ* hybridization.

Two complete ovarioles were taken from a day 7 pupae for each sense and antisense hybridizations, and the follicles were desheathed prior to fixing. Dig-labeled riboprobes were produced from the partial NTP 1-4 cDNA cloned into pBluescript using the T3 and T7 promoters. Panels A through D show hybridizations done with the antisense probe, and panels E and F show hybridizations with the sense control. All photographs were taken with a Zeiss III RS Light Microscope.

Panels A and B clearly show the tight developmental regulation of this transcript over the vitellogenic/choriogenic boundary (from stage -6 to stage +3). The higher magnification image in Panel C demonstrates the follicle-specific honeycomb staining pattern, while panel D shows the exclusion of the NTP 1-4 mRNA from the nurse cells. Panels E and F show the limited extent of background staining seen for the sense probe.

In situ hybridizations for NTP 1-4

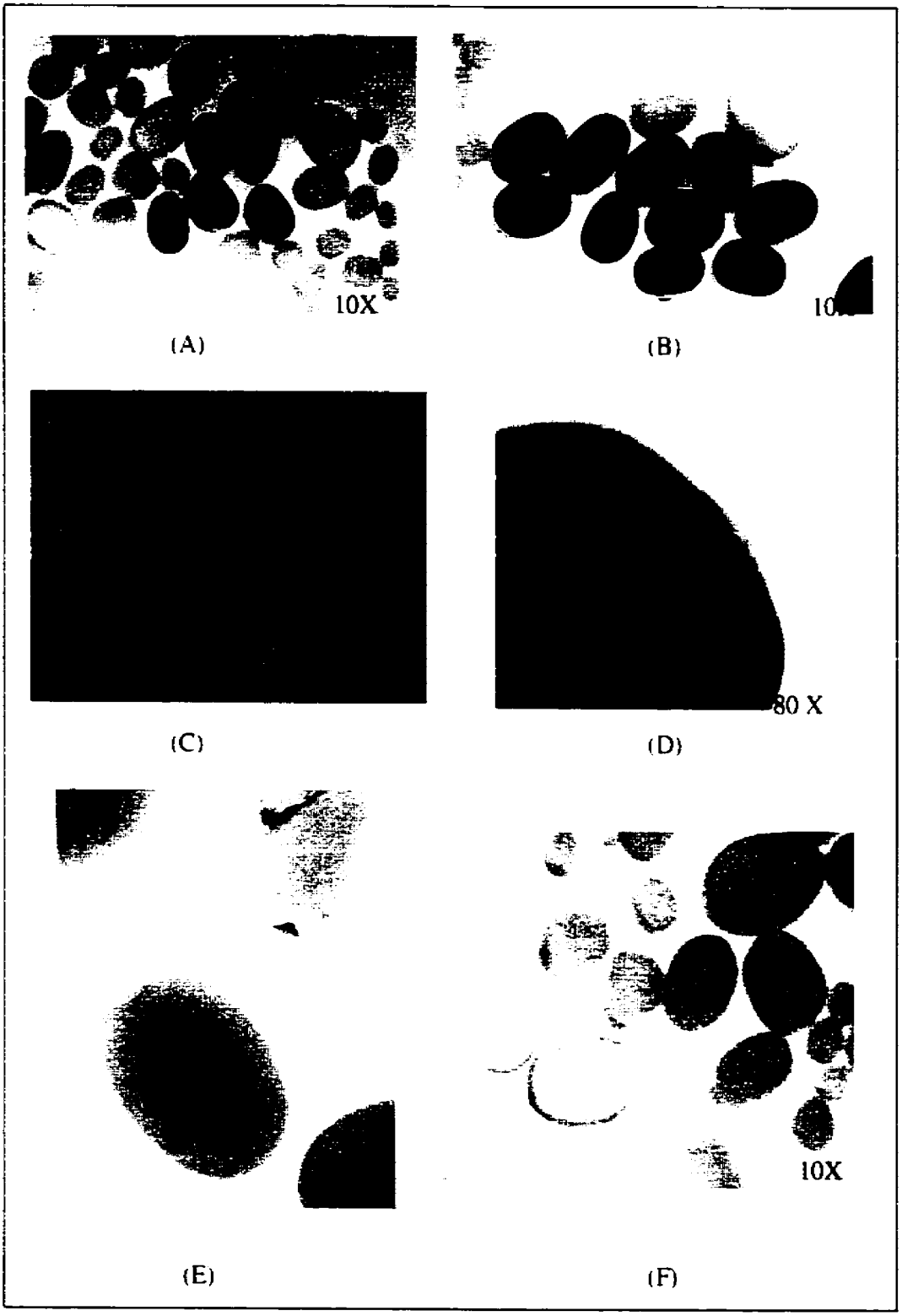
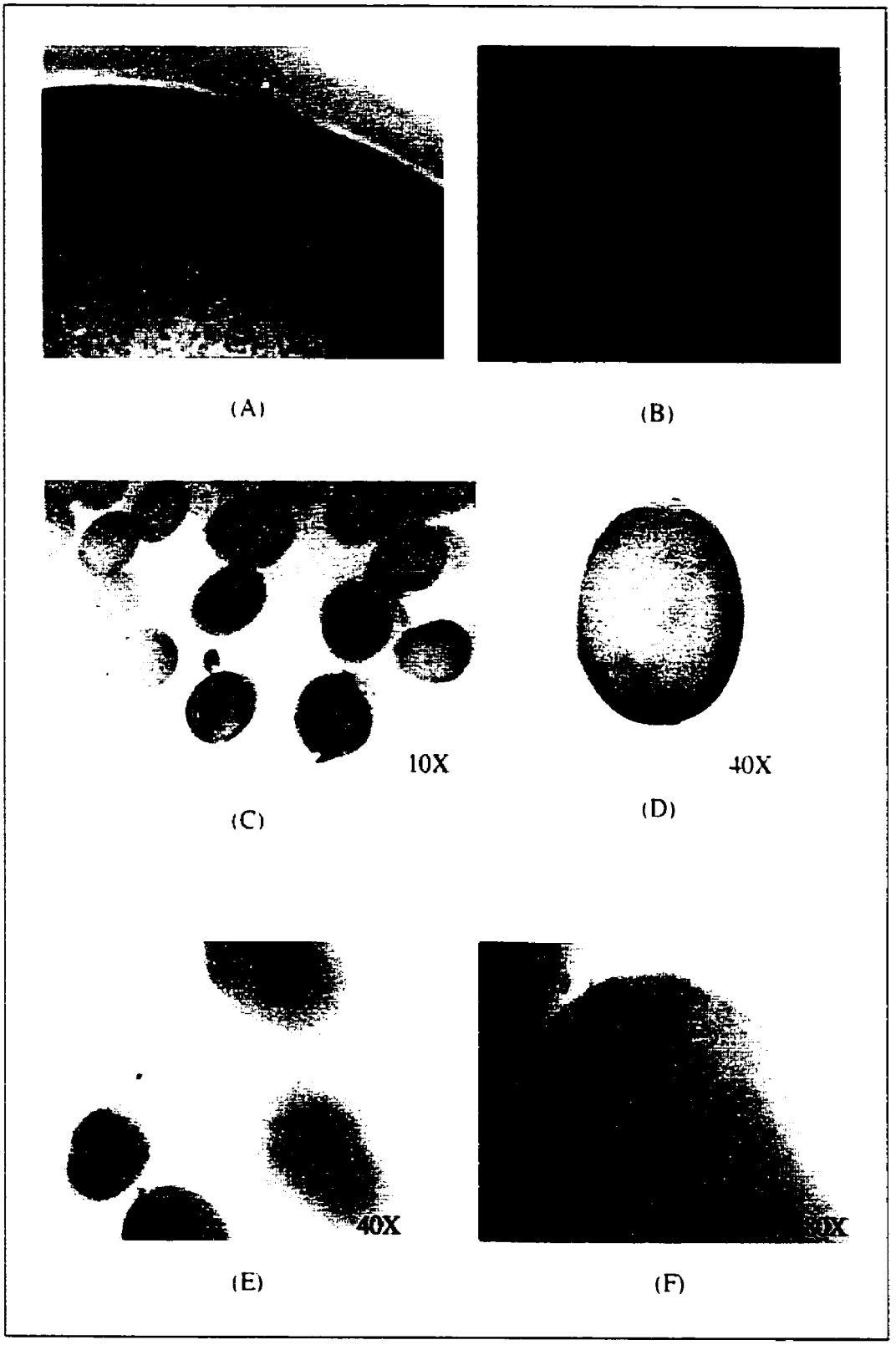


Figure 13 : Localization of CTP 15-1 transcript by *in situ* hybridization

In situ hybridizations were run against total follicle panels prepared from day 8 pupae. Colour reactions were allowed to proceed for 20 minutes due to the low level of expression observed from the Northern analyses.

Panels A and B show high magnification photographs localizing the CTP 15-1 mRNA to the follicular epithelium (honeycomb staining pattern). Panel C shows a lower magnification photo confirming the choriogenic specific expression pattern suggested by the tissue specific Northerns. Panel D demonstrates epithelial cell restriction, as evidenced by the much lighter staining over the nurse cell cap. Panels E and F show medium magnification pictures of the sense controls for choriogenic follicles.

In situ Hybridizations for CTP 15-1



CTP 15-1 mRNA proved more difficult with respect to providing a clean signal relative to NTP 1-4, most likely due to its very low abundance (seen in the Northern analysis, Figure 11). Repeated trials were successful in demonstrating the appropriate expression characteristics, however, and these are given in Figure 13. Figure 13a and 13b demonstrate the characteristic honeycomb staining pattern noted above for epithelium specific transcripts. Confirmation of the developmental Northern results (choriogenic specificity) is shown in Figure 13c, as all choriogenic follicles appear to stain, while those in vitellogenesis do not. Whether this transcript is confined to the follicular epithelium is not completely clear; the follicle shown in Figure 13d is characteristic of most observed early choriogenic follicles, and, while there is clearly less staining in the nurse cell cap than elsewhere in the epithelium, it is not possible to definitively preclude the presence of CTP 15-1 transcripts from the nurse cells. Given the low abundance of this message, the small size of the nurse cells at this stage, and the background problems alluded to above, it may still be reasonable to assume that the expression of this gene may indeed specific to the follicular epithelium. Additional experiments optimizing hybridization conditions or utilizing alternative probes or probes of higher specific activities may resolve this question.

λgt11 phage library screens.

In order to obtain full length sequences for NTP 1-4, NTP 11-3 and CTP 15-1, attempts were made to screen vitellogenic specific and choriogenic specific λgt11 cDNA libraries. Probes generated from the original NTP 1-4 clone were used to screen the vitellogenic specific library (Invitrogen), but sequences within hybridized to all plaques. The positions and signal intensities that resulted suggest that this is caused by a hybridization artifact and does not result from probe interactions with the filter or phage proteins. As CTP 15-1 is not expressed until the onset of choriogenesis, it would not be expected to be present in this library and was not screened for. A probe made from the

entire *EcoRI/BamHI* NTP 11-3 two hybrid fragment was used to screen this library, and this resulted in the recovery of several positive plaques. Three of these were recovered, plaque purified, and taken on to secondary confirmatory screening. The resulting positive plaques were eluted from their plugs and stored at 4 ^{o} C for future analysis.

Screening of 5' RACE Libraries for NTP 1-4 and CTP 15-1

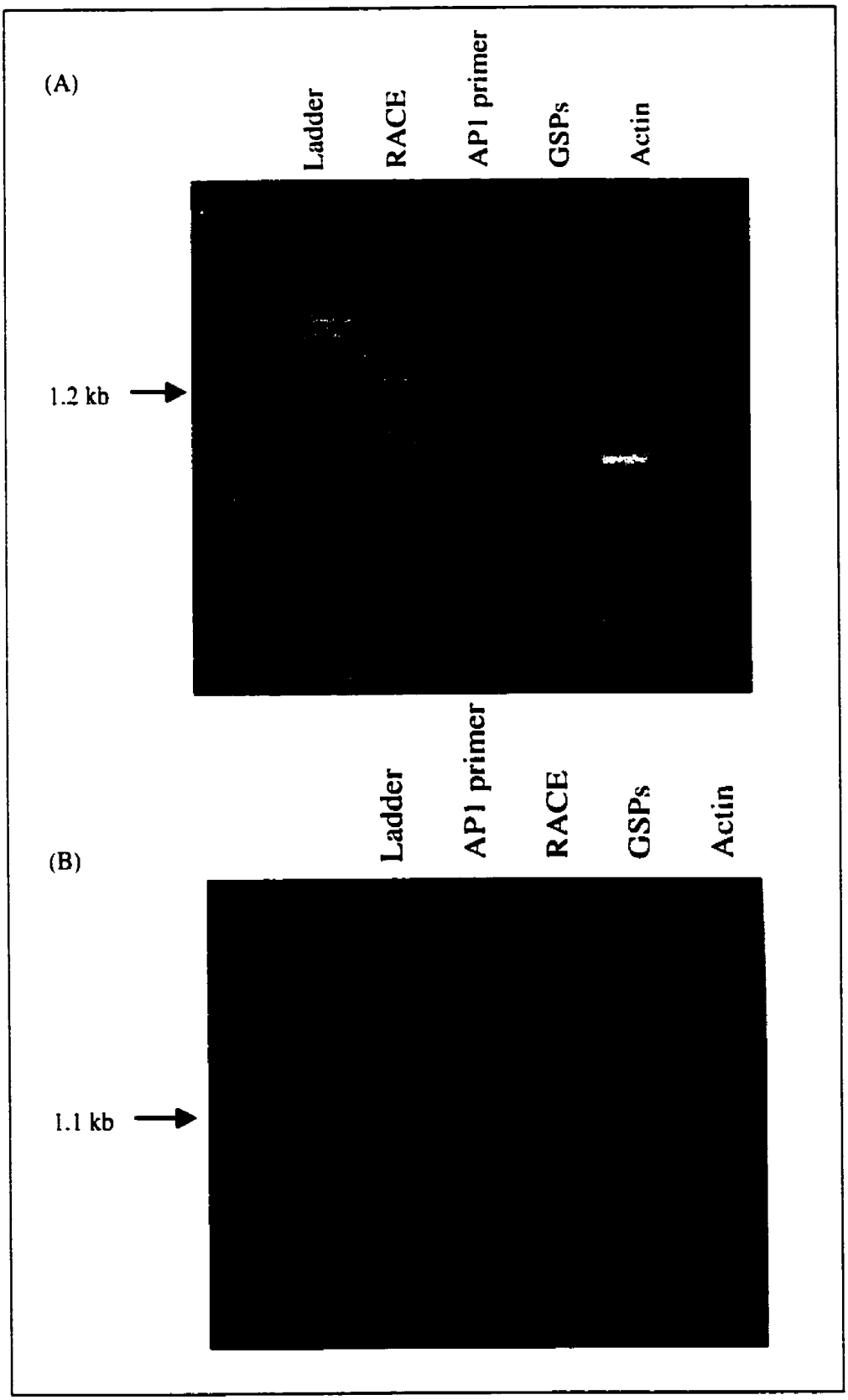
The original clones isolated through the two hybrid screens all appear to represent small 3' fragments of their full length transcripts (Figure 9; Appendix I). The NTP 1-4 fusion contains only 450 bp of unique coding sequence and open reading frame (only a portion of the 3' UTR is included in Appendix I due to the inability to read through the poly A tail). Two hybrid prey CTP 15-1 contains 600 bp of ORF and 480 bp of 3' UTR. In order to extend the available sequences, 5' RACE cDNA libraries were screened for both NTP 1-4 and CTP 15-1 using previously generated vitellogenic and choriogenic specific cDNA libraries (Lunke, 1999; Kendirgi, 2000). Internal nested primers were chosen from the 5' ends of the available clones (Appendix I; highlighted), with T_m temperatures high enough to permit touchdown PCR (Clonetech, 5' RACE Manual). Following optimization, these experiments provided the products shown in Figure 14, which were blunt end cloned into pBluescript prior to sequencing.

Figure 14a shows the expected actin positives, single primer negatives, and products of the RACE amplification for NTP 1-4 with primers 1109 and 1110: a single band migrating at approximately 1300 bp was repeatedly amplified. Cloning and sequencing of this fragment proved to be problematic, but it appears to provide a short extension to the open reading frame at the 5' end (confirmed sequences of the generated PCR product are highlighted in Appendix I; NTP 1-4). Extended sequencing of this fragment was not accomplished prior to the completion of this thesis, and all sequences highlighted in the Appendix are speculative.

Figure 14 : RACE experiments for NTP 1-4 and CTP 15-1

Panel A: 5' RACE experiment for NTP 1-4 using gene specific primers 1109 and 1110 (sequences given in Materials and Methods). Lane 1: lambda *HindIII/HincII* digest marker. Lane 2: 5' RACE experiment; note position of RACE product indicated by the arrow. Lane 3: RACE amplification using the AP1 primer alone. Lane 4: RACE amplification using the gene specific primer 1110 alone. Lane 5: RACE PCR run using actin specific primers. A similar pattern of bands was observed for the actin controls for most of the 5' RACE experiments.

Panel B: 5' RACE for CTP 15-1 using gene specific primers 1583 and 1584. Lane 1: lambda marker DNA. Lane 2: RACE control using the AP1 primer alone. Lane 3: RACE amplification using gene specific primer 1584 and primer AP1. A single band was consistently produced under these conditions, migrating at approximately 1.1 kb. Lane 4: RACE control using the gene specific primer alone. Lane 5 : Actin control for 15-1 5' RACE experiments.


(A)
Ladder
RACE
AP1 primer
GSPs
Actin
1.2 kb
(B)
Ladder
AP1 primer
RACE
GSPs
Actin
1.1 kb

Identical experiments were performed for CTP 15-1 using gene specific primers 1583 and 1584 (Materials and Methods; annealing positions given in Appendix I). As can be seen in Figure 14b, this 5' RACE reaction generated a fragment of 1200 bp; accounting for the overlap with previously identified sequences (and the inability to verify the 5' sequences of the PCR product), this provided 360 bp of additional 5' sequence information (Appendix I). Added to the original 1080 bp long two hybrid clone, this results in a total of 1440 bp of confirmed sequence.

While the RACE experiments were successful in extending the sequence information for these clones, the remaining information (suggested by the size of the mRNA transcripts) still represents the bulk of their sequence (Discussion). Given the time required for generating several new primer sets and optimizing reaction conditions for each round, alternative approaches were pursued to generate longer extension products.

5' Extension Strategy and Amplifications

As the RACE libraries mentioned above appeared to contain primarily short cDNAs, an RT PCR methodology was used. The 5' RACE primers for both NTP 1-4 and CTP 15-1 (1110 and 1583 respectively; Materials and Methods) were used as primers for first strand cDNA synthesis against purified follicle poly $(A)^+$ RNA (kindly provided by Ken Ito). The resulting cDNA:RNA hybrids were blunt ended, cloned into pBluescript, and reamplified using the internal nested RACE primers (1109 and 1584 respectively). The results of these amplifications are shown in Figure 15; these bands were removed from the gel, purified, and ligated into an *EcoRV* blunt ended pBluescript SK+ vector. The recovery and sequencing of these products is currently still not completed.

The original NTP 1-4 two hybrid clone contained an insert of 900 base pairs (Table I), only 450 bp of which appears to be in the open reading frame (Appendix I). Given an estimated transcript size of 4.5 kb, this suggested that there was approximately 3.6 kb of

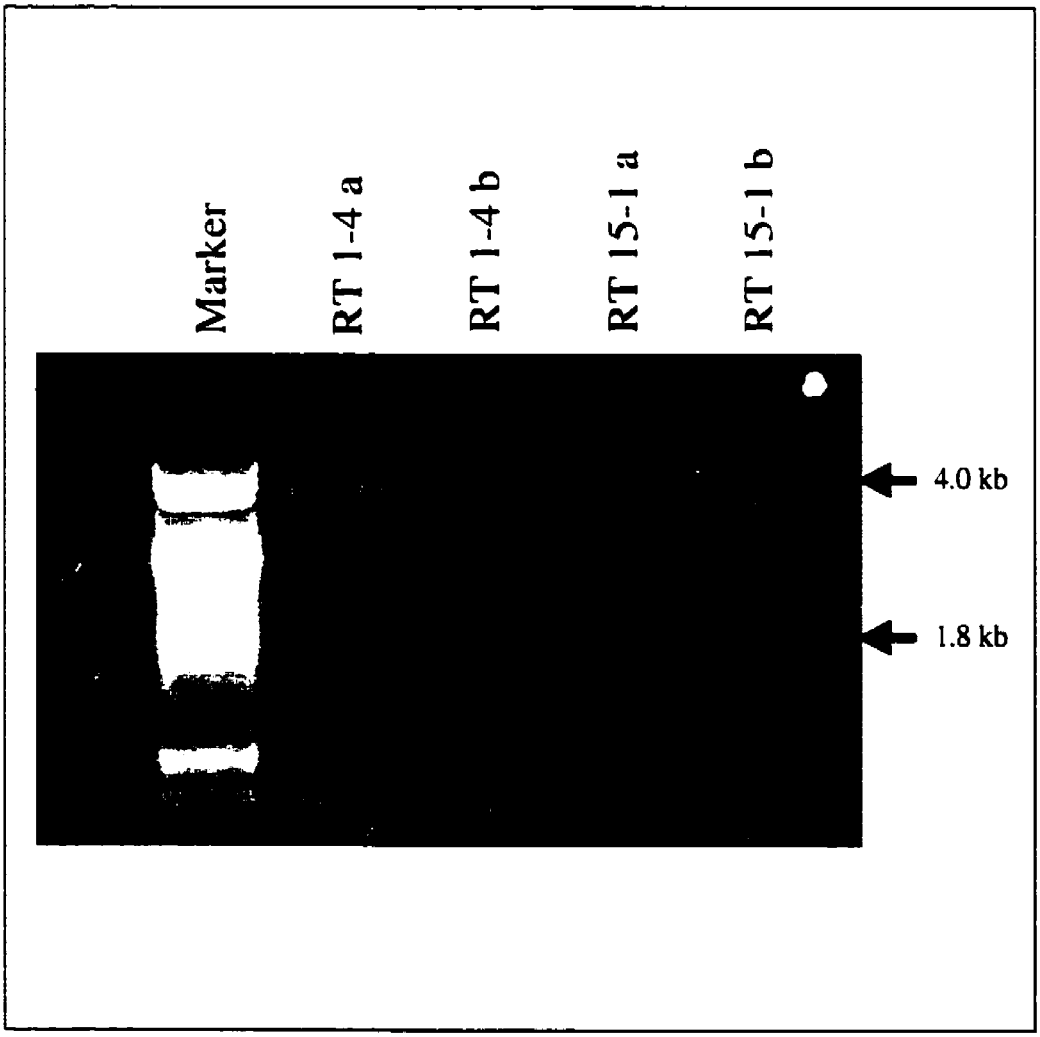
Figure 15 : Production of 5' extensions through RT PCR.

Products shown are the result of PCR amplification of first strand cDNA:RNA hybrid molecules produced by reverse transcription. Arrows indicate the position and size of amplified products.

Lane 1: Lambda DNA ladder

Lanes 2, 3 : PCR amplifications for NTP 1-4, producing a product of approximately 3.8 kb.

Lanes 3,4: PCR amplifications for CTP 15-1 using the gene specific primers from the 5'RACE experiments. The band indicted by the arrow is extremely faint, but was reproducible at 1.8 kb under the conditions used.


Marker
RT 1-4 a
RT 1-4 b
RT 15-1 a
RT 15-1 b
4.0 kb
1.8 kb

remaining sequence upstream of the two hybrid clone. As can be seen in Figure 15, panel A, the extension product amplified by the gene specific primer is between 3.4 and 4.0 kb. Should this product turn out to be a valid 5' extension, it should provide much of the remaining upstream sequences.

CTP 15-1 contained a 1.0 kb insert, with an mRNA estimated to be 3.0 kb in size from Northern hybridization. The use of the 5' extension methodology produced a PCR product 1.5 to 1.7 kb in length, also close to what would remain upstream of the original two hybrid fragment (Figure 15). However, this band is very weak, and attempts to reamplify it were unsuccessful.

The cloning of both of these fragments has proven to be a challenge, possibly due to the addition of pBluescript polylinker sequences into the 5' end of the PCR products during the amplification step. The use of pBluescript primers with annealing sites further upstream of the T3 promoter, or attempts at cloning these products into alternate vectors to overcome these problems is expanded on in the Discussion.

GST "pull downs" and experimentation

Preys identified using yeast two hybrid library screens must eventually be evaluated in an external system, providing verification of the observed interactions in yeast cells. One possible method for obtaining this is through the use of "pull down" experiments. Generally, one protein partner of a suspected pair is expressed as a GST fusion protein and bound to an insoluble support, while the other partner is extracted from native tissue or derived *in vitro*. These extracts or *in vitro* produced proteins are passed over the column containing the fusion, and non-specific proteins are removed in subsequent washing steps. The presence of the second partner is then revealed by Western blotting (if a specific antibody is available) or autoradiography (for radioactively labeled *in vitro* translated proteins).

Figure 16 : Purification Steps for Large Scale Production of GST-BmGATAß Fusion Proteins.

Twenty µL aliquots taken from the purification procedure at various timepoints (indicated above the lanes) were subjected to SDS-PAGE analyses. Expected sizes for the fusion proteins based on clone and transcript length are designated by arrows.

The top gel in this figure demonstrates the production of soluble GST and soluble GST-full length BmGATAß. The fusion protein band was faint but quite obvious prior to drying and scanning of the gel.

The bottom gel demonstrates the purification steps for the C-terminal and N-terminal fusions used in the yeast two hybrid screens. Note the presence of bands of predicted size (arrows) and degradation products.

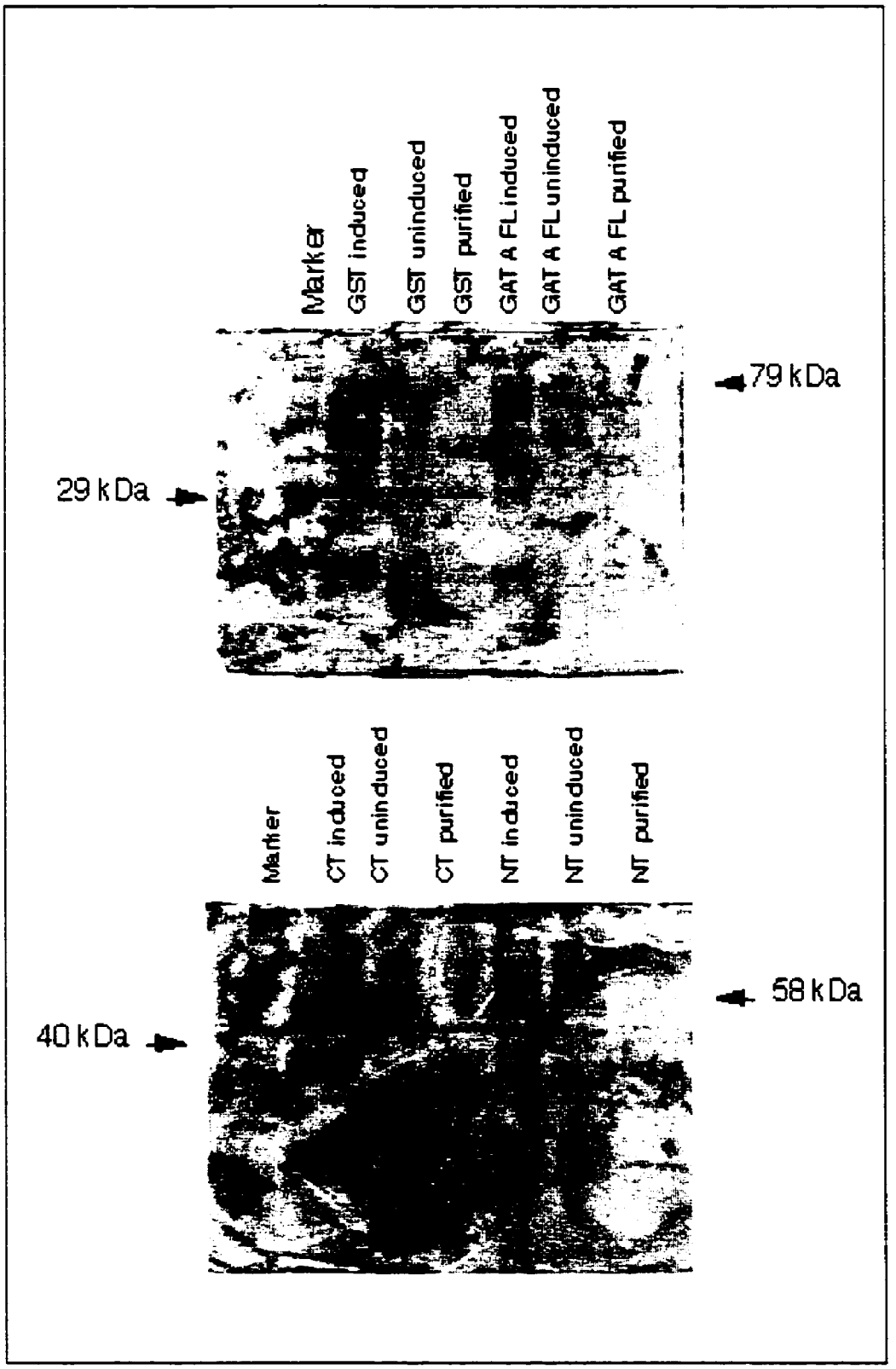
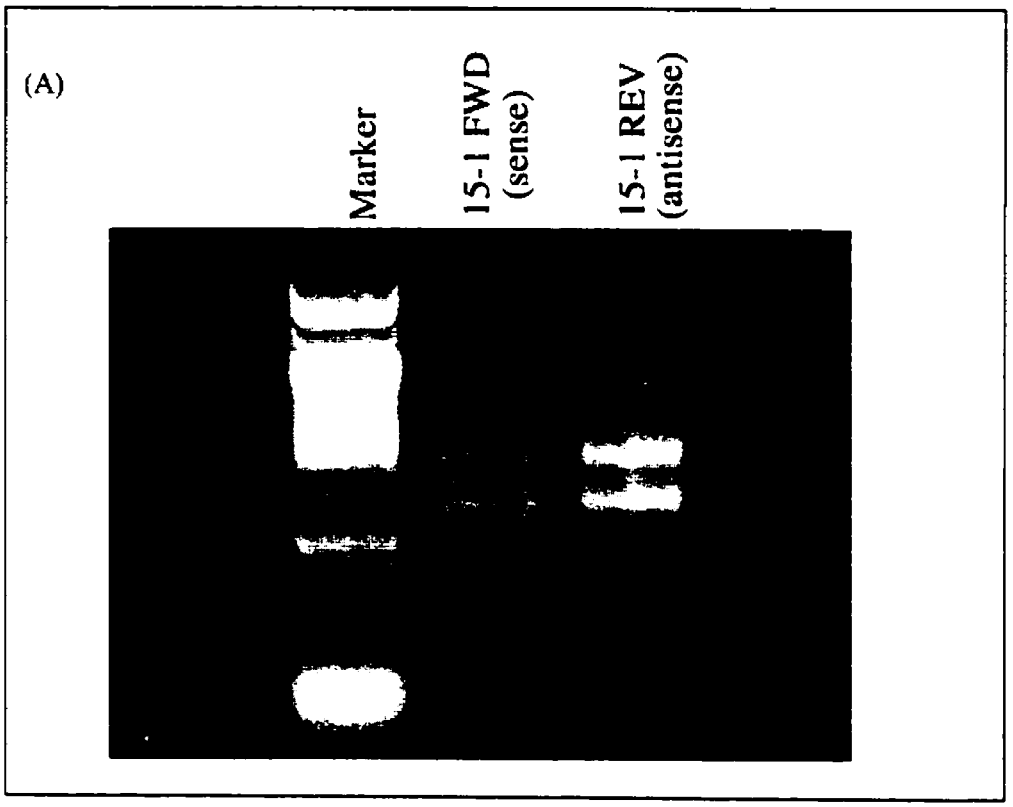
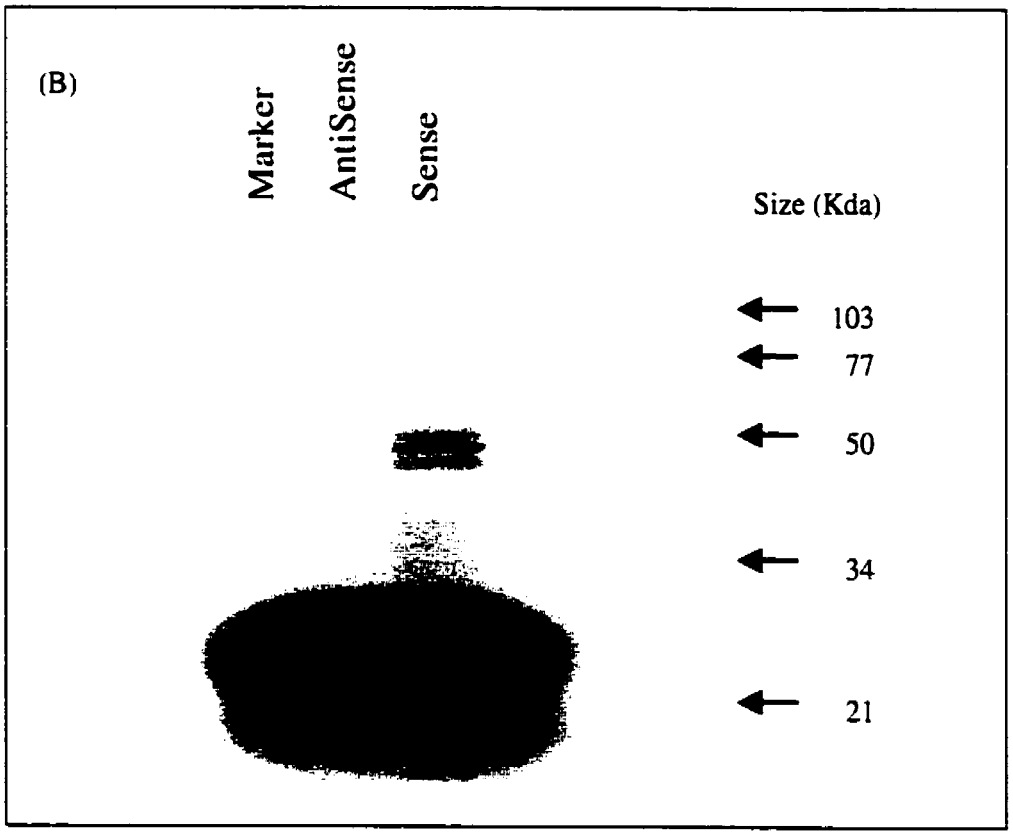

Marker
GST induced
GST uninduced
GST purified
GAT A FL induced
GAT A FL uninduced
GAT A FL purified
29 kDa
79 kDa
Marker
CT induced
CT uninduced
CT purified
NT induced
NT uninduced
NT purified
40 kDa
58 kDa

Figure 17 : Production of *in vitro* Transcribed and Translated CTP 15-1 mRNA

Panel A: Lane 1: DNA ladder. Lane 2: *In vitro* transcription of clone 15-1 sense RNA in pBluescript SK+ using the T3 promoter. Lane 3: In vitro transcription of the antisense RNA using the T7 promoter.

Panel B: Lane 1: Broad range protein ladder. Lane 2: One µg of the antisense product shown in A was added to a rabbit reticulocyte lysate in the presence of ^{35}S-methionine. Lane 3: One µg of the sense product from the *in vitro* transcriptions was used to generate translation products. Positions of protein markers are shown to the right of the figure. The expected size of the product is approximately 22 kDa; the presence of higher molecular weight bands (at 50 kDa) is most likely the result of the formation of higher order complexes as these bands are not present in the antisense control.


(A)
Marker
15-1 FWD (sense)
15-1 REV (antisense)


(B)
Marker
AntiSense
Sense
Size (Kda)
103
77
50
34
21

Full length BmGATAß and the two N-and C-terminal fragments used for library screening were cloned in frame into a pGEX 5X-1 vector in order to generate GST fusions that could be column purified. These plasmids were transformed into *E.coli* DH10b and expressed under the conditions given in Materials and Methods. The purified products and their calculated masses (determined assuming a mean residue weight of 106 Daltons) are shown in Figure 16. While only the GST product alone was produced in significant quantities, bands corresponding to the expected fusion sizes were also detectable in this gel.

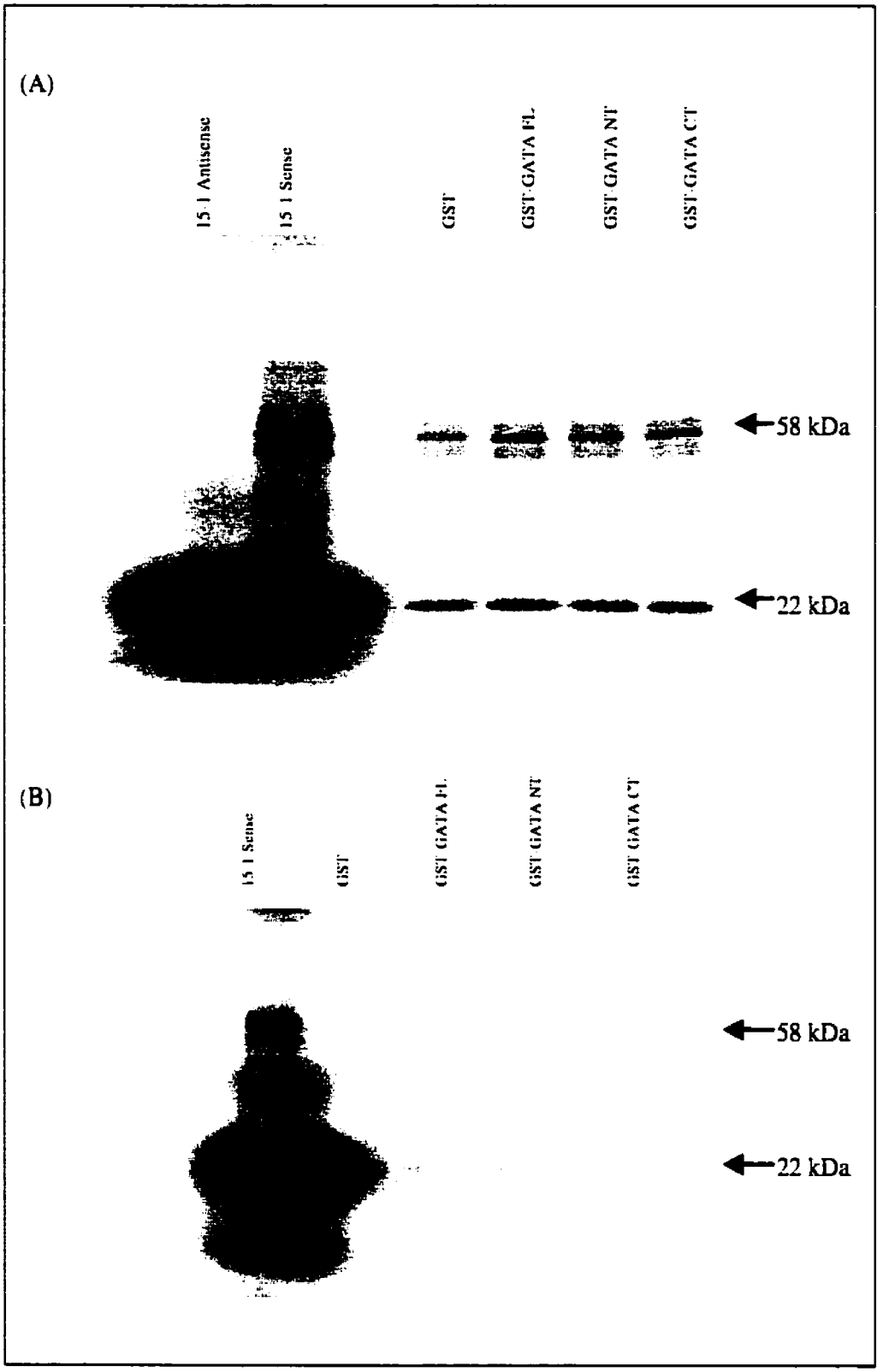
CTP 15-1 was cloned into pBluescript, and *in vitro* transcriptions were run from the T3(sense) and T7 (antisense) promoters. These products were verified by gel analysis (Figure 17a) before being used as templates in *in vitro* translations with ^{35}S-methionine (Figure 17b). Only the translation of the sense strand produced any protein products, but this result is obscured in Figure 17 due to the comigration of unincorporated ^{35}S-methionine. The labeled proteins derived from the sense RNA template were then added to the columns containing the GST- BmGATAß fusions, washed several times, and separated by SDS PAGE. It was found (Figure 18a) that the resulting proteins bound equally to all of the columns, including GST alone. Repeated attempts at washing under higher stringency (higher temperature, lower salt concentrations), and in the presence of blocking agents (BSA) were not successful in solving this problem (Figure 18b). Possible reasons for these observations are detailed in the Discussion.

Figure 18: GST Pull Down Experiments

Panel A : Following purification of the GST-BmGATAß fusion proteins, and binding of these to GST Sepharose 4b, 5 µL aliquots of the *in vitro* translated preys were passed over the matrix and allowed to bind. Following washing these complexes were eluted by boiling in PAGE buffer prior to electrophoresis.

The products separated by SDS-PAGE are the result of the sense transcript translation shown in Lane 2. All of the pull downs demonstrated equal binding for the translated prey. The size of the expected 22 kDa protein is designated by the arrow to the right of the figure. The identities of the selected GST fusions are given at the top of the figure.

Panel B : The identical experiment shown in A, except that 2 µL BSA (10mg/mL) was added to each GST sepharose preparation prior to the addition of the translated preys. This treatment reduced binding to all preparations equally.


(A)
15-1 Antisense
15-1 Sense
GST
GST-GATA FL
GST-GATA NT
GST-GATA CT
58 kDa
22 kDa
(B)
15-1 Sense
GST
GST GATA FL
GST GATA NT
GST GATA CT
58 kDa
22 kDa

SECTION IV : DISCUSSION

One of the difficulties in using atypical organisms like *B. mori* is the unavailability of recombinant techniques in the analysis of gene function. For example, it is currently not possible to evaluate gain of function in this system through misexpression of transcripts of interest or evaluate loss of function through gene knockouts. Within the constraints of our model system, gene (and product) function must be elucidated through the examination of *in vivo* expression and *in vitro* confirmation. Accordingly, the purpose of this thesis has been to detail the procedures used in the analysis of several clones, isolated from a follicular specific cDNA library, whose properties in yeast suggest they may interact with the BmGATAß transcription factor.

BmGATAß is a strong transcriptional activator, displaying both the ability to interact directly with the late chorion promoters (Drevet et al., 1994) and induce transcription from them in the silkmoth - derived Bm5 cell line (Lunke, 1999). These two activities, DNA binding and transcriptional activation, make this factor a poor choice as bait in nuclear based yeast two hybrid systems. The His 3 and ß-galactosidase reporter genes in the yeast host are quite "leaky" and permit basal levels of activation with a wide variety of baits (Vidal and Legrain,1999). This promiscuity is most easily represented by (1) Gal4 DNA binding domain fusions with proteins (or peptides) that have the ability to contact the activation domains of the transcription complex (promiscuous bait), and (2) Gal4 activation domain fusions with proteins that have DNA binding properties (promiscuous preys). Both of these types of "false positives" have been revealed during the course of the research detailed here, and much of the work represents attempts to limit, identify, and remove this background in order to establish appropriate conditions for library screening and verification of the nature of the observed interactions.

The cloning of full length BmGATAß into the bait fusion vector resulted in a product with extremely high activation potential. This activity was strong and consistent enough

to be used as a positive control during selection and colour reaction trials (Figure 6b; Results). This problem was overcome through the use of the Exonuclease III mutagenesis that allowed both the generation of non-activating baits and the tentative identification of activating domains within BmGATAß itself. Neither the N-terminal nor C-terminal baits alone demonstrated any ability to overcome the histidine repression or activate the ß-galactosidase reporter during several screening and confirmatory trials. This argues strongly in favor of the decision to use these deletion products as baits for the library screens, and suggests that the interactions observed in the yeast host are the result of direct interactions with the recovered preys. It must be noted, however, that these interactions may not represent authentic relationships, since the system does not restrict protein misexpression (temporally) or localization (cytoplasmic or membrane proteins that are transported to the yeast nucleus). Thus, the verification of preys isolated from our cDNA library must also demonstrate that the proteins under investigation would normally have the opportunity to interact with BmGATAß at the appropriate time and place. This issue was addressed following sequencing through developmental and tissue specific Northern analyses.

The library screens were optimized and evaluated in order to ensure both proper representation of clones and selection conditions. The libraries (built directionally using follicular cell derived poly $(A)^+$ RNA; K. Ito, unpublished results) were initially screened at a density of approximately 50,000 clones per plate. Forty plates were screened per clone (2×10^6 transformants per library screen) with an average insert size ranging from 0.7 - 2.0 kb . Thus while these libraries were not normalized (except for the smaller chorion transcripts that were excluded from the choriogenic library; K. Ito, unpublished results), one might expect that even relatively rare transcripts will be represented.

In order to establish proper conditions for library screening, pilot experiments were done using activators of various strengths. Conditions were optimized specifically for

the repression of the leaky His 3 selection gene by the histidine pathway antagonist 3-aminotriazole (3-AT). The inclusion of 3-AT into the media at 20 mM concentration was sufficient to lower background colonies to near zero. The colonies surviving the selection should, therefore, result strictly from the activation of the selector gene by promiscuous or genuine bait:prey interactions at the His 3 promoter.

Another manner in which the success of the screens may be evaluated is the number of positives isolated from them. The total number of positives isolated from 3 library screens was 42 (30 for the N-terminal and 12 for the C-terminal bait). While there is no quantitative or statistical device that can be used to evaluate these numbers, 42 surviving colonies from six million transformants does not appear excessive. Indeed, once the promiscuous preys were removed through confirmatory secondary analyses in yeast, this number was reduced to seven. This suggests that the library screens were sensitive enough to recognize possible *bona fide* interactions and selective enough to remove most promiscuous and non-relevant clones.

The final point that should be raised about the bait and library screens concerns the concept of saturation. While we may be able to demonstrate that the all of the isolated clones have the ability to activate the reporters (alone or in combination), there is no way to tell if all of the relevant interactors present in the libraries were identified. One method, which has been suggested to achieve this, is the recovery of the same prey within the context of a single library screen. This condition was not satisfied by any of the screens done with either bait, as upon sequencing, the isolated cDNA clones never revealed the same sequence twice. Given the strong and reproducible reporter activations by the selected clones, these screens were probably not run to saturation and potentially interesting and relevant clones (suggested in the Introduction) have yet to be cloned from these libraries.

Sequencing of the seven clones that were selected for further evaluation revealed a number of interesting, and a few not so interesting preys. CTP 1-1 is a partial cDNA for *Bombyx* alpha tubulin. Previous studies on this protein (Itachouf-Gheras et al., 1998) demonstrated that expression during development is ubiquitous, and that expression is highest as 20E levels reach their peak (but expression does not appear to be dependent on 20E). Tubulins act primarily as structural proteins, forming homo and hetero dimers to form microtubule assemblies. There has yet to appear a published report on the binding of transcription factors to alpha tubulin and, with the notable exception of the proposed BmGATAß cytoplasmic anchor, there does not appear to be a putative function for such an association. However, since this protein is not specifically localized to the cytoplasm and would presumably also form part of the nuclear architecture, its interaction with BmGATAß may be deserving of further examination. Furthermore, sex and tissue specific isoforms of ß-tubulin have been found in *B. mori* (Mita et al., 1995), although the nature and purpose of this expression is unclear.

The protein encoded by cDNA clone NTP 1-2 appears to be the *Bombyx* paralogue of *Drosophila* ribosomal protein L34. This component of the 60 S ribosomal assembly, that has been cloned in several organisms, displays a wide variety of possible functions. In bacteria it appears to regulate the biosynthesis of polyamines and the activity of arginine decarboxylase (Panagiotidis and Canellakis, 1984), while in HeLa cells it has been identified as a potent inhibitor of the cell cycle kinase CDK5 (Moorthamer and Bhabatosh, 1999). Tissue specific Northern analysis indicates that the NTP 1-2 mRNA is expressed ubiquitously in *Bombyx* pupae (Figure 8), while developmental Northern analysis demonstrates a significant downregulation in the follicular epithelium just prior to the onset of choriogenesis (between stages -10 and -5), with expression continuing to drop to very low levels by stage +21 (Figure 7). This is interesting in that it coincides with the induction pattern of BmGATAß expression, but could just as easily be correlated with the commitment to the expression of the chorion genes.

The NTP 8-11 polypeptide is the *Bombyx* paralogue of the t-complex polypeptide 20, a member of the CCT chaperonin complex. This rather generic housekeeping gene is involved in protein folding during translation, but also involved in the regulation of post translational modifications and protein targeting (Leroux and Hartl, 2000). Interestingly, the primary targets of these complexes appear to be the tubulins. They are, however, also involved in the folding of several other classes of proteins. Other functions may include the assembly of protein complexes, because mutations that impair its ability to properly contact its substrates result in non-active complexes; one cited example of such a case is the formation of the VHL tumor suppressor complex, where mutations within the von Hippel Lindeau syndrome map to the TriC/CCT complex (Feldman et al., 1999). Like NTP 1-2, NTP 8-11 is down regulated at the -10 stage of late vitellogenesis. While it is intuitively attractive to link the expression of this protein with the downregulated potential primary substrate alpha tubulin, there is no evidence to support such an association (functionally or physically). It is also unclear at this point what this system may have to do with the control of BmGATAß activity.

The protein encoded by the NTP 11-3 clone is becoming a bit of an enigma. Much like NTP 1-2 and NTP 8-11 , expression of NTP 11-3 is downregulated prior to choriogenesis, but appears to occur very closely to the time of BmGATAß upregulation (approximately stage -5; Figure 7). Sequence database searches originally turned up no homologues or conserved motifs and, since NTP 11-3 mRNA virtually disappears when BmGATAß is presumed to be active, this clone was not pursued with much vigor. Two more recent developments may indicate, however, that this decision was premature. Tissue specific Northern analysis (Figure 8) demonstrated that this gene produces two relatively rare transcripts that are restricted specifically to the ovary (and probably to the follicular epithelium). A very recent database search finally turned up a conserved domain with an uncharacterized genomic sequence within the *Drosophila* databases (Appendix II). Only a small (20 amino acid long) region appears to be similar with a

region of NTP 11-3, but because this occurs at the 5' end of the recovered two hybrid clone it is not possible to relate this domain conservation with gene or protein homology. While PSI and PHI Blast searches (reiterative searches based on conserved domains) did not produce any other matches, links to related sequences reveal a large and expanding class of tumor suppressor genes with DNA helicase activities. The closest related sequence arises from the *xnp-1* gene of *Xenopus laevis*, that has tentatively been designated to be a member of the SWI/SNF family of C_2C_2 zinc finger containing DNA helicases (Biggar and Crabtree, 1999). These factors activate transcription through nucleosome and chromatin remodeling, and are found to be essential for the appropriate expression of a wide variety of genes. It has also been suggested that these complexes are required for the appropriate repression of genes (Sudarsanam and Winston, 2000), and that they are involved in a wide variety of developmental and differentiation pathways. Given the extremely limited amount of sequence information available for NTP 11-3, however, all of this is conjecture and does not immediately suggest a functional association with BmGATAß.

NTP 1-4 was easily the most enthusiastically pursued clone pulled from the two hybrid libraries. Control transformations done for this clone in yeast demonstrated both the fastest and most consistently strong activation of the yeast reporter genes in conjunction with the N-terminus of BmGATAß (Figure 6). The expression profiles of this transcript are dramatic (Figure 10), with a high and abrupt induction (and subsequent down-regulation) precisely at the point of BmGATAß upregulation (vitellogenic stage -6). Tissue specific Northern analysis indicates that its expression is limited to the ovaries during pupal development. Confirmation of these results has also been provided by the *in situ* hybridizations done against its transcript (Figure 12). Expression is localized to the follicular epithelium and the results have demonstrated a consistent and rapid induction over the entire surface of the oocyte (follicles at stage -7 clearly do not express the message while stage -6 follicles are already nearly saturated with signal; Figure 12). This type of expression would be expected for genes presumed to be within the sphere of the

autonomous differentiation model currently held for this system, and its restriction to the follicular epithelium and developmental regulation make it a prime candidate for a product involved in the terminal differentiation (re: chorion synthesis) of these cells.

The relevant question regarding its presumptive role as a BmGATAß interacting and modulating factor has not been answered. Blast and other homology searches produced inconclusive results (Appendix II). While a number of interesting genes are highlighted from the Blast list, no significant regions or domains show homologies with sequences within the current databases, and the conceptually translated product does not yield a recognizable secondary structure with the EMBL or Genomenet software packages. Revelations concerning this clone will have to arise from additional sequence information. The 5' RACE products that were obtained through this work (Appendix I) were not sufficiently long to provide this information, and the products generated from the RT reactions could not be cloned and sequenced prior to the completion of this thesis.

The final clone examined in this study, CTP 15-1, was the single C-terminus interacting clone that was isolated from the choriogenic library and survived the selection procedures. The developmental and tissue specific Northern analyses (Figure 11) have shown this transcript to be choriogenic-specific and restricted to the pupal ovary and brain. *In situ* hybridization analysis (Figure 13), while not as clean as that for NTP 1-4, has clearly demonstrated follicular specific expression (honeycomb pattern), relative lack of staining in the nurse cell cap, and developmental restriction to choriogenic follicles. The reproducibility of the Northern analyses and comparisons with adjacent stained and unstained follicles from the *in situ* hybridizations suggest that this transcript is expressed almost exactly at stage +1. As far as I am aware, no other genes or clones isolated from research on *Bombyx* oogenesis other than the early chorion genes themselves have demonstrated such temporally defined expression patterns. Homology searches against this clone have shown that it is clearly not a chorion polypeptide, an observation that is substanciated by the expression of its transcript throughout the entire spectrum of

choriogenesis (earlier developmental Northerns (not shown) have demonstrated expression of this transcript up to and beyond developmental stage +35).

Blast searches against this sequence revealed a large number of unremarkable genes. One exception , however, is a single curious stretch of amino acids (Appendix II) from one of these Blast members that shows the highest "domain-type" homology with CTP 15-1, and this is the Polycomb type Pc1 homologue of *Xenopus laevis*. The Polycomb group of proteins have long been implicated in the control of gene families or arrays, most notably in their association with the trithorax transcriptional activators and the bithorax gene cluster that controls *Drosophila* segmentation (Orlanda et al., 1998). Although these proteins have never been shown to bind promoters or DNA in general, they appear to function through promoter-specific response elements (PREs), and do so only in association with transcription factors and modulators (Kingston et al., 1996). Current reasoning places the polycomb proteins within multimeric protein complexes that assemble on promoters and designate them as transcriptionally active or repressed. The nature of the factors associating with the Polycomb group proteins suggests that these complexes are involved in chromatin remodeling. CTP 15-1 is the second clone whose closest sequence match appears to be a chromatin remodeling factor (see above, NTP 11-3). These correlations become even more interesting if one considers that the SWI/SNF complex members are also viewed as trithorax group genes, whose members act as chromatin opening complexes. The possible relevance of these coincidences is expanded upon in the Conclusion section of this thesis. As with NTP 1-4, however, although these correlations make an attractive case for CTP 15-1's involvement in the choriogenic cascade, they do not provide any direct evidence for a functional *in vivo* association with BmGATAß.

Given the promiscuity and non-specificity problems that the two hybrid system presents, it was necessary to confirm the physical association seen in yeast by alternative *in vivo* or *in vitro* experiments. Several attempts were made to demonstrate the

BmGATAß-Cterminus: CTP 15-1 interaction using an *in vitro* GST pull down assay. While none of the fusion bait proteins produced were particularly soluble, enough of each was produced to be visible on a Coomassie blue stained gel (Figure 16). *In vitro* transcribed and translated CTP 15-1 (Figure 17) results in one major 22 kDa band corresponding to the approximate size of the expected product, and several higher molecular weight products that may arise through higher order interactions. Unfortunetly, none of the pull down experiments were successful in demonstrating specific interactions between the bait fusions and the labeled preys (Figure 18). It would appear that the *in vitro* translated products are "sticky" in nature and bind equally well to the fusions and GST alone. This is not totally unexpected, as the translation products have no cues or cofactors that could allow them to fold properly, which may result in non native protein conformations. These products might interact with surfaces that stabilize these conformations and, in this respect, it is perhaps not surprising that attempts to limit the observed non specificity by blocking agents such as BSA (which would only remove non-specific protein-protein interactions) were unsuccessful.

The best solution to this problem would make use of available BmGATAß N-terminal and C-terminal specific antibodies (Lunke; 1999). Follicular cell extracts passed over GST-prey fusion columns could reveal BmGATAß specific interactions through Western blotting using these antibodies. As many types of proteins are faithfully produced by the bacterial expression systems, and the BmGATAß would be in native form, this type of analyses has a much greater chance of succeeding.

Conclusions and Perspectives for Future Studies

The primary purpose of this thesis was the isolation and initial characterization of gene products that directly interact with , and modulate the activity of, BmGATAß. As there are no direct methods available to analyze the gain or loss of function of genes in *Bombyx*, realization of these thesis goals was difficult. However, the following objectives have been accomplished:

1) Isolation, testing, and use of various BmGATAß deletion products as baits in two hybrid screens.
2) Recovery and confirmation of several cDNA clones encoding potential BmGATAß interacting proteins.
3) Determination of developmental and tissue specific expression profiles for 5 of these selected clones: NTP 1-2, NTP1-4, NTP 8-11, NTP 11-3 and CTP 15-1.
4) *In situ* hybridizations for two of these clones, NTP 1-4 and CTP 15-1.
5) Isolation of 5' RACE and RTPCR products towards generation of full length cDNAs for NTP 1-4 and CTP 15-1.

Given the list of potential interactors detailed in the Introduction, it would be convenient to be able to assign any or all of these clones into the categories of factors assumed to modulate BmGATAß function, localization, or activity. At this point, however, this is not possible, due to incomplete characterization of the cloned sequences.

Designating the dozens of isolated two hybrid clones into potential interactors or potential background has been the most difficult task faced during this research. Factors whose interactions are only transitory (kinases, phosphatases, etc.) may well have been

missed due to the limitations of sensitivity inherent in the two hybrid system. The DNA binding activity of secondary transcription factors may also have led to their exclusion, since the introduction of a DNA binding domain onto the N-terminus of the Gal4 activation domain would possibly result in a promiscuously active prey. On the basis of the results presented in this thesis, I would suggest that there are three clones whose characteristics make them prime candidates for further analysis: NTP 1-4, NTP 11-3, and CTP 15-1. Although no lengthy homologies have been found for these sequences, the nature of their expression and putative sequence homologues suggest that they may either directly interact with BmGATAß, or are involved more generally in the commitment of the cells of the follicular epithelium to the synthesis of the *Bombyx* chorion.

NTP 1-4, as stated in the Discussion, is a prime candidate for subsequent research. Its developmental expression pattern shows that it is induced at high levels exactly at the same time as BmGATAß upregulation, it activates consistently and specifically the yeast reporter genes only in the presence of BmGATAß, and its expression is both ovary specific and confined to the follicular epithelium. These characteristics would be expected for only two classes of factors suggested in the Introduction, a BmGATAß specific kinase or an early chorion gene transcription factor. The recovery of either of these entities would significantly improve the resolution of the proposed model, and would provide for interesting and relevant research. On this basis, this is probably the most significant clone recovered, and screening for the full length cDNA is the next logical step in this research. Although there was enormous background produced when the lambda gt11 vitellogenic cDNA library was screened, this may have resulted from sequences in the 3' end or UTR of this transcript; a more careful choice of probe would very likely result in the recovery of the full length (or nearly full length) cDNA. Previous screens of this library have been quite successful, and, given the abundance of its transcript during late vitellogenesis, I would assume that the desired clones are there.

I am hesitant to comment on the NTP 11-3 and CTP 15-1 clones due to the very limited amount of available sequence information, and the low homology they show to the sequences to which they have been compared. However, should NTP 11-3 turn out to be a SWI/SNF paralogue and CTP 15-1 turn out to be a Polycomb group gene, this would not only strengthen the results presented here but would greatly expand the model for chorion gene expression control. The chorion loci would be expected to be regulated in at least three different ways: their expression must be restricted physically and temporally to the choriogenic oocyte and follicular epithelium (since these genes are not required by any other *Bombyx* tissue), they must be activated at the appropriate time, (as detailed in the Introduction), and they must be repressed as the subsequent class of chorion genes is activated. In this paradigm the actions of a SWI/SNF complex would be to bind to protein complexes on the chorion promoters and mark them for subsequent expression (Orlando et al., 2000). Once expressed, the actions of a Polycomb group protein would serve to repress this activity and maintain this repression for the duration of choriogenesis. It is also worthwhile to point out that neither of these proteins have known direct promoter binding activity, and that they appear to function only in conjunction with their appropriately expressed transcription factors. BmGATAß would then assume the essential role of simultaneously designating which target genes are to be activated and which are repressed, with post-translational modifications and isoform selection providing the mechanism for stage selection. Phosphorylated BmGATAß or isoforms of it would change the nature of interactions both with their designated promoters, and with the activating and repressing complexes that control chorion gene expression at the level of chromatin structure. This concept also fits nicely with the observations that the late chorion genes have the potential to form secondary structures themselves (Iatrou and Tsitilou, 1983), and thus suggest that their expression is not simply the result of transcription factor binding, but instead the result of interactions within higher order protein and chromatin structure.

The number of extensions required to assess the validity of these proposals is too large to be detailed here. As with NTP 1-4 the first direct extension to this research is the cloning and characterization of the full length cDNAs, which will provide both additional sequence information and designate possible paralogues and associated function. These clones may well fit into bandshift analysis, as they might not be expected to bind the chorion promoters themselves, but would supershift a promoter construct that already had bound BmGATAß. This would serve both to demonstrate BmGATAß specific interactions and demonstrate the formation of higher order structures presumed to form on the chorion promoters *in vivo*.

The final extension to this research is the repetition of the GST pull down assays mentioned in the Discussion section of this thesis. I feel that it would be a reasonably simple and straightforward series of experiments to produce GST-preys, pass follicular extracts over them bound to GST sepharose, and demonstrate binding and specificity with the current BmGATAß antibodies. Along with the cloning of the full length cDNAs, the bandshift and pull down experiments would serve to complete the initial characterization of these clones.

References

Andrews J, Garcia-Estefania D, Delon I, Lu J, Mevel-Ninio M, Spierer A, Payre F, Pauli D and Oliver B. (2000). OVO Transcription Factors Function Antagonistically in the *Drosophila* Female Germline. *Development.* 127: 881 - 892.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K (1999). Current Protocols in Molecular Biology; Supplement 37. John Wiley and Sons, Inc.

Beck T. and Hall M.N. (1999). The TOR Signaling Pathway Controls Nuclear Localization of Nutrient-Regulated Transcription Factors. *Nature.* 402: 689 - 692.

Bock SC, Campo K, and Goldsmith MR. (1986). Specific Protein Synthesis in Cellular Differentiation. Temporal Expression of Chorion Gene Families in *Bombyx mori* Strain C108. *Developmental Biology.* 117: 215 - 225.

Biggar SR, and Crabtree GR. (1999) Continuous and Widespread Roles for the SWI/SNF Complex in Transcription. *EMBO J.* 18(8): 2254 - 2264.

Brewer AC, Guille MJ, Fear DJ, Partington, GA and Patient RK. (1995). Nuclear Translocation of a Maternal CCAAT Factor at the Start of Gastrulation Activates *Xenopus* GATA-2 Transcription. *EMBO J.* 14(4): 757 - 766.

Cardoso C, Timsit S, Villard L, Khrestchatisky M, Fontes M and Colleaux, L. (1998). Specific Interaction Between the XNP/ATR-X Gene Product and the SET Domain of the Human EZH2 Protein. *Human Molecular Genetics.* 7(4): 679 - 684.

Drevet J.R., Skeiky Y.A.W., and Iatrou K. (1994). GATA-Type Zinc Finger Motif Containing Sequences and Chorion Gene Transcription Factors of the Silkworm *Bombyx mori. J. Biol. Chem.* 269 (14): 10660 - 10667.

Eickbush TH and Kafatos FC. (1982). A Walk in the Chorion Locus of *Bombyx mori. Cell* 29: 633 - 643.

Eickbush T.H. and Izzo J.A. (1995). Chorion Genes: Molecular Models of Evolution. In: Molecular Model Systems in the Lepidoptera, M.R. Goldsmith and Wilkins A.S., editors, Cambridge university Press, pp. 217 - 248.

Elefanty AG, Antoniou M, Custodio N, Carmo-Fonseca M, and Grosveld FG. (1996). GATA Transcription Factors Associate With a Novel Class of Nuclear Bodies in Erythroblasts and Megakaryocytes. *EMBO J.* 15(2): 319 - 333.

Feldman DE, Thulasiraman V, Ferreyra RG, and Frydman J. (1999). Formation of the VHL-Elongin BC Tumore Suppressor Complex is Meditated by the Chaperonin TRi-C. *Molecular Cell.* 4(6): 1051 - 1061.

Goldsmith MR. (1989). Organization and Developmental Timing of the *Bombyx mori* Chorion Gene Clusters in Strain C108. *Dev. Genet.* 10: 16 - 23.

Gorner W, Durchschlag E, Martinez-Pastor MT, Estruch F, Ammerer G, Hamilton B, Ruis H, Schuller C. (1998). Nuclear Localization of the C2H2 Zinc Finger Protein Msn2p is Regulated by Stress and Protein Kinase A Activity.
Genes Dev. Feb 15;12(4):586-97.

Haenlin M, Cubadda Y, Blondeau F, Heitzler P, Lutz Y, Simpson P, Ramain P. (1997). Transcriptional Activity of Pannier is Regulated Negatively by Heterodimerization of the GATA DNA-binding Domain with a Cofactor Encoded by the u-shaped Gene of *Drosophila. Genes Dev.* Nov 15;11(22):3096-108.

Hibner, B.L., Burke, W.D., Lecanidou, R., Rodakis, G.C., and Eickbush, T.H. (1988). Organization and Expression of Three Genes from the Silkmoth Early Chorion Locus. *Developmental Biology.* **125**: 423-431.

Iatrou K. and Tsitilou S.G. (1983). Coordinately Expressed Chorion Genes of *Bombyx mori* : Is Developmental Specificity Determined by Secondary Structure Recognition?. *EMBO J.* 2:1431 - 1440.

Jans DA, Hubner S. (1996). Regulation of Protein Transport to the Nucleus: central role of Phosphorylation. *Physiol Rev.* Jul; 76(3):651-85. Review.

Kafatos FC, Regier JC Mazur GD, Nadel MR Blau HM, Petri WH, Wyman AR, Gelinas RE, Moore PB Paul M, Efstratiadis A, Vournakis JN, Goldsmith MR, Hunsley JR, Baker B, Nardi J, and Koehler M. (1977). The Eggshell of Insects: Differentiation Specific Proteins and the Control of their Synthesis and Accumulation During Development. In: Biochemical Differentiation in Insect Glands, W. Beerman, editor, Springer-Verlag, Berlin, pp. 45 - 145.

Kendirgi F (2000). P30 and BmOSP, Two Novel Follicular Cell Specific Proteins Involved in Silkmoth Oogenesis. Ph.D. thesis, the University of Calgary.

Kingston RE, Bunker CA, and Imbalzano AN. (1996). Repression and Activation by Multiprotein Complexes That Alter Chromatin Structure. *Genes and Development.* 10: 905 - 920.

Lajeunesse D, and Shearn A. (1996). E(z): A Polycomb Group Gene or a Trithorax Group Gene ?. *Development.* 122: 2189 - 2197.

Lecanidou R, Rodakis GC, Eickbush TH, and Kafatos FC. (1986). Evolution of the Silkmoth Chorion Gene Superfamily:L Gene Families CA and CB. *Proc. Natl. Acad. Sci.* USA 83: 6514 - 6518.

Leroux MR, and Hartl FU (2000). Protein Folding: Versatility of the Cytosolic Chaperonin TRiC/CCT. *Current Biology*. 10(7): 260 - 264.

Lunke MD (1999). Molecular Characterization of BmGATAB, a Gene Involved in Choriogenesis in the Silkworm *Bombyx mori*. Ph.D. thesis, the University of Calgary.

MacNeill C, Ayres B, Laverriere AC, Burch JB. (1997). Transcripts for Functionally Distinct Isoforms of Chicken GATA-5 are Differentially Expressed from Alternative First Exons. *J Biol Chem*. Mar 28;272 (13): 8396 - 401.

Mitsialis S.A., Veletza S., and Kafatos F.C. (1989) Transgenic Regulation of Moth Chorion Gene Promoters in *Drosophila*: Tissue, Temporal, and Quantitative Control of Four Bidirectional Promoters. *J. Mol. Evol.* 29: 486 - 495.

Moorthamer M and Chaudhuri B. (1999). Identification of Ribosomal Protein L34 as a Novel CDK5 Inhibitor. *Biochem Biophys Res Commun*. 255(3): 631 - 638.

Orlando V, Jane EP, Chinwalla V, Harte PJ, and Paro R. (1998). Binding of Trithorax and Polycomb proteins to the Bithorax Complex: Dynamic Changes During Early *Drosophila* Embryogenesis. *EMBO J.* 17: 5141 - 5150.

Orford RL, Robinson C, Haydon JM, Patient RK, Guille MJ. (1998) The Maternal CCAAT box Transcription Factor Which Controls GATA-2 Expression is Novel and Developmentally Regulated and Contains a Double-Stranded-RNA-Binding Subunit. *Mol. Cell Biol.* 18(9): 5557 - 66.

Partington GA, Bertwistle D, Nicolas RH, Kee W-J, Pizzey JA and Patient RK. (1997) GATA-2 is a Maternal Transcription Factor Present in Xenopus Oocytes as a Nuclear Complex Which is Maintained Throughout Early Development. *Dev.Biol.* 181:144 - 155.

Partington G.A. and Patient R.K. (1999) Phosphorylation of GATA-1 Increases its DNA Binding Affinity and is Correlated with Induction of Human K562 Erythroleukaemia Cells. *Nucleic Acids Research*. 27(4): 1168 - 1175.

Putterill J, Robson F, Lee K, Simon R and Coupland G. (1995). The CONSTANS Gene of *Arabidopsis thaliana* Promotes Flowering and Encodes a Protein Showing Similarites to Zinc Finger Transcription Factors. *Cell*. 80(6): 847 - 857.

Reiger JC, Mazur GD, Kafatos FC, and Paul M. (1982). Morphogenesis of Silkmoth Chorion: Initial Framework Formation and its Relation to Synthesis of Specific Proteins. *Dev. Biol.* 92: 159 - 174.

Skeiky Y.A.W., Drevet J.R., Swevers L. and Iatrou K. (1994). Protein Phosphorylation and Control of Chorion Gene Activation Through Temporal Mobilization of a Promoter DNA Binding Factor From the Cytoplasm Into the Nucleus. *J. Biol. Chem.* 269: 12196 - 12203.

Spoerel, N.A., Nguyen, H.T., Eickbush, T.H., and Kafatos, F.C. (1989). Gene Evolution and Regulation in the Chorion Complex of *Bombyx mori*: Hybridization and Sequence Analysis of Multiple Developmentally Middle A/B Chorion Gene Pairs. *J. Mol. Biol.* **209**: 1-19.

Spoerel N.A., Nguyen H.T., Towne S. and Kafatos F.C. (1993). Negative and Positive Regulators Modulae the Activity of a Silkmoth Chorion Gene During Choriogenesis. *J. Mol. Biol.* 230: 151 - 160.

Svensson EC, Tufts RL, Polk CE, Leiden JM. (1999). Molecular cloning of FOG-2: a Modulator of Transcription Factor GATA-4 in Cardiomyocytes. *Proc Natl Acad Sci* USA. Feb 2;96(3):956-61.

Swevers, L., and Iatrou, K. (1992). Early Establishment and Autonomous Implementation of a Developmental Program Controlling Silkmoth Chorion Gene Expression. *Developmental Biology*. **150**: 12-22.

Swevers.L., and Iatrou, K. (1995). Terminal Differentiation of Ovarian Follicular Cells in the Silkmoth *Bombyx mori*. Molecular Mechanisms of Insect Metamorphosis and Diapause 217-228.

Tsang A.P, Visvader JE, Turner CA ,Fujiwara Y, Yu C, Weiss MJ, Crossley M, and Orkin SH. (1993). FOG, a Mulititype Zinc Finger Protein, Acts as a Cofactor for Transcription Factor GATA-1 in Erythroid and Megakaryocytic Differentiation. *Cell* 90(1): 109 - 119.

Tsuchida K., Nagata M. and Suzuki A. (1987). Hormonal Control of Ovarian Development in the Silkworm *Bombyx mori*. *Archives of Insect Biochem. Physiol.* 5: 167 - 177.

Tzertzinis G, Malecki A and Kafatos FC. (1994) BmCFI, a Bombyx mori RXR-type Receptor Related to the *Drosophila* Ultrascpiricle. *J. Mol. Biol.* 238: 479 - 486.

Vidal M and Legrain P (1999) Survey and Summary : Yeast Forward and Reverse 'n'-hybrid Systems. *Nucleic Acids Research.* 27(4) : 919 - 929.

von Mikecz A, Neu E, Krawinkel U and Hemmerich P. (1999). Human Ribosomal Protein L7 Carries Two Nucleic Acid Binding Domains with Distinct Specificities. *Bioch and Biophys Res Comm.* 258: 530 - 536.

Whyatt DJ, deBoer E and Grosveld F (1993). The Two Zinc Finger-Like Domains of GATA-1 have Different DNA Binding Activities. *EMBO J.* 12: 4993 - 5005.

Yang H.Y. and Evans T. (1992). Distinct Roles for the Two cGATA-1Finger Domains. *Mol.Cell.Biol.* 12: 4562 - 4570.

Zhu J, Fukushige T, McGhee JD and Rothman JH. (1998). Reprograming of Early Embryonic Blastomeres into Endodermal Progenitors by a *Caenorhabtidis elegans* GATA Factor. *Genes and Development.* 12: 3809 - 3814.

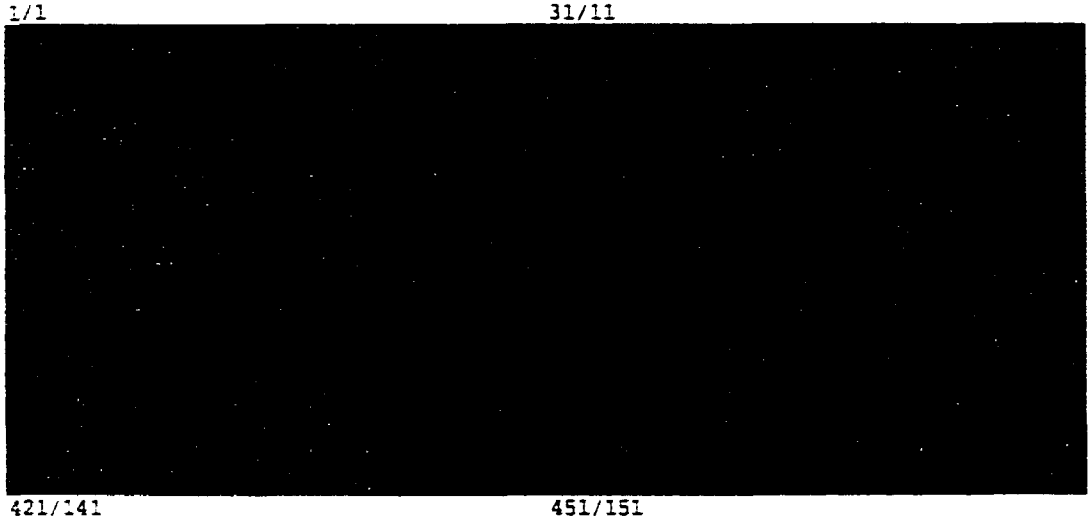
Appendix I: Sequence Information

Sequences and conceptual translations of clones isolated through the yeast two hybrid library screens. Positions of primers used for RACE amplification of NTP 1-4 and CTP 15-1 are included in their sequences, and the 5' sequence information obtained from these experiments is outlined. Only confirmed and reliable sequence information from the RACE amplifications are included.

ORF for NTP 1-2 : Ribosomal Protein L34

```
1/1                                                 31/11
CCG CCG CAC AGA AAA TTT TCA GCA CCC CGT CAT GGG TCT ATG GGA TTC TAT CCC AAA AAG
P   P   H   R   K   F   S   A   P   R   H   G   S   M   G   F   Y   P   K   K
61/21                                               91/31
AGG TCC CGT CGT CAT CGT GGT AAG GTC AAG GCG TTC CCG AAA GAC GAC CCT AGC AAA CCT
R   S   R   R   H   R   G   K   V   K   A   F   P   K   D   D   P   S   K   P
121/41                                              151/51
GTT CAT TTG ACT GCT TTT ATC GGT TAT AAG GCC GGT ATG ACC CAC GTG GTT AGA GAA CCT
V   H   L   T   A   F   I   G   Y   K   A   G   M   T   H   V   V   R   E   P
181/61                                              211/71
GAC CGT CCC GGT TCA AAA ATC AAC AAG AAA GAG ATC GTG GAG GCT GTC ACC ATC ATC GAG
D   R   P   G   S   K   I   N   K   K   E   I   V   E   A   V   T   I   I   E
241/81                                              271/91
ACT CCT CCG ATG GTT TGT GTC GGT GTT GTT GGA TAC ATT GAG ACC CCT CAT GGT CTA CGC
T   P   P   M   V   C   V   G   V   V   G   Y   I   E   T   P   H   G   L   R
301/101                                             331/111
GCT CTG TTG ACT GTC TGG GCG GAG CAT ATG TCT GAA GAT
A   L   L   T   V   W   A   E   H   M   S   E   D
```

ORF and 3' UTR for NTP 1-4



```
1/1                                          31/11

421/141                                      451/151
                                                        GSP : 1109
GCT ATG CCT AAC TCT AAG AAC AAC GAA ATG AGT ACA GCA CAA TCC AAA TTT GAG GAG TCC
A   M   P   N   S   K   N   N   E   M   S   T   A   Q   S   K   F   E   E   S
481/161                                      511/171
                                  GSP : 1110
AAT AAT TTT CGA ATT GCA AAT CAC AAT AGT CAG TCA GTG GGT CAC CAT CAC ACA TCA ACA
N   N   F   R   I   A   N   H   N   S   Q   S   V   G   H   H   H   T   S   T
541/181                                      571/191
ATG AAG GAT GTA AAT CAT TTA GAT ACG ATG AGA ACT CAT ATG CGC ATG GAC CCT GTA CAT
M   K   D   V   N   H   L   D   T   M   R   T   H   M   R   M   D   P   V   H
601/201                                      631/211
AAA AAT TTA AAT GAT CAT TTA ATT CAT GGA CAA GTT TAT CAC GAT GAT CAC ATT GAT TAC
K   N   L   N   D   H   L   I   H   G   Q   V   Y   H   D   D   H   I   D   Y
661/221                                      691/231
TGC TTT TCA TTC CAA CAT GAG GGA TTC CTT AAA TGG CTT TAG ACC TGA TAA TAC AAT GTT
C   F   S   F   Q   H   E   G   F   L   K   W   L   *   T   *   *
721/241                                      751/251
CCA ATG GAA GCC TAG TCA AGA GAG AAC CGT ATA TGG CCC TGC AAT GTC CTC TCA CAC GGG

781/261                                      811/271
ATC AAG TGC TGT CGG CGT ATT TCC AAA GGC AAA CAT CGG GAG TTG TGG AA
```

ORF and 3' UTR for two hybrid clone NTP 8-11

```
1/1                                                 31/11
TCT GTT ACC ATC CTC AAG GGA CCC AAC AAG CAT ACC TTG ACA CAA ATC AAA GAT GCA GTT
S   V   T   I   L   K   G   P   N   K   H   T   L   T   Q   I   K   D   A   V
61/21                                               91/31
CGT GAC GGC CTC CGC GCT ATT AAT AAC GCC ATT GAA GAT AAA TGC GTC ATA CCC GGT GCT
R   D   G   L   R   A   I   N   N   A   I   E   D   K   C   V   I   P   G   A
121/41                                              151/51
GCG GCC TTC GAA ATC AAA GCT AAC ACG GAA CTA CTG AAA TAT AAG GAC ACT GTT AAA GGC
A   A   F   E   I   K   A   N   T   E   L   L   K   Y   K   D   T   V   K   G
181/61                                              211/71
AAG GCT CGT CTT GGA ATC CAA GCT TAT GCC GAA GCT CTT TTG GTT ATT CCT AAA ACT CTG
K   A   R   L   G   I   Q   A   Y   A   E   A   L   L   V   I   P   K   T   L
241/81                                              271/91
GCC GTT AAC TCT GGT TAC GAT GCT CAA GAT ACT ATA GTA AAG TTG CAG GAA GAG TCA CGC
A   V   N   S   G   Y   D   A   Q   D   T   I   V   K   L   Q   E   E   S   R
301/101                                             331/111
CTT AAT CCT GAG CCT ATT GGT TTA GAT TTG AGT ACA GGT GAA GCA TTC AAA CCT ACA GAC
L   N   P   E   P   I   G   L   D   L   S   T   G   E   A   F   K   P   T   D
361/121                                             391/131
TTA GGT ATT TTG GAC AAC TAT ATC GTG AAA AAA CAA ATA TTG AAT TCC TGT TCA GTG ATC
L   G   I   L   D   N   Y   I   V   K   K   Q   I   L   N   S   C   S   V   I
421/141                                             451/151
GCA AGC AAC CTC CTT GTA GAC GAG ATT ATG AGA GCA GGA ATG AGC AGC CTC AAA GGA TAA
A   S   N   L   L   V   D   E   I   M   R   A   G   M   S   S   L   K   G   *
481/161                                             511/171
TAA TAA ACA AAT GTT TTA TTT TGA GGT CAT TCG AGT TTC TGC TAT TAC TTT CAC TCA TTT
*   *
541/181                                             571/191
ATC ATT TTG CTT TCA ACT GTA ACT GCA TTT TTA TGT AGA ATT GTC TTA TTA CAA TCT ACG

601/201
CCA AAC TTC GCA TTT CTG ATT A
```

ORF for NTP 11-3

```
1/1                                                 31/11
CGC CTG ATC AAG AAG GAA CAA CAT CTG GAC CGG GTC ACC ACC AAA ATG TTC AAG AAA GTC
R   L   I   K   K   E   Q   H   L   D   R   V   T   T   K   M   F   K   K   V
61/21                                               91/31
ACT ACT GCT GAA AAA GAA AAT CAA TGG CGT GAA GAG ATG AGT GAA GGT CTG CCG AAA CCA
T   T   A   E   K   E   N   Q   W   R   E   E   M   S   E   G   L   P   K   P
121/41                                              151/51
CAT AAC CCA GCG GAA CAT GAC TCC GAA TCC GAC GGA GAG TAC AAA GCC ATC AAC CCT CCC
H   N   P   A   E   H   D   S   E   S   D   G   E   Y   K   A   I   N   P   P
181/61                                              211/71
GTT AAG AAC AAG AAG AAA GAT CAC AAG GCT CGA AGG AAA CAG AAG GAG AGA CTG TTG GAA
V   K   N   K   K   K   D   H   K   A   R   R   K   Q   K   E   R   L   L   E
241/81                                              271/91
AAG GAA CGC CTG AAA AGG GAA AAG ATT GAC AAA AAG AAG ATA ACT GAT ATT TAC AAG CTT
K   E   R   L   K   R   E   K   I   D   K   K   K   I   T   D   I   Y   K   L
301/101                                             331/111
CGG AAA CTG GAG TCT TCG ATC AAG AAG GGC GAG AGG GTT CGC GAG GCT CAG CGC GCA CAG
R   K   L   E   S   S   I   K   K   G   E   R   V   R   E   A   Q   R   A   Q
361/121                                             391/131
CGA GCC GCC CGC CGC GCC GCC CGC GCC CGC GCC GCG CCC CCC GCA CTC AGC GCG CAC CCC
R   A   A   R   R   A   A   R   A   R   A   A   P   P   A   L   S   A   H   P
421/141                                             451/151
GCG CCC GCC CCC GCC CCG CCG CTC GTC GCG CCC GCG CAG CTC TCC GGC GAC CTG CGC CGC
A   P   A   P   A   P   P   L   V   A   P   A   Q   L   S   G   D   L   R   R
481/161                                             511/171
GTC GCG CCC GCC GTG AAC CTT CTC CGC GAA CGT TTC GAG
V   A   P   A   V   N   L   L   R   E   R   F   E
```

ORF for CTP 1-1 : *Bombyx mori* alpha-tubulin

```
1/1                                                 31/11
TTT AAT TCT ACA TTT AAG CAT AGA AAA AAA CTC AAA TCA AAA TGC GTG AGT GCA TCT CTG
F   N   S   T   F   K   H   R   K   K   L   K   S   K   C   V   S   A   S   L
61/21                                               91/31
GTA CAC GTT GGC CAA GCC GGA GTC CAG ATC GGT AAT GCC TGC TGG GAG CTT TAC TGC CTG
V   H   V   G   Q   A   G   V   Q   I   G   N   A   C   W   E   L   Y   C   L
121/41                                              151/51
GAG CAC GGC ATC CAG CCT GAT GGC CAG ATG CCC ACA GAC AAG ACC ATC GGG GGT GGA GAC
E   H   G   I   Q   P   D   G   Q   M   P   T   D   K   T   I   G   G   G   D
181/61                                              211/71
GAT TCT TTC AAC ACT TTC TTC AGC GAG ACC GGA GCT GGC AAG CAC GTA CCC CGT GCT GTC
D   S   F   N   T   F   F   S   E   T   G   A   G   K   H   V   P   R   A   V
241/81                                              271/91
TTC GTC GAT CTT GAA CCT ACT GTT GTT GAT GAG GTC CGC ACT GGC ACA TAC AGA CAG TTG
F   V   D   L   E   P   T   V   V   D   E   V   R   T   G   T   Y   R   Q   L
301/101                                             331/111
TTT CAT CCA GAA CAA CTT ATT ACT GGT AAG GAA GAT GCG GCC AAC AAC TAT GCC CGT GGT
F   H   P   E   Q   L   I   T   G   K   E   D   A   A   N   N   Y   A   R   G
361/121                                             391/131
CAC TAC ACC ATT GGA AAG GAA ATC GTA GAT TTG GTT TTG GAC AGA ATC CGC AAG CTC GCT
H   Y   T   I   G   K   E   I   V   D   L   V   L   D   R   I   R   K   L   A
421/141                                             451/151
GAC CAG TGT ACC GGT CTG CAA GGA TTC CTG ATC TTC CAC TCC TTC GGT GGA GGT ACC GGC
D   Q   C   T   G   L   Q   G   F   L   I   F   H   S   F   G   G   G   T   G
481/161                                             511/171
TCT GGG TTC ACT TCC TTA TTG ATG GAG CGT CTC TCC GTT GAC TAC GGC
S   G   F   T   S   L   L   M   E   R   L   S   V   D   Y   G
```

CTP 5-1

```
1/1                                                 31/11
CCT GAT CAC CGC CAC ACT CAC AAT TTC TTT CTG CAA CAC CAG TTT CAA TCT TGT TGT GAT
P   D   H   R   H   T   H   N   F   F   L   Q   H   Q   F   Q   S   C   C   D
61/21                                               91/31
TTA TCT GAA AAA GTT GAA AAT ACC GTG TGC CAT TCG TTG ATA AAG AAA TTG TGA GTT GTT
L   S   E   K   V   E   N   T   V   C   H   S   L   I   K   K   L   *   V   V
121/41                                              151/51
ATC CTT AGA CAA AGC GGT GCG CTG CAC TCG AGG TTT CTT AGA ATT AAA TGT TCT GTA GGC
I   L   R   Q   S   G   A   L   H   S   R   F   L   R   I   K   C   S   V   G
181/61                                              211/71
ACG AAG CAA TAA TTG GAT TTA TCA TTC GCT TTG TAG TTC AAA ATG ATT ATT TGA ATC GAC
T   K   Q   *   L   D   L   S   F   A   L   *   F   K   M   I   I   *   I   D
241/81                                              271/91
TTA TTT AAC AAA TTA CCA CTG CTT GTG ATC GAC TAA CGC ACG AAA TTT AGT TAC ACG TTG
L   F   N   K   L   P   L   L   V   I   D   *   R   T   K   F   S   Y   T   L
301/101                                             331/111
TAG ATA GAT TTT TTT GTG ACG CCT GAT TAG TTC ATA CAC AAG AAT TTC GTA TCC CTT GCT
*   I   D   F   F   V   T   P   D   *   F   I   H   K   N   F   V   S   L   A
361/121
AAG ATA TAA CGA CGA CAA TAC CGC TTA ACA
K   I   *   R   R   Q   Y   R   L   T
```

ORF and 3' UTR for CTP 15-1

```
1/1                                                 31/11
361/121                                             391/131
                                                                  GSP : 1584
AAA GAG CAT GAA GAC AAC AAT CTC AGA AAA AGA CGA TCT ACA GAA ACA CCA GTC GAT AGC
K   E   H   E   D   N   N   L   R   K   R   R   S   T   E   T   P   V   D   S
421/141                                             451/151
AGT CAG TTG ATT ATG AGG CTT CTG AAA CAC ATA AAG GCC ACT AAC GAA TTT CAA AAC ATA
S   Q   L   I   M   R   L   L   K   H   I   K   A   T   N   E   F   Q   N   I
481/161                                             511/171
                        GSP : 1583
GCT ATT GAC AAA ATG ATG ACC GCA CAA CAA ATC GCC GAT AAA TTT GGA ATC GAA TTT AAT
A   I   D   K   M   M   T   A   Q   Q   I   A   D   K   F   G   I   E   F   N
541/181                                             571/191
CCT GAC CCA GAA ATC CTG ACT GAT TTA GCC GTA GCC ACT AAT AAG AAC ACA AAT GAT CTT
P   D   P   E   I   L   T   D   L   A   V   A   T   N   K   N   T   N   D   L
601/201                                             631/211
GCA TCT ATG CTG AGT GAA GTT TGT GAT CTG AGC AAA AAT AAA ACA TGT CAT AAC ATA AGT
A   S   M   L   S   E   V   C   D   L   S   K   N   K   T   C   H   N   I   S
661/221                                             691/231
TTG GAT CAG GAA GAC GAA GAC GAA ATA GAA ATG CCA ATT GAC AAT AGC ACT TAC TAT GTT
L   D   Q   E   D   E   D   E   I   E   M   P   I   D   N   S   T   Y   Y   V
721/241                                             751/251
TAC GCG CTT TAC TAT CCC GAT GAT GAT GAT GAC TAT ACC CAG CCG CCT TAC GAA TTT GTG
Y   A   L   Y   Y   P   D   D   D   D   D   Y   T   Q   P   P   Y   E   F   V
781/261                                             811/271
CCA CAG ACC TCT AGT CCG ACG GCT CCC AAC AAC TAT CCG TAC GAA ACA CAG ATC GAT CGC
P   Q   T   S   S   P   T   A   P   N   N   Y   P   Y   E   T   Q   I   D   R
841/281                                             871/291
CGC TTA GTG AAC GTG AAC ACT ATA AAG CAC CAG TTA GTC AGA AGC CAT CAT TTT ACG AGG
R   L   V   N   V   N   T   I   K   H   Q   L   V   R   S   H   H   F   T   R
901/301                                             931/311
CGA TTT CGA TGC TAC GAG AAT GTT CAA TGG AGC CAT TTT CTG ATC CAG AGC CAG AAT TGG
R   F   R   C   Y   E   N   V   Q   W   S   H   F   L   I   Q   S   Q   N   W
961/321                                             991/331
TAG GCG AAG TGT ATG AGG AAT CAG TTT CTA ACA AAG TAA TTG TTG AGA AAG GCG ACG AAC

1021/341                                            1051/351
CCG GAG CGG CTA CAG TCA ATC ACA TTA TGA CGT ATT CTG TAT CAG AAA AAT CAC ACT ACA

1081/361                                            1111/371
AGA CGC CAC AAA TCG AAA GAC TAC CAC AAC AGA TGC AGT ACT ATT TTT ATT TAA TCT AAC

1141/381                                            1171/391
TGT TTG TTT TTT AGA ATA AGG ACT TGT ACA TAT AAC TCT AAC ATA AAA TGT ATG ATT AAG

1201/401                                            1231/411
TTA AAA TGT TGA AAT TGC TGT CTA TGA AAA TTT GAG TTT TAC CGA ATC ACT CAC GAT ATA

1261/421                                            1291/431
```

TTC CTT AAA AAG AAA CTT AGA ACA TTT CTT TGG TTT TAT AAA AAC ATT CAT TTA TTT
ATA

1321/441 1351/451
TTT TCA TAT TCA CTA ATA AAA TTC ACT GCA AAA CAC TTC TAA CAA TAT AAC TTG TAT TTT

1381/461 1411/471
TGA AAA ACA TGA TGA ATA TAG GAA AGA TAT TAT ATT TAT GTA AAT AAA AGT GCT ATT TCG

1441/481
TTT ACT

Appendix II: Homologies and Search Results

Matches for the closest homologous sequences for clones NTP 1-2, NTP 8-11, and CTP 1-1 are given. Also represented is the small region for which NTP 11-3 finds partial homology. Blast search results for the remaining clones, NTP 1-4 and CTP 15-1, are also included.

NTP 1-2

dbj|BAA89259.1| (AB024901) ribosomal protein L3 [Bombyx mori]
Length = 171
Score = 236 bits (596), Expect = 7e-62
Identities = 111/111 (100%), Positives = 111/111 (100%)
Frame = -1

HRKFSAPRHGSMGFYPKKRSRRHRGKVKAFPKDDPSKPVHLTAFIGYKAGMTHVVREPDR 186
HRKFSAPRHGSMGFYPKKRSRRHRGKVKAFPKDDPSKPVHLTAFIGYKAGMTHVVREPDR
HRKFSAPRHGSMGFYPKKRSRRHRGKVKAFPKDDPSKPVHLTAFIGYKAGMTHVVREPDR 62

PGSKINKKEIVEAVTIIETPPMVCVGVVGYIETPHGLRALLTVWAEHMSED 339
PGSKINKKEIVEAVTIIETPPMVCVGVVGYIETPHGLRALLTVWAEHMSED
PGSKINKKEIVEAVTIIETPPMVCVGVVGYIETPHGLRALLTVWAEHMSED 113

NTP 8-11

emb|CAA77160.1| (Y18419) t-complex polypeptide 20 [Drosophila virilis]
Length = 532
Score = 250 bits (633), Expect = 7e-66
Identities = 125/161 (77%), Positives = 144/161 (88%), Gaps = 2/161 (1%)
Frame = -1

SVTIL-KGPNKHTLTQIKDAVRDGLRAINNAIEDKCVIPGAAAFEIKANTELLKYKDTVK 177
SVTIL KGPNKHT-TQIKDA-RDGLRAINNI DK+-PGA AFE+-A-ELKYK- TVK
SVTILIKGPNKHTITQIKDAIRDGLRAINNTIGDKALVPGAGAFEVRAHNELTKYKETVK 431

GKARLGIQAYAEALLVIPKTLAVNSGYDAQDTIVKLQEESRLNPEPIGLDLSTGEAFKPT 357
GKARL -QA-A-ALLVIPKTLA-NSGYDAQDTIVKL ERL+PE -GLDL+TGE KP
GKARLAVQAFADALLVIPKTLAINSGYDAQDTIVKLTVEDRLSPELVGLDLATGEPMKPA 491

DLGILDNYIVKKQILNSCSVIASN-LLVDEIMRAGMSSLKG 477
DLG- DNYIVKKQILNSCS-IASN LLVDE-MRAGM-SLKG
DLGVYDNYIVKKQILNSCSIIASNLLLVDEVMRAGMTSLKG 532

CTP 1-1

emb|CAA58465.1| (X83429) alpha-tubulin [Bombyx mori]

Score = 298 bits (764), Expect = 3e-80
Identities = 145/163 (88%), Positives = 148/163 (89%)
Frame = -1

KCVSASLVHVGQAGVQIGNACWELYCLEHGIQPDGQMPTDKTIGGGDDSFNTFFSETGAG
-C-S VHVGQAGVQIGNACWELYCLEHGIQPDGQMPTDKTIGGGDDSFNTFFSETGAG
ECIS---VHVGQAGVQIGNACWELYCLEHGIQPDGQMPTDKTIGGGDDSFNTFFSETGAG

KHVPRAVFVDLEPTVVDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVDLVL
KHVPRA-FVDLEPTVVDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVDLVL
KHVPRALFVDLEPTVVDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVDLVL

DRIRKLADQCTGLQGFLIFHXXXXXXXXXXXXLLMERLSVDYG
DRIRKLADQCTGLQGFLIFH LLMERLSVDYG
DRIRKLADQCTGLQGFLIFHSFGGGTGSGFTSLLMERLSVDYG

NTP 11-3

gb|AAF46409.1| (AE003444) CG1785 gene product [Drosophila melanogaster]
Length = 478

Score = 50.4 bits (118), Expect = 1e-05
Identities = 27/70 (38%), Positives = 38/70 (53%), Gaps = 10/70 (14%)
Frame = -1

LIKKEQHLDRVTTKMFKKVTTAEKENQWREEMSEGLPKPH----------NPAEHDSESD 153
-IKKEQHL RVTTMF KVT E - -EMS-G+ - N + + + +
IIKKEQHLKRVTTSMFSKVTPEERDRRRLNEMSQGMDEEEGSELDEDVQTNGKKKEDDDE 299

GEYKAINPPV 183
IN PV
KPYHTINAPV 309

Closest matching sequence for CTP 15-1

>gb|AAC72981.1| (AF101438) Polycomb homolog Pc1 [Xenopus laevis]
Length = 471

```
 Score = 32.1 bits (71), Expect = 9.9
 Identities = 23/63 (36%), Positives = 32/63 (50%), Gaps = 4/63 (6%)

Query: 51  EGKQDWLDFFLGPNSDLYPDHRPT-SSVAMMNQATQSDKPIEEQLEDI---KIAANKITQ
           EG  D  D   G  S L  DH +  SS M   QA+Q  KP    LE+   +   AN IT
Sbjct: 395 EGQRDVADLSTGDDSSLDSDHDSSLSSQDMAVQASQDWKPARSLLEHVFVTDVTANLITV
454

Query: 107 AIQ 109
           ++
Sbjct: 455 TVK 457
```

NTP 1-4 BLAST

Sequences producing significant alignments:
(bits) Value

sp|Q05854|YL78_YEAST	PUTATIVE 151.3 KD TRANSCRIPTIONAL REG...	35
sp|P27625|RPC1_PLAFA	DNA-DIRECTED RNA POLYMERASE III LARGES...	35
sp|P33748|MSN2_YEAST	ZINC FINGER PROTEIN MSN2 (MULTICOPY S...	34
sp|Q03164|HRX_HUMAN	ZINC FINGER PROTEIN HRX (ALL-1) (TRITHO...	31
sp|P79145|CREM_CANFA	CAMP-RESPONSIVE ELEMENT MODULATOR	30
sp|P52963|NBL4_MOUSE	BAND 4.1-LIKE PROTEIN 4 (NBL4 PROTEIN)	30
sp|Q02308|HLES_DROME	HAIRLESS PROTEIN	29
sp|Q91661|GLI4_XENLA	ZINC FINGER PROTEIN GLI4 (NEURAL SPECI...	29

Alignments

>sp|Q05854|YL78_YEAST PUTATIVE 151.3 KD TRANSCRIPTIONAL REGULATORY PROTEIN

Score = 35.2 bits (79), Expect = 0.12
Identities = 25/95 (26%), Positives = 45/95 (47%)
Frame = +1

PTENDNVTKVSVQMNNDKNEGLMKSADIEHMPKINEPAQHLSGTNDNIKKAAMPNSKNNE
P+ + K++V NND N + S + EH N + + N+N + N+ NN
PSSQEKNLKINV--NNDNNGMTISSVNREHNNNHNNNNDNNNNNNNNNNNSNNNNNVNNN

MSTAQSKFEESNNFRIANHNSQSVGHHHTSTMKDV 303
+ + S+ +N+ N NSQ V +++ + DV
DNESNSRSTTNNSCNNGN-NSQYVRNNNVTMENDV 981

>sp|P33748|MSN2_YEAST ZINC FINGER PROTEIN MSN2 (MULTICOPY SUPPRESSOR OF SNF1 PROTEIN 2)

Score = 34.0 bits (76), Expect = 0.26
Identities = 26/96 (27%), Positives = 45/96 (46%)
Frame = +1

NVTKVSVQMNNDKNEGLMKSADIEHMPKINEPAQHLSGTNDNIKKAAMPNSKNNEMSTAQ
N T +S Q +N NE L+ +H + + GN N+ P N+N ++
NETNLSPQTSNG-NETLISPRAQQHTSIKDNRLSLPNGANSNLFIDTNPNNLNEKLRNQL

SKFEESNNFRIANHNSQSVGHHHTSTMKDVNHLDTM 321
+ S + I+N NS S G+++ S ++ D+M
NSDTNSYSNSISNSNSNSTGNLNSSYFNSLN-IDSM 262

```
>sp|P23250|RPI1_YEAST  NEGATIVE RAS PROTEIN REGULATOR PROTEIN
          Length = 407

 Score = 32.1 bits (71), Expect = 1.00
 Identities = 18/59 (30%), Positives = 25/59 (41%)
 Frame = -1

SGTNDNIKKAAMPNSKNNEMSTAQSKFEESNNFRIANHNSQSVGHHHTSTMKDVNHLDT
+GTNDNI      N  NN  +   S  +  NN    N NS    +H T+      H+ +
NGTNDNINNHYYNNCNNNNNNINNSNNSNNNNSNNINRNS----NHSTNVFSTPEHIQS

>sp|Q03164|HRX_HUMAN ZINC FINGER PROTEIN HRX (ALL-1) (TRITHORAX-LIKE
PROTEIN)

 Score = 30.9 bits (68), Expect = 2.2
 Identities = 29/134 (21%), Positives = 57/134 (41%), Gaps = 4/134 (2%)
 Frame = -1

TENDNVTKVSVQMNNDKNEGLMKSADIEHMPKINEPAQHLSGTNDNIKKAAMPNSKNNEM
TE+D     + + ++   +   K    E++ KI+ P    +G  +++ K++++  +M
TESDTSVTATTRKSSQIPKRNGKENGTENL-KIDRPED--AGEKEHVTKSSVGHKNEPKM

STAQS----KFEESNNFRIANHNSQSVGHHHTSTMKDVNHLDTMRTHMRMDPVHKNLNDH
    S    K +  ++      + +S    HTST  D N LDT  T   +     N +D
DNCHSVSRVKTQGQDSLEAQLSSLESSRRVHTSTPSDKNLLDTYNTELLKSDSDNNNSDD

LIHGQVYHDDHIDY
   G +   D +D+
C--GNILPSDIMDF

>sp|Q02308|HLES_DROME HAIRLESS PROTEIN

 Score = 29.3 bits (64), Expect = 6.6
 Identities = 15/58 (25%), Positives = 29/58 (49%), Gaps = 2/58 (3%)
 Frame = -1

EHMPKINEPAQHLSGTNDNIKKAAMPNSKNNEMSTAQSKFEE--SNNFRIANHNSQSV
EH    IN  + HS    N+N   +  N+ N++ +++ +      SN   + N+N+ S+
EHKSNINSNSSHSSNNNNNGSSSNNDNNSNDDAASSSNSKNNNTSNESSHSNNNTSSI


>sp|Q91661|GLI4_XENLA ZINC FINGER PROTEIN GLI4 (NEURAL SPECIFIC DNA
BINDING PROTEIN XGLI4) (XGLI-4)

 Score = 29.0 bits (63), Expect = 8.7
 Identities = 19/75 (25%), Positives = 33/75 (43%), Gaps = 5/75 (6%)
 Frame = -1

NDNVTKVSVQMNNDKNEGLMKSADIEHMPKINEPAQH-----LSGTNDNIKKAAMPNSKN
ND + K S+   N  NE   K +  EH   ++   +H       T DN+  +  P  ++
NDIIQKPSLPKENGDNEASAKLSGREHSDSVSRDQEHCLQTRTIKTEDNMMHQSSPGGQS

NEMSTAQSKFEESNN
+  S+  S +  +NN
S-CSSEPSPYGNTNN
```