IMPROVEMENT OF EXPRESSION OF RECOMBINANT HUMAN PROTEIN C IN THE MILK OF TRANSGENIC ANIMALS USING A NOVEL TRANSGENE CONSTRUCT

by

Christopher Glyn Russell

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical Engineering

APPROVED:

W. H. Velander, Chairman

J. L. Johnson

R. M. Davis

E. A. Wong

W. N. Drohan

December, 1993

Blacksburg, Virginia

IMPROVEMENT OF EXPRESSION OF RECOMBINANT HUMAN PROTEIN C IN THE MILK OF TRANSGENIC ANIMALS USING A NOVEL TRANSGENE CONSTRUCT

by

Christopher Glyn Russell

Committee Chairman: William H. Velander
Chemical Engineering

(ABSTRACT)

Past studies of mammary tissue specific expression of transgenes using the murine whey acidic protein (WAP) promoter have shown widely variable, position-dependent and copy number-dependent expression. This study evaluates a series of three WAP transgenes containing the cDNA of human protein C (hPC) for the expression of human protein C in the milk of mice. In two of the transgenes studied, the cDNA of (hPC) was inserted at the translational start site of a 7.8 kbp mouse WAP genomic DNA Eco RI fragment containing 2.6 kbp of 5' flanking, 3.9 kbp WAP coding (exons and introns), and 1.3 kbp 3' untranslated region (UTR) and flanking sequences (designated WAPPC1 and WAPPC2). A third transgene consisted of only the 2.6 kbp of WAP 5' UTR and flanking DNA, 1.4 kbp hPC cDNA, and 1.3 kbp of 3' WAP UTR and flanking DNA with no linker sequences (designated WAPPC3). The WAPPC1 and WAPPC2 transgenes expressed up to about 10 μg/ml recombinant hPC in mouse milk while WAPPC3 expressed 30-300 ($n=10$, $n=5$, $n=11$, number of founder lines evaluated for each transgene, respectively). In

contrast to past studies with WAP-cDNA fusion transgenes where the maximal expression was about 5% of endogenous WAP expression, the WAPPC3 transgene gave maximal expression which was about 30% of endogenous WAP expression. Thus, results from the combination in WAPPC3 of intact 5' and 3' WAP UTR with the cDNA of hPC suggests that introns are not necessary to enable high level expression in the mammary gland when using WAP regulatory elements. Relative specific transcript and protein levels in the transgenic animals studied suggest that the rates of translation initiation may be different for the mRNAs of each of the transgenes studied.

ACKNOWLEDGEMENTS

At the risk of overlooking persons who have helped me in the pursuit of my Ph.D., I wish to specifically thank the many who immediately come to mind. I am indebted to my committee members for their excellent advice and criticism, and especially to John Johnson and Bill Velander for their close supervision of my work. I also thank Bill for bringing me to Virginia Tech and for his friendship and many good basketball and softball games. Also, I especially appreciate the many conversations and patient instruction of John Johnson and his ability to make everything bad seem "not so bad".

I appreciate the assistance and abilities of all the personnel who have helped me with aspects of my work such as pig and mouse handling and protein analyses, including Barry Williams, Gene Ball, Yopo Canseco, Ray Page, Steve Butler, Amy Larsen, Diane Moody, Steve Ellis, Jan Lowe, Tulin Morcol, Anu Subramanian, Art Degener, Tim Moore, Krista Hess, and Lillian Crow. I thank Julianna Toth and Georgia Hammond for advice and help in the lab and for their frienship.

I especially thank my family for their love and for supporting me in my graduate work. I thank my wife Tracy for being the best in the world and especially for the last year-and-a-half since she has been here with me. Tracy, thank you for the walks and the talks and for being my best friend.

Lastly, I wish to thank and praise God for the wonderful world He has created. Science at its best is our attempt to understand the Creation and thus to better understand the Creator (Proverbs 3:5-6).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES, continued

INTRODUCTION

Transgenic technology has revolutionized the study of eukaryotic gene regulation and gene expression. Because natural genes, heterologous genes, or fusion genes (originating from multiple genes or multiple species) can be synthesized in vitro and then inserted into the genomic DNA of an animal by transgenic technology, the regulation of virtually any gene can be studied in mammals. The production of foreign proteins for pharmaceutical use or for determining protein function are important areas of transgenic animal research. In both of these fields of study, greater insight into the mechanisms of gene expression is being rapidly made.

The gene is the most fundamental storage unit of information for the temporal regulation and structural features necessary to synthesize proteins. Proteins act as regulators and catalysts for the biochemical reactions of life. Many of the biochemical reactions of life are well-conserved throughout the processes of catabolic and anabolic metabolism which define and sustain the structure of any given organism. As a result, most proteins have taxonomically conserved regions of biochemical structure and function. Thus, many genes also have well-conserved regions, structures, and functions even across species barriers. The gene transmits this information by acting as an asymmetric, macromolecular substrate which consists of a sequence of polydeoxyribonucleotides (deoxyribonucleic acid, or DNA) with distinct 5' and 3' phosphorylated ribose moieties at the ends. The molecular

asymmetry of DNA consists of a four letter alphabet of nucleotides (nt) containing adenine (A), thymine (T), cytosine (C), and guanine (G), which make up the units of information for inheritance of genetic information. DNA is replicated with fidelity to allow transfer of this genetic information to daughter cells or to offspring.

To initiate the mechanism of gene expression, DNA must be sequentially processed or "transcribed" into polyribonucleotides (ribonucleic acid or RNA) with fidelity [with regard to sequence, except that uracil (U) is always substituted for thymine (t)] in a 5' to 3' direction. When complexed together with a specific system of proteins (termed a transcription complex) the DNA acts as a template which transmits, by biochemical reaction, the regulatory and structural information of the gene into a series of gene products. Genes can be considered catalysts because they are not used up and are regenerated to their original form during the biochemical reaction. The transmittal process is a series of reactions where genes (within the transcription complex) generate RNA and proteins which in turn catalyze other reactions of life as well as regulate the gene itself. This circuitous process is referred to as gene regulation and expression. Because of the biochemical diversity of organisms, the structure and function of many genes are species- and tissue- specific, but are surprisingly well-conserved within classes of prokaryotes and eukaryotes. The work presented here strives to elucidate a better understanding of this conservation as applied to mammary tissue expression of a foreign protein in the milk of mice and pigs. The process of gene regulation and expression is viewed here in the context of

a system of rate processes defined by each respective biochemical reaction or reaction step.

Vast numbers of genes are concatenated and grouped into very high molecular weight structures called chromosomes. The highly dense nature of the DNA structure within the cell is referred to as chromatin DNA. While most elements of a naturally functioning gene are spatially juxtaposed and are termed "cis" elements, some elements may be at distant locations within the same chromosome. Other factors required for gene function are produced at distant sites (on other chromosomes within the cell or even outside the particular cell or tissue in which they may act) and are termed "trans"-acting factors. For the purposes of this work, the gene is defined as the smallest functional (transcription) unit of DNA which can transmit information to either the RNA or protein level. For a functional gene in higher eukaryotes, the known steps **required** for expression of a protein are: 1) Transcription of the DNA into a nascent RNA; 2) Capping of the mRNA at the 5' end by the (unique) addition of guanosine triphosphate (GTP) by a 5'-5' linkage and subsequent methylation of N-7 of the GTP cap; 3) 3' end processing or modification; 4) Removal of introns by splicing; 5) Export of the mature mRNA to the cytoplasm; and 6) Translation of the mRNA into a polypeptide. The exact order of steps 2-5 has not been established but no non-viral uncapped mRNAs have been identified in eukaryotes (1).

The general structure of a functional eukaryotic gene is given in Figure 1. A variety of DNA sequences or "elements" are required for a fully functioning gene.

Each of these elements acts as a substrate recognition site or binding site for regulatory proteins or enzymes involved with the regulation, expression, or maintenance of the gene or gene product. The most notable features of eukaryotic genes listed in Figure 1 include exons, introns, promoter sequences, RNA termination signal(s), RNA cleavage and polyadenylation signal(s), translational start and stop signals, and possible hormonal response and enhancer elements.

Transcription is an enzymatically mediated reaction by which a proximal sequence of gene regulatory and protein encoding elements are processed to generate a sequence of RNA (Figure 2). Unlike DNA, RNA is subject to extensive chemical transformation and degradation as part of the gene expression process so that it behaves kinetically not as a catalyst, but as a true chemical substrate. This nascent RNA product contains less information than the original gene since only elements for RNA regulation and protein code are needed for subsequent reactions. Thus, from a chemical engineering viewpoint, transcription (especially initiation of transcription) represents a distinct reaction step (throttling point) where process control may be logically exerted. Specifically, the DNA gene is "transcribed" into an RNA transcript by a macromolecular complex that includes an RNA Polymerase (RNA Pol II for nuclear genes) and several proteinaceous transcription factors. The transcription of DNA into RNA is catalyzed by a coordinated reaction between DNA promoter elements and transcription factors that recognize these promoter sequences. "Consensus" promoter DNA sequences in eukaryotic genes include the "TATA" box

and the "CAAT" box, which are located at about -25 and -75 nt (upstream), respectively, relative to the transcriptional start site. The initiation of transcription may also be amplified or induced by the presence of so-called "enhancer" elements and/or "Hormone Response Elements" (HREs). The structure of an enhancer is believed to create a domain which lowers the activation energy for entry of the transcription complex into specific areas of chromatin DNA structure. Thus, enhancer structures exert a kinetic control over transcriptional gene units. The presence of HREs allows another level of coordinated induction or attenuation of transcription in response to a sequence of DNA binding proteins which are the result of hormonal stimuli. For example, the process of lactogenesis consists of an endocrine controlled secretion of insulin, prolactin, and glucocorticosteroids which together form an extracellular signal which is transmitted into the cells of mammary tissue through a mechanism which results in DNA-binding proteins (hormone plus hormone receptor) complexed to HRE sites. The many gene products which coordinately arise from the signals provided by extracellular lactogenic hormones result in transformation of the structure of the mammary tissue and the subsequent production of specific proteins which are secreted in the milk.

The transcription process is concluded by a putative "terminator" element which has been postulated to cause RNA Polymerase to disengage the DNA substrate once the transcription enzyme complex reaches the terminator element. While this DNA sequence is not well-characterized, it may be dispensable due to specific RNA

cleavage sites which generate precise and well-characterized 3' ends of transcripts (Figure 3). The oligonucleotide sequence "AATAAA" is common in eukaryotic genes and is a "consensus signal" for cleavage and modification at the 3' end of RNA transcripts. Although the "AATAAA" is a consensus signal, it is apparently not sufficient for cleavage and polyadenylation. Other recognition signals are believed to be contained in the sequences just downstream of the AATAAA consensus sequence, although these other sequences are less well-characterized. After the primary or nascent RNA transcript is cleaved at the 3' end, a polyadenylate tail is added to the RNA (except for certain classes of RNAs, such as histones, that do not have polyA tails). This polyA tail may be 200 to 300 nt long, but is often heterogeneous in size. The polyA tail, in conjunction with a polyA binding protein, probably protects the RNA from degradation at the 3' end. The 5' end of the transcript is also enzymatically modified to have a "cap" structure. This cap structure is a GTP residue attached to the 5' nucleotide of the transcript by a 5' to 5' linkage and subsequently methylated. It is believed that the cap protects the RNA from degradation at the 5' end, and is also probably required for recognition of the mRNA transcript by the ribosome complex.

The DNA and nascent RNA transcripts have protein encoding regions which are interrupted by intervening sequences which do not code for protein. Exons are the protein-encoding sequences of the DNA and are present in the final, mature version of the RNA transcript, which is termed the messenger RNA, or mRNA.

Introns, also referred to as intervening sequences, are thought to act as molecular spacers between the exons in the gene and in the pre-mRNA. Introns may have a function of helping to control gene expression at the transcriptional (and possibly the post-transcriptional) level. Since the introns must be removed to form a functional protein, the removal of introns and joining of the exons (or "splicing") can logically serve as a distinct kinetic step from which a regulatory effect is exerted at the post-transcriptional level. After the primary or nascent RNA transcript is formed and modified at the 5' and 3' ends, the introns are removed by a complex known as the "spliceosome" (Figure 4). The biochemical machinery of the spliceosome is intricate and requires several proteins and ribonucleoprotein factors to remove the introns and join the exons which code for the protein.

When the mature mRNA has been formed by the above reactions, it must then be transported across the nuclear membrane to the cytoplasm (Figure 4), where it may be translated into protein. That is, within the exons are contained the three-nucleotide "codons" which are recognized by the ribosome complex and are then "translated" into individual amino acids joined by peptide bonds to form specific polypeptides and proteins (Figure 5). The act of translation can also be regulated at the levels of ribosome (complex) binding, translation initiation, or possibly even polypeptide elongation (see ref. 2 for a detailed discussion of gene regulation and expression).

In summary, each of the individual steps mentioned above is a point at which regulation or control of the gene may be performed (Figure 6). This regulation can

be exerted by induction, attenuation, or complete inhibition of gene activity. The slowest step in the chain of reactions would be the rate-limiting step for the formation of the gene product, just as in ordinary multiple-step chemical reactions. Besides the regulatory step described above, the nature of the gene expression reaction pathway is further complicated by degradation of intermediate reaction products. Expression of a native gene or transcription unit is ordinarily done in a coordinated combination of induction, attenuation, inhibition, and degradation pathways which are determined by the molecular interaction of distinct elements of the gene sequence and the species- or tissue-specific recognition of these elements by regulatory factors. Because some genetic elements serve to attenuate or enhance the reactions of gene processing, their presence or absence will give differential gene expression. If particular regulatory step(s) at which the expression of a transgene is rate-limited can be determined, appropriate deletions or additions of genetic elements may be made to benefit regulation. In this pursuit, the goal may be higher (or lower) levels of expression from the transgene or better developmental control of the transgene.

LITERATURE REVIEW

Transgenic technology has revolutionized the study of eukaryotic gene regulation and gene expression. In addition to the ability to use transgenic animals for the production of foreign proteins or for the study of heterologous protein function, the transgenic animal allows the in vivo study of heterologous gene regulation and expression. The capacity to study the regulation of virtually any gene in a higher animal lends more insight into the mechanisms of the expression of that gene, whether it is a natural gene, a heterologous gene, or a fusion gene originating from multiple genes or multiple species.

This ability to insert a gene into the genomic DNA of an animal allows the investigator to study, in detail, the many different cis-acting factors of genes which give rise to gene expression. The transgenic animal is inherently facile to studies of mutant genes, such as gene deletions and insertions, which had only previously been possible in cell culture systems, either prokaryotic or eukaryotic.

The study of gene expression in transgenic animals is not without limitation. For example, random integration of the transgene results from pronuclear microinjection -- the most common method for generating animals. This method is believed to result in a random number of transgene copies integrated at a random site within the chromosomal DNA (3). However, it is possible to generate transgenic animals with transgene recombination occurring at a specific locus after pronuclear

microinjection (13); however, this methodology is very inefficient at generating homologous site recombination (0.2% of born mice) and one must be able to distinguish between random integration (which occurs in 20-30% of mice born from pronuclear microinjection) and the homologous integration. Because of the randomness of gene integration, a particular transgene may not express equally in different transgenic animals made with that transgene, due to the effect of the chromatin structure (the surrounding chromosomal domain) around the transgene. Therefore, a transgene integrating in a "good" site will more likely result in high expression levels and vice-versa. However, what gives rise to "good" or "bad" integration sites is largely unknown, and there is still the limitation of randomness of integration.

One method of overcoming this randomness is the technique of homologous recombination of a transgene into an embryo (4, 5). However, at present, this technique is limited to systems for which an embryonic stem cell (ES) line has been developed. At this time, mice are the only animals for which an ES line is available, though many groups are working on ES lines for large animals such as pigs and cattle.

Another approach to overcoming the randomness of transgene integration is to incorporate enough "extragenic" elements into the transgene to give the transgene domain sufficient structure in the chromatin for proper gene expression. Unfortunately, this approach is limited to the known elements, such as enhancers (6),

matrix attachment regions (MARs, 7), and locus control regions (LCRs, 8). Another limitation with incorporating many elements as above is that the transgenes can be excessively large, resulting in extra difficulty cloning, handling, and injecting. This limitation must be taken into consideration in light of the extra time and effort required, even though it can be overcome by alternate cloning methods (e.g. Bacteriophage P1, 9; YACs, 10; cosmids, 11 and 12) and injection methods which result in extrachromosomal homologous recombination of injected transgene fragments (14).

The ability to make and use transgenic animals spurred an immediate interest in the field of protein production, and the subsequent rush to the end product, the protein, often led to a shotgun approach to transgene design. That is, researchers often used a "cut-paste-and-test" methodology for designing transgene constructs with simplified regulatory formats. Further, the choice of the "best" transgene was often rooted only in the construct that yielded the highest expression levels. It has been found by many groups (15-18) that a transgene with "genomic" sequences (that is, exons and introns) will usually express higher levels of protein than will a transgene with the cDNA sequence for the same gene. Further, it was found that in some cases heterologous introns can increase the expression from transgenes (19). These observations have led many groups to choose a genomic transgene over the cDNA transgene whenever possible. However, high expression levels from cDNAs have been obtained from cDNAs (20-22), though often with greater variability in

expression rates and expression levels (23, 24).

The relationship between genetic regulatory elements and their corresponding tissue specificity can complicate targeted expression of a foreign protein. Tissue specific recognition may be inefficient because transgenes are often chimeric, with structures having promoter, coding region, and 3' elements with different origins. In many cases, the "host" transgenic animal may not normally contain any of the native elements of these genes. Thus, some cis- or trans- acting regulatory elements for the transgene may not be present in the host animal. In light of this information, it is easy to see that predicting the behavior of a transgene can be very difficult, if not impossible.

As a result of the above methodology, the design of transgenes has often been based solely on the end product and little regard was given to transgene design from the standpoint of natural gene expression and the regulation of the transgene. Though this approach accomplished the goal of protein production in many cases, the intermediate steps of gene regulation were often overlooked and left unoptimized. This approach has led to a largely incomplete understanding of transgene regulation and expression.

Gene Regulation and Expression

Study of the molecular biology of organisms has yielded an invaluable storehouse of information with regard to "proper", natural gene expression. This

information on endogenous gene expression should serve as a guide to transgene design. In other words, past studies have shown that natural gene structures are organized in a manner that enables tightly controlled regulation in terms of induction and attenuation, which results in correct temporal and tissue-specific expression of gene products.

It is often accepted that the main level of gene control is at the initiation of transcription (25). However, it is also apparent from the literature that transcription initiation is only a small part of the overall picture of gene regulation. As early as 1973, measurements of mRNA half-lives in cultured HeLa cells indicated that the production of mRNA is probably not the rate-limiting or regulating factor (26). In fact, Singer and Penman determined that about 67% of steady state mRNA had a long half-life of about 24 hours average. The remaining 33% had a relatively short half-life of about 7 hours. On the basis of these findings, Singer and Penman hypothesized that the main regulation was likely at the level of initiation of translation. Berger and Cooper (27) reported similar results for mRNAs from resting human lymphocytes. However, the half-lives they reported were about 17 min for the short-lived component and greater than 24 hours for the stable component. Berger and Cooper theorized that protein synthesis might be regulated by controlling the supply of specific messages within the pool of mRNA. Though these theories are not entirely false, it is apparent that several levels of gene regulation may exist, including transcriptional, post-transcriptional, translational, and post-translational. These ideas

are discussed below.

Transcription

The rate of transcription of a particular gene is controlled primarily by the rate of initiation of transcription, which is determined by promoter "strength" in concert with enhancers, LCRs, MARs, or other elements that affect the chromatin structure around that gene. Presence of response elements (e.g. glucocorticoid response elements, GRE) in the promoter allow another type of transcriptional regulation in response to hormonal stimuli. Further, sequences in the 3' end of the gene have been implicated in tissue- and development- specific expression of genes. These regions have been characterized as DNAse I hypersensitive sites (6 and 28) and show the same effect and mode of action as enhancers that are located 5' to the gene, and thus can probably be characterized as enhancers, though the authors did not choose this designation.

The transcription of the gene into RNA requires several trans-acting factors, from the proteins ("transcription factors"; see ref. 29 for discussion) that recognize the enhancer(s), GREs (if present), and promoter sequence(s) to the enzyme that forms the RNA transcript (RNA Polymerase II). The action of cis- and trans- acting factors on the DNA sequence in the promoter region often directs the expression of a gene in a tissue-specific manner. Also, upstream and intragenic DNA sequences have been shown to direct tissue specificity for immunoglobulins heavy and κ light chain

genes (30) and immunoglobulin heavy chain μ (31), though in the latter case the sequences are undefined and the mode of action is unclear.

The first intron of the α1(I) collagen gene was implicated in transcriptional control (32). Specifically, the authors suggested that the promoter-intron interactions, mediated by DNA-binding proteins, regulate transcription of the collagen gene. The intronic element had both positively and negatively acting elements and the effect of the intron was dependent upon both position and orientation of the element.

Post-Transcriptional: 5' and 3' end processing

Processing of the 5' end of mRNAs must happen very rapidly after (or perhaps during) transcription. Although it has not been proven, it appears that the 5' cap (and possibly methylation of the cap) is crucial not only for translation but for mRNA stability and transport as well (1, 2, and 47). No uncapped eukaryotic mRNAs have been identified (1). Processing of the 3' end and polyadenylation (for most eukaryotic mRNAs) is believed to occur soon after the nascent transcript is formed (33) and is reliant on the consensus sequence "AATAAA", though this sequence is not sufficient for cleavage and polyadenylation. In yeast, it appears that the polyadenylation of mRNAs may be linked to transcription termination, but the 3' processing of non-polyadenylated histone mRNAs is not linked to termination. Also, it appears that the 5' ends of yeast mRNAs must be capped for efficient 3' processing. Further, the yeast polyA addition sites are located within a short range of nucleotides (34).

Termination of transcription in higher eukaryotes is not well understood, but may be unimportant in light of the nature of 3' end processing of mRNA, which appears to be critical. The steps of RNA cleavage and polyA addition may be a concerted reaction in higher eukaryotes and the processing requires specific sequences in the 3' untranslated region (UTR) of the transcript as well as several trans-acting enzymes to cleave the transcript and add the polyA tail, which is usually added at a precise site (for a review, see ref. 34), though the site can be altered by mutations in the mRNA downstream of the polyadenylation site (35). For those mRNAs with polyA tails, the literature most often implicates polyA in mRNA stability. This stability is probably a result of the interaction of the polyA with a specific polyA binding protein, which protects the 3' end from degradative attack (36).

Post-Transcriptional: Intron Removal

The spliceosome machinery is very complex and many proteins and ribonucleoproteins (RNPs) are involved in splicing of introns (for a review, see ref. 37). Even the polyA tail has been hypothesized to play a role in splicing of nascent transcripts (36) though this hypothesis has not been experimentally proven. Further, the spliceosome may play a role in retaining the pre-mRNA in the nucleus (38) or in transport of the mRNA to the cytoplasm (18, 37). Merrill et al (39) report that an intragenic region is sufficient to specify the regulation of thymidine kinase (tk) enzyme within the appropriate cell type and during the appropriate cell stage.

Further, they propose that the regulatory effect may be at the transcriptional level or post-transcriptional level. For the post-transcriptional model, they theorized that the regulatory effect might be due to either 1) a destabilization of the hnRNA in the improper cell type; or, 2) an inability of the improper cell type to correctly splice the tk transcripts. Thus, introns might play a role in the either transcriptional or post-transcriptional regulation of some genes.

After the above modifications occur and the transcript has become the mature mRNA, it must be transported to the cytoplasm to be translated into protein. The mechanism of this transport is not well understood (37, 38).

Post-Transcriptional: mRNA Stability

Once the mRNA is in the cytoplasm, regulation may occur by selective degradation (by specific RNAse activity) or by stabilization of the mRNA (by specific factors). The 5' cap is believed to offer stability by protecting the transcript from enzymatic attack at the 5' end. Protection at the 3' end is believed to arise from the polyA tail (in polyadenylated mRNAs) and associated proteins, or from the 3' UTR sequence and its secondary structure (and any associated proteins) for non-polyA mRNAs (e.g. for histones and actin). The degradative attack on the mRNA is then begun by removal of the protective polyA tail or 3' end by an endonuclease which likely is directed by other trans-acting factors to a specific region near the 3' terminus, just upstream of the polyA tail or 3' secondary structure. Following the

removal of the protective 3' end, the mRNA is then rapidly degraded 3' to 5', probably by an exonuclease that behaves in a processive fashion. Thus, the degradation is generally believed to occur in the 3' to 5' direction with the major elements leading to the instability determined by the sequence near the 3' terminus of the mRNA (25, 36, 40-44). Further, the 3' UTRs of mRNAs have been implicated in cell cycle or developmental regulation of histones (45) and dihydrofolate reductase (42).

Translational: 5' end and Regulation

Since translation is mechanistically and kinetically a separate class of biochemical reactions, it is another tier at which regulation of gene expression may occur. All eukaryotic mRNAs that have been examined have a 5' cap, and evidence supports the idea that the cap may be crucial for translation as well as for mRNA stability and transport (1). By comparison of all available sequences of vertebrate mRNAs at the 5' end, M. Kozak found highly conserved nucleotides immediately around the AUG start codon and postulated their importance in the initiation of translation. The context of these conserved regions indicates that they may be important in creating a structure that allows the ribosome complex to recognize the initiator codon, which is the first AUG in about 90% of known vertebrate mRNAs.

From review of this data and experimental data in the literature, Kozak inferred that certain classes of genes may be regulated by "throttling" at the level of

initiation of translation (46; reviewed in ref. 1; see also ref. 47). That is, the genes appear to be designed such that the mRNAs are not good substrates for translation initiation. This observation came from noticing aberrant patterns (see below) in the 5' UTRs of many mRNAs in the classes of oncoproteins, transcription factors and DNA-binding proteins, growth factors, ligand receptors, immune/inflammation mediators, and signal transducers. In contrast to the first AUG being the start (translation initiator) codon, as mentioned above, two-thirds of the oncogene mRNAs have AUG codons upstream of the start of the major open reading frame, indicating that a different type or context of translation initiation is in effect. Specifically, the nucleotide sequences around the upstream AUG "codons" do not match the consensus sequence and appear to be unfavorable for initiation of translation. Also, most of the mRNAs from genes with "upstream" (non-initiator) AUG codons are predicted to have highly structured 5' UTRs, which could also encumber translation. In fact, Kozak pointed out that these classes of genes are generally tightly regulated and not highly expressed; thus, the 5'-end "throttling" mechanism may be a general regulatory motif for these types of genes.

Post-Transcriptional and Translational: 3' UTR and Regulation

The 3' UTR can regulate genes by changes in message stability and/or the efficiency at which the message is translated. Zaret and Sherman (48) described a yeast gene in which mutations in the 3' end altered either the stability of the transcript

or the translational efficiency of the transcript. The stability of mRNA was quantified as the steady-state amount of specific mRNA relative to that of the normal endogenous gene. Translational efficiency was similarly assessed by comparison of the specific mRNA:protein ratio of revertants of a mutant system to that of the normal yeast system. The yeast systems characterized were revertants of a mutant that was deficient in iso-1-cytochrome c (CYC 1 gene). Reversion mutations were either spontaneous or induced by ultraviolet or x-ray irradiation.

For revertant strains containing the normal 3' end, the relationship between specific iso-1-cytochrome c mRNA and protein levels was linear even though the absolute levels were between 10 and 100% of normal. That is, the mRNA% : protein% ratio (expressed as percentages of normal endogenous levels) was 1:1 in 18 of 18 strains. However, for 5 of 7 strains with new 3' ends on the gene in question, the linear relationship was disrupted; i.e. the measured ratio of mRNA:protein was no longer 1:1. For example, four of these strains had relative mRNA:protein levels from 2:1 to 7:1. One outstanding example of the opposite relationship showed a protein level that was 100% of normal but an mRNA level that was only 64% of normal. Therefore, in the former cases the translational efficiency of the mRNA was lower than that of the normal gene (14 to 50% of normal), whereas in the latter case the translational efficiency of the mRNA was higher (156% of normal). Since the only changes in the gene in question were in the 3' end, we see that the 3' UTR can affect the translational efficiency of the mRNA in either a positive or a negative manner.

Summary of Gene Regulation Themes

So, it is apparent that there are many levels at which a gene may be regulated, and several of these may be in effect for any one gene. Further, the types of regulatory mechanisms appear to be so diverse as to preclude the concept of specific motifs for every type of eukaryotic gene regulation. Since it is clear that there are so many different types and levels of gene regulation, a thorough understanding of the types of regulation that may come into play is essential in the proper design of transgenes.

Milk Gene and Transgene Expression

One benefit of better understanding gene regulation would be to better use transgenic animals as bioreactors for production of foreign proteins. The mammary gland of a livestock animal, due to its ability to produce and secrete tremendous amounts of protein, is desirable for use as a bioreactor. There have been many foreign proteins produced in the milk of transgenic animals, proving the utility of the mammary gland as a production system. Early studies used small animals (e.g. rabbits, mice, and rats) to prove that targeted genetic expression of foreign proteins in the mammary gland was possible. Since then, successful expression of foreign proteins in the milk of livestock such as pigs, goats, and sheep has been achieved. However, though transgenic cows have been made (including a few with mammary specific promoters), foreign proteins have not yet been produced in the mammary gland of transgenic cows.

The genes for milk proteins are in general regulated by the presence or absence of hormones. In particular, hydrocortisone, prolactin, and insulin have been shown to be important for the expression of milk proteins (49, 50). These hormones were implicated especially in the transcription of milk genes, but have also been shown to be important in the post-transcriptional regulation of these genes.

Milk consists of proteins from multigene and single gene families. The caseins consist of several closely related proteins (and genes) which occur assembled into micelles in the milk. The whey proteins are in general unrelated to the caseins

and to each other and they are aqueous soluble. While both caseins and whey proteins are secreted during lactation, it appears that their genetic control elements and regulation mechanisms are significantly different. Several milk protein gene promoters have been used to drive the expression of foreign proteins in the milk. The milk promoters used include those for ovine ß-lactoglobulin (BLG, ref. 16), mouse and rat Whey Acidic Protein (WAP; ref. 15 , 51, 52), rat ß-casein (53), and bovine αS1-casein (54).

In vitro culture of mammary cell explants and immortalized mammary cells have been used to study gene regulation of milk proteins. In cell culture experiments investigating the expression of casein, it was determined that the casein gene(s) responded well to the presence of the lactogenic hormones. However, it was also seen that a major mode of casein regulation must be the stabilizing effect of prolactin on mRNA, although it appears that the coordination of regulation at both the transcriptional and post-transcriptional levels is required for efficient expression of the caseins (55). In cultured cells, it had also been determined that the cell substratum has a major effect on gene expression. The effect of cell substratum has been proposed to be due to differences in cellular morphology on different extracellular matrices, but WAP could not be expressed in cell culture under some conditions where casein was expressed (50, 55, 56). In (50), the authors noted that WAP mRNA was expressed in a constitutive manner, but appeared to be so unstable that no WAP was produced. Results from Schonenberger, et al (56) indicated that the

impediment to WAP expression was not the matrix effect itself, but rather the lack of cell-cell interactions that are present in the functioning mammary gland. The results from these experiments suggested early on that milk gene expression is not coordinately controlled between different genes. Further, the casein genes are located on the same chromosome but the WAP gene in rodents is on a different chromosome, implicating different mechanisms of control (57). This observation does not preclude control by trans-acting factors, but suggests that the effects of cis-acting elements (e.g. enhancers, MARs, LCRs) will not be equal for different loci on different chromosomes. Also, the observation that the casein genes are coordinately controlled as a gene family indicates that their regulation may be different than the regulation of the (unrelated) genes for whey proteins. Caseins and whey proteins do not appear to be controlled in the same manner in cell culture as they are in vivo. Thus, transgenic expression may have certain advantages in determining the regulatory mechanisms of genes for milk proteins.

The genes for mouse and rat WAP have been compared extensively by direct sequence comparison (49) and by in vitro culturing (50, 56-58) and transgene studies using both intact WAP genes (mouse WAP in transgenic mice, ref. 59; mouse WAP in transgenic pigs, ref. 15; and rat WAP in transgenic mice, ref. 60) and fusion WAP genes (discussed below). There are four exons in the WAP gene, coding for a 134 (in mouse) amino acid polypeptide. Endogenous WAP mRNA accounts for about 10-15% of polyA containing mRNA during peak lactation and results in expression of

WAP at about 1 mg/ml in the milk of rabbits, mice, and rats.

The WAP gene has been used in several transgenes for the production of foreign proteins in the mammary gland of transgenic animals (Human Protein C in mice, ref. 61; Human Protein C in Pigs, ref. 20; WAP in Pigs, ref. 15; Human t-PA in Mice, ref. 62-64; rabbit WAP, human α_1-antitrypsin in mice, ref 65). Some of these transgenes were reviewed in (57). It was seen that, in general, heterologous WAP-driven fusion genes resulted in much lower expression levels than that from the endogenous WAP gene. Also, the transgenes often showed a somewhat different pattern of expression in terms of both temporal regulation and tissue-specific expression. For example, the fusion genes were often expressed earlier in pregnancy than is the endogenous WAP gene, and in addition to expression in the mammary gland, the fusion genes were often expressed in the tissues where the endogenous gene is not. The reasons for the lower expression levels were not readily apparent and were speculated to be due to presence or absence or enhancer elements or attenuating elements (59). Further analysis of the transgenes provides some insight into these differences.

Human Protein C (hPC) is a regulator of hemostasis, and is thus a candidate for use in therapy of many disease states (66). Protein C is a zymogen of a very complex serine protease. The critical post-translational modifications of hPC include vitamin K-dependent carboxylation of glutamic acid, proteolytic cleavage to remove propeptide sequence and to yield the heterodimeric mature form of hPC, and

glycosylation of several amino acids (66). The present study is an investigation of the molecular biology of the regulation of WAP-hPC transgenes as part of our transgenic project to produce hPC in transgenic animals.

Goals of the Present Study

In summary, the results of the previous studies catalyze several important questions about gene regulation, the answers of which would enable more predictable expression of a foreign protein of interest:

Are endogenous genetic elements required to achieve the same tissue specific and temporal regulation for foreign coding sequences as seen for the natural coding sequence?

Can heterologous fusion cDNAs (lacking endogenous introns) be expressed at levels approaching those of the normal coding sequence from which the regulatory elements were taken?

Is proximity of the coding sequence to 5' and 3' regulatory elements important for high level expression?

What are the trans-species effects on the above questions?

What is rate limiting for a given regulatory format at the transcriptional and translational level?

In an effort to answer some of these questions, we have selected regulatory

elements taken from mouse to direct expression of a foreign protein in transgenic mice. Thus, we have eliminated trans-species effects upon the promoter in the regulation of a foreign coding sequence. However, comparison will be given from results in some of the transgenes in different species. Our previous results (61), combined with others from the literature, led us to a transgene design and philosophy which yield insight into transgene expression mechanisms. We sought to make a transgene that looked more like a natural gene in terms of the "exactness" at the junctures of the upstream untranslated region (5' UTR) and downstream untranslated regions (3' UTR) with the coding region of a cDNA (that is, lacking linker DNA and having proper positioning of the coding region). From results of this transgene and comparison to other transgenes, we desired to gain insight into mechanisms of transgene regulation. Another goal from this process was to develop an expression cassette for the insertion of a cDNA or genomic sequence under the juncture constraints listed above.

We believed from observation of the spatial relations between the coding and noncoding junctions of genes that it would be desirable to eliminate synthetic linker DNA from the transgene. Though some transgenes with linker DNA have been expressed at high levels, the effect of the presence of linker DNA cannot be predicted *a priori*.

From further observation, we believed that the 5' and 3' UTRs in a transgene should be from the same gene. Especially for tissue specific expression, it seems

likely that the 3'UTR of genes may an important part in the regulation of that gene. This type of regulation can be either at the transcriptional or post-transcriptional stage.

It also seems intuitively correct to use eukaryotic regulatory elements that are related as closely as possible to the species and tissue in which the coding sequence is to be expressed. This observation is also borne out in the literature indicating undesirable expression results using 5' and 3' elements from viral genes (60). This study presents an approach to the molecular biology of transgene design based on that of natural genes.

MATERIALS AND METHODS

Design and Cloning of Transgene Constructs

The "WAPPC1" transgene (mouse Whey Acidic Protein gene promoter driving the cDNA to human Protein C, construct number 1), was cloned at National Institutes of Health (NIH) under contract for the American Red Cross (ARC) and is shown in Figure 8. The vector was reported to be pBS (Stratagene).

The second transgene construct was cloned in the lab of John L. Johnson and was designated WAPPC2 (Figure 9). WAPPC2 is very much like WAPPC1, with only two minor changes. These differences are the deletion of the 27 "extra" base pairs (bp) of DNA between the upstream Kpn I site and the ATG of hPC and the deletion from 10 bp upstream of the poly A to the downstream Kpn I site added by the synthetic linker from WAPPC1. This change effectively removed all of the poly A that remained from the hPC cDNA. The change was made using PCR (primer positions shown on Figure 7: hPC5'KpnI 5'-CTA GGT ACC ATG TGG CAG CTC ACA AGC CTC CTG-3' and hPC3'KpnI 5'-GAT GGT ACC TTT AAT GTC CCA TCC ATT GCC ATG-3') to generate the fragment from the ATG (plus a Kpn I site) down to 16 bp upstream of the poly A of WAPPC1 (plus a Kpn I site). This fragment was then digested with Kpn I and cloned into the Kpn I site of pUC19 (with BamHI of pUC deleted). The 1.2 kbp Bam HI fragment of hPC from the WAPPC1 construct was then used to replace that fragment in the DNA derived from PCR. The

nucleotide sequence was then checked only from the upper Kpn I through the upper Bam HI (in hPC) junction and from the lower Kpn I junction upwards through the lower Bam HI (in hPC) junction. Thus, the DNA from the PCR generated fragment was found to have no inadvertent changes from the PCR. The Kpn I fragment was then replaced into the pWAPPC1 "vector" with the WAPPC1 Kpn I fragment removed, and the Kpn I junctions verified by sequencing.

The WAPPC3 construct (Figure 10) was made starting from the WAPPC2 construct. First, pUC18 was digested with Hinc II and Bam HI. A fragment (ProCS2 5'-CTC CTG CAG TGT CAC CCC GCA GT-3'+ ProCA4 5'-AAG GTG CCC AGC TCT TCT GGG GGG-3'; see Figure 7 for primer locations) from 110 bp upstream of the Sst II site of hPC down to the stop site was generated by PCR and phosphorylated. This fragment was digested with BamHI and cloned into the pUC18(Hinc II-Bam HI deletion) to generate a pUC18 with an Sst II site (pUC/Sst II). The same ProCS2 + ProCA4 fragment was also double digested with Sst II and Bam HI. A second PCR fragment (WAP3'S1 5'-GAG CAG CCG GCC CTG GCA TC-3' + WAP3'A1 5'-TAG CAG CAG ATT GAA AGC ATT ATG-3'), containing the sequence from the WAP stop codon to ~35 bp downstream of the Bam HI site of WAP, was generated by PCR and then phosphorylated and digested with Bam HI. The pUC/Sst II vector was double digested with Sst II and Bam HI and the three fragments of interest (pUC/Sst II-Bam HI vector, hPC fragment from Sst II to the stop codon, and WAP fragment from the stop codon to the Bam HI site) were isolated

from an LMPA gel and then ligated together to generate the junction of hPC with the WAP 3'.

This strategy prevented the propagation of self-ligation products of vector-vector and selected only for the insert of the two correct fragments, which are required to be blunt-end ligated together. The desired plasmid was chosen by restriction mapping the fragments inserted into the pUC vector and the sequence verified by automated sequencing of the entire region of interest (from the upstream BamHI site across the stop codon to the downstream Bam HI site. The Bam HI fragment was then isolated by restriction digestion followed by electrophoresis and recovery in LMPA.

The final vector was prepared by partial Bam HI restriction digestion of WAPPC2 and selection of the fragment containing vector plus 2.6 kbp WAP 5' plus ~1.3 kbp of hPC cDNA down to the second BamHI site plus the ~1.2 kbp of WAP 3' up to the Bam HI site of WAP. This fragment was selected by electrophoresis and the fragment recovered in LMPA. The Bam HI fragment containing the hPC-stop-WAP junction was ligated into the vector/WAPPC Bam HI fragment and selected by restriction mapping and PCR across the insert. The final clone was verified by sequencing the entire insert from the downstream Bam HI site of hPC to the Bam HI site of WAP. No inadvertent changes in DNA sequence were detected. It was assumed that no sequence changes arose from the DNA manipulation and cloning.

The transgene designated p238.18 (Figure 11) was designed and cloned at

ARC in the laboratory of Henryk Lubon.

Host Cells

E. coli DH5α was generally used as the host cell line for plasmid constructs. If it was determined that a particular plasmid was inhibitory to cell growth (due to "leaky" gene expression from the plasmid), then E. coli JM109 was used to take advantage of its tighter control of expression from the plasmid expression promoter. Both strains were treated the same for all cell handling procedures.

Preparation of Electrocompetent Cells

Host cells were prepared for electrotransformation by the method of Dower, et al (67 and 68).

Selection and Propagation of Plasmid-bearing Cells

Luria-Bertani medium with agar (LB-Agar, ref. 69) plus 100 μg/ml ampicillin was used for the selection of plasmid-bearing cells (colonies) in Petri dishes (Fisher 100 mm x 15 mm standard) after transformation of the host cell. Either LB or Terrific Broth (TB, ref. 69) were used for cell growth in the cloning and proliferation of plasmids (69). Ampicillin was used at 100 μg/ml for LB-Amp or 200 μg/ml for TB-Amp.

Whenever possible, TB-Amp was used to obtain higher cell densities and thus

higher plasmid production. However, for very large plasmids [over about 10 kilobase pairs (kbp)], TB-Amp sometimes did not work well, resulting in very low plasmid yield. When this problem was encountered, LB-Amp was used.

Cloning

For restriction endonuclease digestion, the manufacturer's directions were followed in general. Most enzymes were purchased from Gibco BRL (Bethesda, MD), with the exception of Not I (New England Biolabs, Beverly, MA).

After restriction digestion, the DNA fragments were electrophoretically resolved on a 1% agarose (SeaKem LE agarose, FMC Bioproducts, Rockland, ME) gel in 1xTris-Acetate-EDTA (TAE, ref. 69) buffer containing 0.5 μg/ml Ethidium Bromide (EtBr). The DNA was visualized on a long-wave (320 nm) ultraviolet (UV) light box and photographed. A "well" was cut (using a clean scalpel) in the agarose below the desired band of DNA, and liquified low melting point agarose [LMPA, SeaPlaque low gelling temperature agarose (FMC Bioproducts) in 1xTAE, with EtBr] was poured into the well. After the LMPA gelled, the DNA was then electrophoresed into the LMPA. Again the DNA was visualized on the light box, and the DNA in the LMPA was excised and placed into a tared microcentrifuge tube and weighed to allow calculation of the concentration of DNA in the agarose. To use this DNA in a ligation reaction, the agarose was melted at 65°C and the desired amount (5-10 pmol DNA total; vector:insert molar ratio of 1:1 to 1:3) pipetted into another

tube containing other DNA fragments or reaction components. If, upon restriction digestion, there was only one fragment present (e.g. linearizing a vector) then the restriction enzyme was removed by Phenol-Chloroform extraction and the DNA was concentrated by ethanol precipitation, redissolved in TE or H_2O and used in the ligation reaction.

Bacteriophage T_4 DNA Ligase (Gibco BRL, Gaithersburg, MD) was used to join DNA fragments (as per the manufacturer's instructions for both blunt- and cohesive-end ligations) for the construction of transgenes.

Transformation of Host Bacteria and Propagation of Cells

DNA from plasmid preparations or from ligation reactions was used for transforming bacteria by electroporation. One μl of DNA (about 2 to 100 picograms) from a "clean" plasmid preparation was used as a positive control for the electroporation or to grow cells for the generation of larger amounts of a particular plasmid if the culture bearing the plasmid was not readily available. If DNA from a ligation reaction was being used, 80 μl of H_2O was added to a 20 μl ligation, and the mixture was heated to 65°C for 15 minutes to dissolve any LMPA present and one μl of this solution was used in the electroporation process.

For electroporation, one μl of DNA (as above) was mixed with a pipette into 25 μl of cells in a microcentrifuge tube held on ice. The DNA/cell mixture was then pipetted between the electrodes of a BTX Transfector 100 (Biotechnologies &

Experimental Research, Inc., San Diego, CA) and pulsed at 650 V for 5 ms. The cell mixture was then washed from the electrode with 100 μl of SOC into 900 μl SOC and mixed gently, and the cells allowed to recover for about 1 hour at 37°C and 150 rpm rotary shaking. After the recovery period, the cells were selected by plating onto LB-Amp Agar. 100 to 500 μl of cells in the SOC recovery medium were plated onto the agar using a sterile Pasteur pipette shaped into an "L" with a Bunsen burner. The plates were dried slightly after plating by allowing to stand uncovered in the 37°C incubator for about 15 minutes. The plates were then covered and inverted and incubated overnight at 37°C.

When possible, blue/white selection of cells was used to determine the colonies containing insert-bearing plasmids. For this procedure, 50 μl of X-Gal (Gold Biotechnology, St. Louis, MO)/IPTG (Gibco BRL, Gaithersburg, MD) solution (40 μl of 2% X-Gal in dimethyl formamide plus 10 μl 0.1M IPTG in H_2O) was spread onto the top of a LB-Amp agar plate and allowed to dry uncovered at 37°C for about 30 min prior to plating. Following overnight growth of the cells, the plates were placed at 4°C to enhance the blue color.

Colonies were quantitated and then individual colonies were picked using a sterile Pasteur pipette and inoculated into separate 15 ml tubes containing 2 ml of LB-Amp or TB-Amp and grown from 10 hours to overnight.

Plasmid DNA isolation

The "Rapid Plasmid" alkaline lysis procedure (70), a modification of the Birnboim and Doly method (71), was used to isolate plasmid DNA from the overnight culture.

This procedure was scaled directly for large culture volumes, except 5M Acetate from Potassium Acetate was used for the first Ammonium Acetate step. If the plasmid preparation was to be used for preparation of DNA for microinjection or other critical procedures, the preparation was treated with a mixture of RNAse [50 μg/ml RNAse A (Sigma, St. Louis, MO) plus 50 U/ml T_1 RNAse (Sigma) final], followed by extraction with phenol-chloroform, then chloroform, and the final precipitation with ethanol and sodium acetate. These steps were used in place of the second ammonium acetate precipitation and ethanol precipitation described in the above references.

DNA Sequence Analysis

For checking the actual DNA sequence of recombinant plasmids, the protocol for the DuPont Genesis automated sequencer was followed. This method employs the use of fluorescent dideoxynucleotide tags for labelling and laser detection of the tags.

In general, the double stranded (ds) DNA was cut with an appropriate restriction enzyme and then treated with Bacteriophage T7 Gene 6 Exonuclease (T7G6, USB) according to the manufacturer. This enzyme degrades only ds DNA in

the 5' to 3' direction, leaving only single stranded (ss) DNA. After the T7G6 digestion (and inactivation by 15 min at 65°C), the ss DNA was precipitated with sodium acetate and ethanol, centrifuged, washed with 80% ethanol, and dried at 37°C. The resultant ss DNA was then used as template for the sequencing reaction.

For one sequencing reaction, about 1 pmol of template was used, with about 15 ng of primer. DNA sequence was analyzed by comparison to published sequences using the Macvector program (IBI).

For radioactive "manual" sequencing to compare to the reverse transcriptase analysis (in determining the transcription start position) the Sequenase 2.0 (USB) protocol and solutions were used. Template DNA (~ 1 pmol) was prepared by restriction digestion and T7G6 exonuclease digestion as described above.

Normally, the manual sequencing reactions were done just prior to their electrophoresis, since the band resolution deteriorated after storing at -20°C for more than 1 day. The samples were boiled for 2 min prior to running on an 8 M urea/8% polyacrylamide gel. After electrophoresis, the gel was fixed for 30 min in 10% Acetic Acid/10% methanol, and dried onto 3MM chromatography paper (Whatman, Maidstone, England) under vacuum for 1½ hr at 85°C. The gel was then exposed to x-ray film overnight at room temperature.

Preparation of DNA Constructs for Microinjection

After verification of the desired sequence of a transgene construct, a large amount of the cells containing the plasmid was grown as described above. Usually, for one construct, six 2 liter flasks containing 300 ml each of LB-Amp or TB-Amp were used for generating a large amount of plasmid. The flasks were inoculated at 1:50 from a fresh overnight culture of the plasmid-bearing cells. Plasmid from the cells was then isolated as described above. This procedure usually yielded over one mg of the desired plasmid.

A large batch (100 μg to 1 mg) of the plasmid was digested with the appropriate restriction enzyme to cut at the ends of the transgene for purification from the vector DNA. After verification of complete digestion, the enzyme was removed by phenol-chloroform and then chloroform extraction, followed by sodium acetate-ethanol precipitation of the DNA. The DNA for microinjection was then purified by one of four methods: 1) Anion exchange High Performance Liquid Chromatography (HPLC), 2) Sodium chloride gradient ultracentrifugation, 3) Bio-Rad Prep-A-Gene Kit, or 4) Agarose gel electrophoresis followed by purification using an "agarase" enzyme.

For HPLC purification, a Waters (Milford, MA) Gen-Pak FAX column was used on a Altex HPLC system (Model 334, controller model 421, 2 model 110A pumps), with Buffer A (25 mM Tris-HCl, pH 7.5, and 1 mM EDTA) as the loading buffer and Buffer B (Buffer A plus 1 M NaCl) as the elution buffer, and a total flow

rate of 1 ml/min. The separations were monitored using an ISCO (Lincoln, NE) V^4 UV scanner and chart recorder. The column was held at 50°C by an HPLC column heater (Bio-Rad, Hercules, CA). DNA (about 100 μg) was loaded onto the column at $\leq$ 55% Buffer B for gradient HPLC, and the concentration of Buffer B was increased at about one-to-2½ %B per minute until the DNA was completely eluted from the column. The elution usually occurred at about 70% B, as indicated by the HPLC monitor. This value does not account for the delay in the buffer between the mixer and the column. The peaks were collected in microcentrifuge tubes and analyzed on an agarose gel before combining correct samples and precipitating two to three times with sodium acetate and ethanol, with washes of 80% ethanol after each centrifugation step. Filtered Brinster's buffer (10 mM Tris, pH 7.5 plus 0.25 mM EDTA) was used to redissolve the precipitated DNA after each precipitation.

After each gradient run, the column was run at 100% Buffer B for about 5 minutes, followed by re-equilibration of the column at the desired percentage of Buffer B for about 15 to 30 minutes. The column was used continuously in this manner for the same DNA sample until excessive pressure buildup occurred (above 4000 psi). At that time, the column was washed with 100% Buffer B for 5 minutes, followed by a 10 to 15 minute wash with 0.2 M NaOH (the manufacturer's original instructions), and then Buffer B for 5 minutes more before re-equilibrating the column at the loading concentration of Buffer B. The manufacturer's instructions for washing were later changed to 0.04 M phosphoric acid in place of the NaOH, and this

protocol was followed after that time.

For some DNA fragments, an isocratic elution was used in place of the gradient chromatography. The optimal elution condition was usually about 62 to 65% Buffer B. These conditions worked well for the case of the WAPPC1 and WAPPC2 constructs (vector $\approx$ 3.1 kbp and inserts $\approx$ 9.3 kbp) since the difference in the sizes of the fragments was great enough and the maximum size was not exceeded. (The manufacturer recommends this column for DNA up to about 5 kbp, but it worked well for DNA at least up to about 10 kbp.)

However, for large fragments ($\approx$ 12 kbp or greater) some shearing of the DNA occurred, resulting in low yields of high quality, full-length DNA and a high percentage of degraded DNA, as seen upon gel electrophoresis of these samples (p238.18 and p225.11, from American Red Cross, ARC). Also, the column failed whenever there was contaminating genomic (High Molecular Weight, HMW) DNA in the digested plasmid preparation. This contaminating DNA resulted in the plugging of the column and high pressure buildup (greater than 4000 psi) accompanied by low flow rates. The column could be restored to normal use only by washing at 80°C overnight with the acid, followed by several hours washing with the alkali.

If the desired fragment was too large or the two fragments too close in size to be resolved by HPLC, either Sodium Chloride gradient ultracentrifugation, the Prep-A-Gene kit (Bio-Rad), or Gel electrophoresis and agarase digestion was used to purify the transgene for microinjection.

For the sodium chloride gradient ultracentrifugation (72), step gradients of five steps over the range of about 5 to 20% w/v NaCl in TE were used. The actual percentages used were determined empirically and adjusted for the construct being purified. Conditions used for WAPPC3 purification were 5 equal volume steps of 7.5, 10, 12.5, 15, and 17.5% NaCl in TE. The step gradients were formed by underlaying each subsequent (higher percentage NaCl) step below the previous one, using a syringe with a long tube attached that reached to the bottom of the centrifuge tube. Each centrifuge tube (containing the NaCl step gradient solutions) was then gently overlaid with 100-200 μg of restricted plasmid DNA. The tubes were spun at 25 krpm overnight in a swinging bucket rotor (Sorvall SW41 Ti, 11.5 ml tubes). Fractions of 500 μl were removed from the bottom of the tube using a centrifuge tube fractionator (Buchler Instruments, Fort Lee, NJ) and were analyzed by gel electrophoresis to determine the fractions containing the desired DNA. The correct samples were then precipitated three times with ethanol as described above, using Brinster's Buffer to redissolve the DNA pellet each time. This ultracentrifugation method is also limited in its ability to resolve DNA fragments that are close in size, e.g. 3 kbp and 6 kbp for the vector and insert of WAPPC3. (The resolution problem with WAPPC3 and its vector was circumvented by restricting the vector at a site that cut the vector approximately in half.) However, this method works for DNA as large as 20 kbp and possibly larger.

The Prep-A-Gene kit (Bio-Rad) was used, as per the manufacturer, for the

constructs that could not be purified by HPLC originally (p238.18 and p225.11 from ARC). This method yielded transgene DNA free of vector, but resulted in some shearing of the DNA, as determined by gel electrophoresis of the purified transgene. The shearing was probably due to the large size of the inserts (~ 12 kbp) and the manufacturer only recommends this kit up to sizes of about 5 kbp.

Another method, electrophoresis followed by "agarase" digestion, was used for purifying construct DNA, and this method proved facile for isolating any size transgene from the vector. The only constraint in the system is that the vector (or vector fragments) must be able to be resolved from the insert, using agarose gel electrophoresis.

Briefly, an agarose gel (LE, with EtBr) was poured using a comb with most of the teeth taped together so that there was only one well extending almost across the width of the gel. The transgene/vector mixture (after restriction digestion and removing the restriction enzyme by phenol/chloroform extraction, followed by ethanol precipitation to concentrate the DNA) was loaded onto the gel. Electrophoresis was carried out (~ 50 V for 2 hr to overnight) until the transgene insert DNA was well-resolved from any contaminating DNA fragment(s), as viewed by long wavelength UV irradiation. Using a scalpel, a large well was then cut out below the desired band, and the well filled with LMPA (containing EtBr) and allowed to gel for about 30 min at 4°C. After the gelling was complete, electrophoresis was again done for a short period of time (~ 30 min) until the DNA of interest was contained in the LMPA

portion of the gel. The agarose containing the DNA was then visualized by long wavelength UV light and cut out using a scalpel. The volume was estimated by weighing the gel slice in a tared tube and agarose digestion was carried out as instructed by the manufacturer of the "agarase" enzyme ("Gelase", Epicentre, Madison, WI, or ß-Agarase I, New England Biolabs, Beverly, MA) with some modifications in the recovery procedure.

Briefly, following digestion, 0.1 volume of 3M Sodium Acetate (pH 6.0) was added, mixed, and the tube placed on ice for 15 min. The tube was then centrifuged to collect any undigested agarose in the bottom, and the upper phase transferred to a clean tube. The mixture was then extracted two times with ½ volume of phenol/$CHCl_3$ and ethanol precipitated two times (using 0.1 volume of 3M Sodium Acetate for the precipitation), washing the pellet with 80% ethanol after each precipitation and drying the pellet at 37°C for 10-15 min.

Filtered Brinster's buffer (10 mM Tris-Cl, pH 7.5 plus 0.25 mM EDTA) was used to redissolve the precipitated DNA after each precipitation for all the above transgene purifications before microinjection. After the final precipitation and resuspension, the solution was filtered using a 0.45 μm spin filter (PGC Scientifics, Gaithersburg, MD). The purified DNA concentration was estimated by visual comparison of the fluorescence of a dilution series of the purified DNA to that of a known amount of Bacteriophage λ DNA, Hind III digested, after gel electrophoresis with EtBr and photography of the gel exposed to UV irradiation at 300 nm. This

estimated value was occasionally compared to the value given by spectrophotometric measurement of the 260 nm absorbance of a dilution of the sample and the values for the two methods were always in agreement.

Mice and Embryo Manipulation

The mice used in this study were CD-1 white Swiss mice (Charles River Laboratories, Madison, WI). Immature female mice 24 to 30 days old were superovulated by intraperitoneal (i.p.) injection of 10 International Units (I.U.) of Pregnant Mare's Serum Gonadotropin (PMSG, Dyosynth, Chicago, IL) followed by i.p. injection of 5 I.U. of human Chorionic Gonadotropin (hCG, Henry Schein, Inc., Port Washington, NY) 46 to 48 hours later. The superovulated mice were mated with stud males (3 to 6 months age) by placing one female with one male at the time of the hCG injection. One cell embryos were collected 21 to 24 hours after the hCG injection by standard methods (73).

Embryos were visualized using Hoffman Modulation Contrast Optics at 200x magnification fitted to an inverted microscope (Carl Zeiss, Inc, Hanover, MD, Model IM 35). The embryos were held using a heat-polished pipette made from filamented glass capillaries (World Precision Instruments, Inc., Sarasota, FL) made on a micropipette puller (David Kopf Instruments, Inc., Tujunga, CA, Model 720A) and polished using a microforge (Narashige USA, Greenvale, NY, Model MF-83).

About 1-3 pl (200-600 copies) of DNA solution was microinjected (74) into

one pronucleus of embryos collected as above. Embryos surviving the microinjection were transferred into pseudopregnant recipient female mice (30-40 embryos per mouse).

Purification of Genomic DNA

Genomic DNA from tissue of animals was isolated using the method of J.L. Johnson (75), a variation of the Marmur procedure (76). The tissue of interest was first placed in a tube and frozen in liquid nitrogen. When possible, the samples were processed directly from the liquid nitrogen. If necessary, the samples were stored at -90°C until processing, then they were placed back into liquid nitrogen just prior to beginning the purification procedure. The procedure is given below for DNA isolation from mouse tails or other tissue of about 0.25 g mass. This procedure can be scaled directly up or down with only slight modifications, which will be mentioned at the end of this section.

For isolating DNA from mouse tails, 840 μl of lysing solution was added directly to the frozen mouse tail in a 1.5 ml microcentrifuge tube and the cap closed. [Lysing solution is four parts lysing buffer (50 mM Tris-Cl, pH 8.0, 150 mM NaCl, 10 mM EDTA, 1% Sodium dodecyl sulfate, 1% ß-mercaptoethanol, and 100 μg/ml Proteinase K) plus one part 5 M Sodium perchlorate. Note: The lysing buffer is stored as a 2X solution in water. The ß-mercaptoethanol and Proteinase K are added to the solution just before using. It is important that the activity of the Proteinase K

be at least 10 U/mg, giving a final activity in the lysing solution of about 1 U/ml.]

The small cross-section of the mouse tail allows good digestion of the tissue without any processing prior to the addition of the lysing solution. However, for other tissues (e.g., pig tails or mammary gland) the tissue was crushed using a table-top vise with a pre-cooled, removable jaw made of aluminum. The jaw was placed into the flask of liquid nitrogen for a few minutes prior to crushing tissues and then placed onto the vise. The tissue was then removed from the liquid nitrogen and placed in a folded weigh paper to contain the crushed pieces and prevent them from sticking to the vise. The vise was then used to crush the tissue into pieces that were more completely digested in the lysing solution. As the aluminum jaws heated up, they were placed back into the liquid nitrogen to keep them very cold. This process was necessary to keep the tissue very cold so that it crushed well. The crushed tissue was then placed into a tube using a spatula and lysing solution was added.

Tissues in the lysing solution were then incubated overnight at 50-55°C, then removed and allowed to cool to room temperature. If necessary, the digested tissue was frozen at -20°C until it was able to be processed. For tissues digested in microcentrifuge tubes, the phenol-chloroform extractions were facilitated by shaking the tubes on a Mini-Beadbeater (Biospec Products, Bartlesville, OK) for 15 seconds each. Otherwise, the procedure in (75) was followed.

For isolating DNA from blood, the method of John, et al (77) was followed for the initial step of isolating the nuclei from leukocytes. Briefly, blood was

collected in a vacutainer tube (Becton Dickinson) containing EDTA or heparin as an anticoagulant. For 5 ml of blood, an equal volume of blood buffer (10 mM Tris-Cl pH 7.6, 10 mM KCl, 10 mM $MgCl_2$) was added and mixed gently, followed by the addition of 120 μl Nonidet P40 (Sigma) to lyse the cells. The solution was mixed gently by inverting several times. The nuclei were then pelleted by centrifugation at 2000 rpm for 10 min. The supernatant was carefully poured off and then the nuclear pellet resuspended in 3 ml of the lysing solution used for the DNA isolation from tissue. The above procedure (75) was then used from this point to complete the isolation of genomic DNA from blood. This method yielded about 30-50 μg of DNA per ml of blood.

Oligonucleotide Primer Selection

Primers were selected by visual inspection of the DNA sequence from the region of interest. Locations of primers used in this study for PCR and Primer Extension analysis are shown in Figure 7. When possible (e.g., for screening purposes), the %G+C was kept as high as possible, and the length was chosen such the predicted T_m was about 75-80°C [calculated by $T_m = 4*(G+C) + 2*(A+T)$]. Also, the primer pairs to be used together were chosen to be matched in T_m as closely as possible to minimize the preferential amplification of one strand of DNA over the other. It was empirically determined, by optimization of PCR with different primer sets, that most PCR primers will give the best yield when the annealing temperature is

set at about 20°C below the predicted T_m.

For primers with restriction sites built into the 5' end, four nucleotides were added immediately 5' to the restriction sequence. These additional bases increase the efficiency of restriction digestion of PCR products made with this type of primer.

Oligonucleotide Purification

Primers were synthesized in the laboratory of Tim Larson (Department of Biochemistry and Anaerobic Microbiology, Virginia Tech) and were trityl off. The purification scheme consisted merely of cleavage from the column and deprotection, followed by two precipitations to concentrate and de-salt; as the final step, oligonucleotide solutions were spin-filtered to remove particulate matter. To cleave the oligonucleotide from the column resin, the column was opened by removing the label and carefully separating the top and bottom sections using two pliers. The resin was poured into a 1.5 ml microcentrifuge tube and the tube nearly filled with concentrated Ammonium Hydroxide (Fisher). After capping the tube and placing in a boiling rack to secure the caps, the tube was allowed to stand at room temperature for one hour. The tubes were either shaken several times during this period or placed on a rotary shaker to mix the resin with the alkali.

After cleaving the oligonucleotide from the resin, the resin was allowed to settle to the bottom of the tube and the solution pipetted into a clean 1.5 ml microcentrifuge tube, leaving behind all the column resin. The tube was again capped

and secured in a boiling rack and placed at 80°C for one hour to deprotect the oligonucleotide. Following this period, the tubes were placed on ice to cool before opening them.

The solution was then pipetted into a 50 ml polypropylene (Oak Ridge type) centrifuge tube and the volume brought up to 4.5 ml with TE. One-half ml of 3M Sodium Acetate was added and mixed and then 3 volumes (15 ml) 95% ethanol was added and mixed to precipitate the oligonucleotide. The tubes were then placed at -20°C for at least 30 min, followed by centrifugation at 12,000 rpm for 45 min. After carefully decanting the supernatant, the pellet and tube were carefully rinsed with several ml of 80% ethanol and then inverted to drain for about 15 min. The tubes were then righted and placed at 37°C to evaporate the last drops of liquid. The precipitation step was then repeated with smaller volumes by dissolving the pellet in 2 ml TE, adding 0.2 ml 3M Sodium Acetate and 6.6 ml 95% ethanol and mixing. After standing at -20°C for at least 30 min, the tubes were centrifuged, emptied, rinsed, and dried as above. The oligonucleotide pellet was then redissolved in 500 μl of TE and filtered through a 0.45 μm spin filter (PGC Scientific). Optical density of a 1:100 dilution of each sample was measured at 260 nm, and the concentration calculated by the following equation:

$$\text{concentration } (\mu g/\mu l) = \frac{OD_{260} * 100}{35}$$

For use in PCR, a sample of the concentrated stock primer was diluted to 40 μM,

accounting for the length of the oligonucleotide, but not the individual base composition. i.e. the molecular mass for the primer was assigned the value of 331*(number of nucleotides in length).

Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used for initial screening of all potential transgenic animals. DNA oligonucleotide primers were chosen that corresponded to the transgene of interest; in general, primers that resulted in a PCR product between 200 and 600 bp were used for screening animals. Details for the selection, synthesis, and purification of primers are given above.

General reaction conditions were those of Saiki, et al (78). Reaction volumes were usually 25 μl plus 25 μl of light mineral oil (Sigma) overlay. Components of the reaction were 10 mM Tris-Cl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.1% Triton X-100, 0.5 μM each primer, 0.025 U/μl Taq DNA Polymerase (total 0.625 U per 25 μl reaction, Promega, Madison, WI, or Perkin-Elmer Cetus, Norwalk, CT). One μl of template DNA (300 ng, or $\sim$100,000 genomes) was used for initial screening of the DNA preparations. For pig DNAs, the reaction was usually repeated using 30 ng template. Thermocycling was performed on a Precision Scientific GTC-1, with conditions as follows:

40 cycles of:

1) Denaturation at 96°C for 30 sec

2) Annealing at ~55°C for 1 min (as described in **Oligonucleotide Primer Selection**)

3) Elongation at 75°C for 15 sec to 5 min, depending on target size.

For detection of transgenic animals bearing the WAPPC1, WAPPC2, or WAPPC3 constructs, the primer pair ProCS2 (5'-CTC CTG CAG TGT CAC CCC GCA GT-3') plus ProCA3 (5'-CAC CAT GTT GCT CAT GAC CTC GCT G-3') was used, resulting in a 630 bp product from DNA of transgenic animals carrying an hPC cDNA.

For generation of DNA fragments to be cloned, the reaction conditions were adjusted to minimize the introduction of mutations into the amplified fragment. That is, the original template concentration used was high (on the order of 10^6-10^7 copies) and the number of cycles was kept low (~20 cycles).

PCR products were visualized by electrophoresing a sample (5-10 μl) through a 1% agarose gel containing EtBr and subsequent photography of the gel.

Southern Analysis -- In-gel hybridization

In lieu of Southern blot and analysis, genomic DNA from potential transgenic animals was analyzed by an analogous in-gel hybridization (79). DNA was cut with a restriction enzyme as per the manufacturer. Usually, the digestion was carried out overnight, using 2-5 Units of enzyme per μg of genomic DNA. The DNA was then precipitated with sodium acetate and ethanol and redissolved in TE. The DNA

solution was then heated to 50°C for 15 min immediately prior to loading on a gel. Electrophoresis was carried out in Tris-Acetate-EDTA buffer (1X TAE = 40 mM Tris-Acetate, 2 mM EDTA, pH 8.1) and 1% LE agarose. Ethidium bromide was not used during the electrophoresis step.

Following electrophoresis, the DNA was stained with a solution of 0.5 μg/ml EtBr in TAE for 30 min with gentle shaking. The gel was photographed and then dried onto 3 MM Whatman paper, using a Bio-Rad gel dryer. Drying conditions were 30 min under vacuum at room temperature, followed by 1½ hr under vacuum at 60°C. The gel was then stored at room temperature until it was used for hybridization. To denature the DNA prior to hybridization, the gel was soaked in 0.5 M NaOH, 0.15 M NaCl for 40 min, then neutralized in 0.5 M Tris-Cl, pH 8.0, 0.15 M NaCl for an equal length of time.

Hybridization was carried out overnight in a Techne (Cambridge, England) hybridization oven, with conditions as follows: 6X SSPE (69), 5X Denhardt's solution (69), 0.5% SDS, 40% formamide, and 100 μg/ml denatured fragmented salmon sperm DNA, at 50°C for the detection of human Protein C (hPC) DNA (or RNA). The Probe DNA was hPC cDNA radiolabelled with ^{32}P-dATP or -dCTP, using the Prime-It or Prime It II Kit as per the manufacturer (Stratagene, La Jolla, CA). Following the labelling reaction, the labelled probe was purified by gel filtration through Sephadex G-25 (Pharmacia LKB Biotechnology, Piscataway, NJ), eluting with 2 ml of TE-0.1%SDS. Two μl of the purified probe were counted on a

Beckman (Fullerton, CA) LS 8100 scintillation counter, using Ecolume (ICN Biomedicals, Irvine, CA) as the scintillant. The labelled probe from a 1X reaction was denatured by boiling 5 min, then cooled on ice and the entire amount used in one 10 ml hybridization ($\sim 50x10^6$ to $100x10^6$ CPM total, $\sim 10^9$ CPM/μg). If the molecular weight standards (e.g. 1 kb ladder) were also to be detected, that particular DNA was labelled as above and included in the hybridization at (25-50)$x10^3$ CPM/ml. This amount of probe for the standards generated a good signal in an overnight exposure, whether the actual target amount was low (30 ng total) or high (5 μg).

After the overnight hybridization, the gels were first rinsed and then washed 15 min at room temperature in 2X SSC, 0.5% SDS, then washed two times for 30 min at 37°C in 0.1X SSC, 0.5% SDS, followed by two 30 min washes at 68°C in 0.1X SSC, 0.5% SDS. [Note: It was later found that the high temperature (68°C) wash may actually cause background problems by slightly melting the gel and then re-gelling and entrapping the probe in the gel matrix. This problem was much more pronounced in the RNA gels. The problem was not apparent when gels were washed at 50°C.] After the wash periods, the gels were rinsed briefly in 0.1X SSC and blotted slightly to remove excess liquid. The gels were then placed on old x-ray film and covered with plastic wrap or seal-a-meal bags and exposed to x-ray film (Kodak GBX) overnight at -70°C, with an intensifying screen. The film was developed by 5 min each in Kodak GBX developer and fixer, with a water rinse between the two and a 10 min water wash following the fixer.

RNA Isolation

RNA was isolated from animal tissue by the method of Puissant and Houdebine (80). Pieces of tissues (about 0.25 g) from biopsy or necropsy were immediately frozen in liquid nitrogen and either processed directly from the liquid nitrogen or stored at -90°C until they were able to be processed (at which time they were placed back into liquid nitrogen before processing).

Tissues were crushed by the same method used for DNA isolation and then placed directly in a 2 ml microcentrifuge tube, and the Mini-Beadbeater was used for mixing the lysing and extracting solutions as in **DNA Isolation.**

Poly-A$^+$ selection of RNA

Poly-A$^+$ RNA was enriched from total RNA by the use of oligo(dT) cellulose (Clontech, Palo Alto, CA) and the method of Celano, Vertino, and Casero (81). Total RNA was loaded onto the oligo(dT) cellulose at about 1/10th binding capacity based on the given OD_{260} capacity for poly-A$^+$ RNA, assuming 1-2.5% poly-A$^+$ RNA in total RNA (i.e., 10-25 μg poly-A$^+$ RNA per 1 mg total RNA). For example, an oligo(dT) cellulose binding capacity of 58 OD_{260} units/(g oligo(dT)cellulose) corresponds to a capacity of about 2.5 mg poly-A$^+$RNA/(g oligo(dT)cellulose). Thus, 1 mg of total RNA (10-25 μg poly-A$^+$ RNA) requires 4-10 mg of (58 OD binding capacity) oligo(dT) cellulose at 100% of binding capacity. For this procedure, 15 μg poly-A$^+$ RNA/mg total RNA was assumed (requiring 6 mg of the above oligo(dT) at

100% binding capacity) and was loaded onto 60 mg oligo(dT) cellulose. The entire procedure was carried out in microcentrifuge tubes.

The RNA samples were dissolved in TE-0.1%SDS and adjusted to 0.5 M NaCl with 5 M NaCl and then applied to oligo(dT) cellulose that had been hydrated in the same buffer with 0.5 M NaCl. The tubes were mixed well by inverting and incubated at 37°C for 10 min, mixing occasionally. After this binding period, the tubes were centrifuged at 12,000 rpm for 5 min and the supernatant removed. An equal volume of the loading buffer was added and mixed and again the tube centrifuged and the supernatant removed. Bound RNA was then eluted by three successive rounds of adding elution buffer (TE), incubating at 55°C 5 min, centrifuging as above, and removing the supernatant. The eluted RNA was then precipitated with sodium acetate and ethanol, followed by dissolving in water and adding ethanol for storage as in **RNA isolation**.

RNA gels and hybridizations

RNA electrophoresis was carried out in formaldehyde/agarose gels as in (82). The gels were subsequently dried and used for hybridization (83) using the same drying and hybridization conditions as described for the DNA in-gel hybridizations, omitting the denaturation/neutralization steps prior to hybridization. For molecular weight markers, the 1 Kb ladder DNA fragments were treated as RNA samples (84) and about 30 ng per lane was used. The denaturing conditions of the sample and

loading buffers was sufficient to completely denature the DNA fragments. An hPC PCR fragment of known size was also used as a molecular weight marker and a positive control for hybridization.

A computer program was used to estimate the sizes of the unknown fragments by the mobility relative to that of known standards. The program fits a semi-log curve to the mobilities of the standards and then estimates the unknown sizes by comparison of their mobilities to the standard curve.

RNA Dot Blots

For quantitation of relative amounts of RNA transcripts, dilution series of total RNA were applied to a nitrocellulose membrane, using a Mini-fold (Schleicher & Schuell, Keene, NH) dot-blot apparatus. To ensure that background DNA was not giving rise to a signal, a sample of WAPPC3 mouse RNA was treated with 0.25 M NaOH (15 min, 65°C) to hydrolyze the RNA and then neutralized and applied to the membrane as a control. The RNA, in water, was heated to 65°C for 15 min, then ice cold 20X SSC was added to 6X SSC final. The RNA was then applied to a pre-wet nitrocellulose membrane (that had been equilibrated in 6X SSC) and then each well was washed with 3 volumes of 6X SSC equal to the application volume. The RNA was crosslinked to the membrane using a Stratagene crosslinker on the automatic setting and then baked overnight at 68°C. The membrane was pre-wet in 6X SSC followed by hybridization, washing, and exposing to x-ray film as above. The

individual dots were then cut out using a hole punch and scintillation counted as above.

RNA primer extension analysis

Primer extension was done to determine the positions of the 5' ends of the various transcripts. RNA was precipitated and then redissolved in 5.5 μl of 5X Sequenase 2.0 labelling mix (7.5 μM each dGTP, dCTP, dTTP, for use with radiolabelled dATP). One μl of primer (2.0 pmol) was added and mixed and the solution placed in a boiling water bath for 5 min. The tube was spun briefly to collect the condensate and then 2 μl of 5X AMV reverse transcriptase (AMV RT) buffer (Promega) was added and the primers allowed to anneal for 10 min at 37°C. One μl of ^{32}P dATP (3000 Ci/mmol) was added and mixed with the pipette tip, followed by the addition and mixing of 12 U AMV RT. The primer extension was done for 30 min at 37°C, followed by a chase of 3 μl of 2.5 mM each dNTP for 30 min at 37°C. The reaction was stopped by the addition of 13 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% each of Bromphenol Blue and Xylene Cyanol FF). The tube was placed in a boiling water bath for 5 min prior to loading 2 to 3 μl per lane on an 8M urea/8% polyacrylamide gel. Reactions were stored at -20°C until use, and were found to be stable for over one week.

The oligonucleotides used for the primer extension analyses were ProCA8 5'-GGT GGC CAC GAA CAG CAG-3' and WAPA1 5'-GAC TTG TTC CTC TAG

GTT CTG AGC G-3'.

RNAse H analysis

RNAse H analysis (44 and 85) was done to assess the length of the 5' leader and 3' tailer sequences, as well as the length of the poly A tail (Figure 12). A combination of the methods in (44 and 85) was used, with the chosen conditions being 40 mM Tris-Cl, pH 7.5, 1 mM DTT, 10 mM $MgCl_2$, and 75 mM KCl. From 1 to 50 μg of total RNA or polyA$^+$ RNA (equivalent to 100 μg total RNA) was used, with either 200 or 500 ng sequence specific oligonucleotide (ProCA5, 5'-GTG GCG GGC TGG GCC AG-3') or oligo$(dT)_{25}$, respectively, per 10 μg RNA in a 50 μl reaction volume.

The RNA/primer mixture in water was heated to 65°C for 15 minutes, followed by the addition of 10X buffer to the final concentrations specified above. One unit of RNAse H (Promega) was added and mixed, and the reaction allowed to continue overnight at 37°C. EDTA was added to stop the reaction, and the mixture extracted with phenol/chloroform and then precipitated with sodium acetate and ethanol. The RNA was then electrophoresed on a 2.5% agarose gel, and dried and hybridized as described above.

Mouse Milk Collection

Lactating mice were anesthetized using methoxyfluorothane (Metofane;

Pitman-Moore, Washington Crossing, NJ) and injected intramuscularly with 0.2 ml (5.0 I.U.) of oxytocin (Vedco, Inc., St. Joseph, MO) in saline. Milk was collected into microcentrifuge tubes through flame-polished glass capillary tubes (Kimax brand, 2.0 mm i.d.) at a vacuum of 12 cm H_2O (61). The whole milk was then placed on ice and either processed immediately as below or frozen until further processing.

Frozen milk was thawed on ice and the thawed milk (or fresh milk) was mixed with an equal volume of ice-cold 2X milk buffer (2X milk buffer is 100 mM Tris-Cl, pH 7.0, 200 mM EDTA, 300 mM NaCl). All subsequent manipulations were done on ice or in a cold room (4°C). The diluted milk was then centrifuged at 15,000 rpm for 15 min (4°C) to separate the fat (upper), whey (middle), and precipitate phases. The whey was then removed to a clean microcentrifuge tube and frozen (-90°C) until analyzed by ELISA.

Protein C Detection and Analysis

Recombinant Protein C in whey was detected using a polyclonal enzyme-linked immunosorbent assay (ELISA). Immulon II plates (Fisher Scientific) were coated overnight with 100 μl/well of 5 μg/ml of rabbit anti-human Protein C (Sigma Chemical Co) in 0.1 M $NaHCO_3$ (Sigma), pH 9.3 at 4°C. Wells were washed with TBS/Tween (12.5 mM Tris, 50 mM NaCl, pH 7.0, plus 0.5% vol/vol Tween 20, Sigma). 100 μl of standard and samples in the dilution buffer [TBS /0.1% BSA (Sigma), pH 7.0] was added to all the wells and incubated at 37°C for 20 minutes.

Wells were washed four times with TBS/Tween and the bound Protein C was detected by a sandwich of goat anti-human protein C IgG (American Diagnostics, Parsippany, NJ) and HRP conjugated sheep anti-goat IgG (Sigma). Bound chromophore was detected at 490 nm using an EL308 Bio-Tek Microplate reader.

Protein C and recombinant Protein C were analyzed by electrophoresis through sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE), followed by staining (86) or Western blotting (87).

Protein C Activity Assay

The biological anticlotting activity of recombinant Protein C from milk was assayed by an Activated Partial Thromboplastin Time (APTT) Assay (61).

RESULTS

TRANSGENE CONSTRUCTS

In WAPPC1 (Figure 8), the promoter of the mouse Whey Acidic Protein (WAP) gene [about 7.2 kilobase pairs (kbp) contained on the Eco RI fragment] was used to drive the production of human Protein C (hPC) in the mammary gland of transgenic animals. The cDNA for hPC was cloned into the unique Kpn I site (K#1) of the mouse (WAP) gene immediately upstream of the ATG translation start site of WAP. The 5' element or promoter of the WAP gene contains about 2.6 kbp from the upper Eco RI site down to the Kpn I site (K#1). In WAPPC1, there are 27 "extra" base pairs (from 22 bp of linker DNA and 5 bp residual hPC cDNA 5' sequence) between the WAP Kpn I site and the start codon of the hPC cDNA. Downstream of the hPC stop site (TAG), there are 68 base pairs (bp) of hPC 3' DNA, 76 bp of poly A retained from the cDNA, and 28 bp of synthetic linker including a Kpn I site (K#2). From this region down to the end of the transgene is the rest of the WAP gene, including the coding region (exons and introns) and about 1.3 kbp of 3' DNA down to the natural Eco RI site.

WAPPC2 (Figure 9) is very much like WAPPC1, with only two minor changes. These differences are the deletion of the 27 "extra" base pairs (bp) of DNA between K#1 and the ATG of hPC and the deletion from 16 bp upstream of the poly A down to the Kpn I site (K#2) added by the synthetic linker from WAPPC1. This

change effectively removed all of the poly A that remained from the hPC cDNA.

The third transgene construct, WAPPC3 (Figure 10), was originally proposed by John L. Johnson and can be described as a single exon gene, with the promoter and downstream elements of WAP, but coding for hPC. The upstream WAP DNA spatially matches the start codon of hPC as it would match that of WAP, and the downstream WAP DNA similarly matches the 3' end of the coding region of hPC as it would normally match the coding region of the WAP gene. However, the stop codon of the WAP gene (TGA) was used in place of the stop codon of the hPC gene (TAG). The entire coding region (exons and introns) of WAP was removed. Comparing WAPPC2 and WAPPC3, one can see that the sequence at the start codon is the same. However, no introns are contained within WAPPC3. The total size of the WAPPC3 transgene is about 5.3 kbp.

The p238.18 transgene (Figure 11) is also a WAP-hPC fusion gene, but it contains the genomic coding sequence for hPC. p238.18 contains about 4.2 kbp of WAP 5', joined to the ~9 kbp hPC genomic sequence (exons and introns) by 26 bp of linker DNA. The 3' end is about 400 bp of hPC 3' sequence, and the transgene is bounded by Not I linker DNA.

The sequences at the heterologous junctures of the transgenes are given on Figures 8 to 11. Any regions that had been generated by PCR were verified by sequencing. No inadvertent changes were detected. All junction points were verified by sequencing.

GENERATION OF TRANSGENIC MICE

Note: Founder animals are designated G_0; subsequent generations are designated F_1, F_2, etc. For the WAPPC1 construct, 105 animals were born (from embryos microinjected into one pronucleus) and 30 were transgenic (28.6%) as determined by PCR. Ten of the founder females were tested for rhPC expression in milk, with levels ranging from 10 ng/ml to 3.4 μg/ml. One of these lines, #74, has been carried out to the F_7 generation, and an F_6 mouse was used for most of the WAPPC1 RNA characterization.

For WAPPC3, a total of 60 mice were born from pronuclear microinjected embryos, and 10 were transgenic (16.7%) as determined by PCR. In a separate study (88), 22 WAPPC3 transgenic mice were made by cytoplasmic injection of the transgene. Two mice generated by pronuclear injection and two mice generated by cytoplasmic injection were used for the RNA characterization.

The p238.18 mouse used in this study was received as a gift from Henryk Lubon, American Red Cross (ARC).

DNA In-gel Hybridization

Figure 13 is a picture of a (mammary gland) DNA in-gel hybridization, probed with the hPC cDNA, for a control mouse, a WAPPC1 mouse, six WAPPC3 mice, and a p238.18 mouse. Lanes 1 and 20 are 1 kb ladder size markers. Lanes 2-4 are 250, 25, and 5 pg of WAPPC1 DNA, respectively. Lane 5 is 10 μg Eco RI digested

control mouse DNA, showing no signal in the same range as the WAPPC1 DNA. All mice, except #2-10 in lane 14, show a band at about 1500 bp that is apparently from hybridization to an endogenous mouse sequence. Lane 6 is 10 μg Eco RI digested DNA from WAPPC1 mouse F_6-1 (F_6 of line 74). The "band" is not clear, and is of higher molecular weight than the transgene; this indicates either partial digestion of the genomic DNA at the transgene junction points, or possibly altered Eco RI sites at the ends of the transgene at the genomic insertion points.

Lanes 7-9 contain 250, 25, and 5 pg of WAPPC3 DNA, respectively. The full-length transgene is about 5.3 kbp. Lanes 10-15 are 10 μg each of Eco RI digested DNA from six WAPPC3 mice -- Mouse 7-3-9, 21-14, 23-15, R46-2-2, 2-10, and 23-16, respectively. Each WAPPC3 mouse exhibited a clear band at the same size as the transgene. Transgene copy numbers were estimated to be 3, 11, 4, 11, 4, and 5 for mouse 7-3-9, 21-14, 23-15, R46-2-2, 2-10, and 23-16, respectively. Additionally, three of the mice (Lanes 11, 13, and 14) showed a band at about 11 kbp, indicating partial digestion of concatemers of the integrated transgenes that resulted in some intact dimers of WAPPC3. Alternatively, these bands could be junction fragments of the transgene into the genomic DNA, where the transgene Eco RI sites were lost.

Lane 19 is 10 μg of Eco RI digested DNA from a p238.18 mouse. The band at about 6.2 kbp is the Eco RI fragment from intron B to intron G of the hPC genomic sequence. The faint band at about 4 kbp is the EcoRI fragment from intron

B upstream to the EcoRI site in the 5' region of WAP. The strong band at just above 3 kbp is probably from the EcoRI fragment of a tail-to-tail concatemer ($\sim$1.51 kbp from the EcoRI in intron G to the 3' end of the transgene), or could possibly be a junction fragment of the transgene into the genomic DNA. Though not shown on this picture, the EcoRI restriction pattern of the p238.18 mouse DNA is consistent with that of an EcoRI digest of the p238.18 construct.

RECOMBINANT PROTEIN C IN MILK: LEVELS AND CHARACTERIZATION

As expected, there were great differences in the expression levels of recombinant human Protein C (rhPC) in milk of transgenic animals made with the various transgene constructs. Information regarding the average expression levels and range of expression levels for the constructs in the milk of transgenic mice is given in Table 1. One notices immediately that the highest expression levels were obtained, not unexpectedly, from the genomic human Protein C (hPC) construct, p238.18. Expression of the endogenous whey acidic protein (WAP) is about 1 mg/ml; thus, the p238.18 transgene produced rhPC at levels in some cases in excess of the endogenous WAP. The lowest levels observed were from the WAPPC2 construct, with the WAPPC1 and WAPPC3 constructs intermediate.

WAPPC1

The WAPPC1 construct in mice produced rhPC at levels ranging from less than 1 μg/ml to about 10 μg/ml (for details, see ref. 20). Immunopurified rhPC from milk of WAPPC1 mice was found to be 30-80% active (% of theoretical), as determined by Activated Partial Thromboplastin Time (APTT) of rhPC from pooled milk of six independent lines. The rhPC appeared to be very similar to Protein C from human plasma (hPC), as judged by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western analyses.

When tested in pigs, the WAPPC1 construct produced much higher maximum levels than it produced in mice. The WAPPC1 levels in pigs ranged from about 1 μg/ml to greater than 1 mg/ml (20, and unpublished observations). Recombinant hPC purified from the highest expressing pigs exhibited activity similar to that observed for hPC. It was also found that there are various populations of the porcine rhPC that are hyper- or hypo-active (based on APTT relative to theoretical activity of 250 U/mg), but the overall activity of the entire population was similar to that of hPC. However, the apparent molecular mass (M_r) of the pig rhPC was less than that from hPC or mouse rhPC, as determined by PAGE or Western analyses. Additionally, more single chain form of Protein C was seen in some populations of the porcine rhPC (and in the overall population) than in hPC, but this observation was not connected to the incidence of hyper- or hypo-active populations of rhPC (see ref. 20 for details).

WAPPC2

WAPPC2 was tested only to a limited extent in mice. The expression levels of rhPC observed for WAPPC2 in mice ranged from less than 1 μg/ml to about 5 μg/ml -- no better (and perhaps slightly worse) than those for WAPPC1, therefore it was not used for further experimentation and activity was not tested for WAPPC2 rhPC. WAPPC2 was not included in this study (except for the expression level) since no WAPPC2 mice were available at the time of the comparison.

WAPPC3

The results presented below are data from five founder animals and from eight F_1 and F_2 mice of four independent lines generated with WAPPC3. WAPPC3 in mice yielded substantially higher rhPC levels than did the other two cDNA constructs (WAPPC1 & 2) in mice (Table 1). Expression of WAPPC3 in the first lactations of founder mice was about 3-8 times that of the WAPPC1 maximal expression level in mice (WAPPC3 G_0 = 30-80 μg/ml; average 39 μg/ml; see Table 2); further, rhPC expression in the first lactations of F_1 and F_2 generations of WAPPC3 mice averaged about 10-15 times that of WAPPC1 mice (average 162 μg/ml; range 73-299 μg/ml; see Table 3).

It was found in subsequent lactations of the WAPPC3 mice that the expression levels were greatly reduced -- about 10-fold lower than in the first lactation. For some multiple-lactation mice, hPC antigen was undetectable. No multiple-lactation

WAPPC3 mouse expressed rhPC at more than about 20 μg/ml, with most mice expressing at about 5 μg/ml (see Tables 2 and 3). Although the second and third lactation results were discouraging, the apparent improvement of expression from WAPPC3 led us to continue the analysis of this construct. The reason for the lower expression levels was not determined, but it was found that all WAPPC3 mice expressed rhPC (at the higher levels described above for mice beyond the founder generation) in their first lactations. All animals (except a few founder animals that appeared to be mosaic) were able to transmit the transgene to the offspring.

Recombinant Protein C from a pool of milk from founder WAPPC3 mice was isolated by hPC 12A-8 (ARC) immunocapture and the activity was determined by APTT and was observed to be about 10-30% of theoretical anticoagulant activity. WAPPC3 rhPC was seen to be about the same M_r as rhPC from WAPPC1 (and hPC) as judged by PAGE and Western analyses. WAPPC3 has not been tested in pigs.

p238.18 -- WAP-genomic hPC

As shown in Table 1, the p238.18 transgene produced extremely high levels of rhPC in the milk of mice (up to about 10 mg/ml as determined by polyclonal ELISA), though this recombinant protein from mice was also largely inactive, at about 5% of theoretical.

In two pigs with the p238.18 construct, the expression ranged from about 100-4000 μg/ml. However, for one pig the temporal pattern of expression did not follow

that of other transgenes with the WAP promoter in mice or pigs (52, 64, and unpublished observations). That is, endogenous WAP follows a curve similar to those for most milk proteins, with expression increasing from very low levels during early pregnancy to very high levels at peak expression, which occurs at about mid-lactation. In mice, WAP RNA induction is about 1000-fold between 8 days pre-parturition and mid-lactation (59). But for p238.18 in pig 110-1, the expression was maximal ($\sim$ 1000 μg/ml) in the first few days of lactation and then dropped off rapidly to the lower levels above (100-200 μg/ml). A second p238.18 pig (110-3) showed temporal regulation similar to normal milk expression and the expression levels have reached about 4000 μg/ml. The activity of the p238.18 rhPC from pigs appears to be better than from mice, at about 30% of theoretical. Additionally, whereas the M_r of rhPC from WAPPC1 mice looked markedly different than the M_r of rhPC from WAPPC1 pigs (as judged by PAGE and Western Analysis), the M_r of rhPC from p238.18 mice and pigs appeared to be the same. This result indicates there may be a difference in primary sequence of the rhPCs from the WAPPC1 and p238.18 constructs. Alternatively, the extremely high expression levels of p238.18 in mice may change the carbohydrate amount or structure and thus the M_r and the activity. These questions are presently being addressed in the laboratory of WHV.

RNA ANALYSES

RNA from several tissues (10 μg of total RNA per lane, by OD_{260} measurement) of lactating transgenic mice and pigs was analyzed by electrophoresis and in-gel hybridizations (equivalent to Northern analysis, Figures 14 through 19). All gels were probed with ^{32}P-labelled hPC cDNA. The hPC cDNA probe was generated by PCR, using WAPPC2 as template, and was then purified only by phenol-chloroform extraction and precipitation. Due to the lack of further purification, the remnant of WAPPC2 template (including WAP coding sequences) was included in the labelling reaction and was thus available as probe for WAP mRNA. This is clearly seen in the control mouse mammary gland and salivary gland samples in the in-gel hybridizations (discussed below). However, for the dot blot quantitation, the probe was generated by another PCR using the original PCR-generated probe, and was further purified by gel electrophoresis and subsequent agarase digestion to remove the probe DNA from the gel matrix.

Total RNA from pig 29-2 lactating mammary gland had been previously analyzed (20) and was used as a positive control for the hybridization of each gel (Lane 3, "pig M"), as was the hPC cDNA (Lanes 2 and 22, "PC", same DNA as used for probe, 1456 bp) and human liver total RNA (Lane 4, "hu L"). Note that in gel 5, Figure 18, the cDNA did not denature and remained at the top of the gel. For size markers, the 1 kb ladder (Gibco BRL, Bethesda, MD) was treated exactly as RNA samples and was probed with ^{32}P-labelled 1 kb ladder. This marker spans from

0.5 kb to 12 kb.

In-gel hybridizations of RNA from a lactating control mouse and a p238.18 mouse are presented in RNA gel 1, Figures 14 A and B. Figure 14 A is a short autoradiographic exposure of about 20 hours at room temperature, and Figure 14 B is an autoradiograph that was exposed for about 84 hours at -70°C with an intensifying screen (B exposure $\approx$ 30 times more than in A, for RNA gels 1-5.).

RNA from pig 29-2 mammary gland had at least 5 bands (Lane 3 "pig M"), corresponding to sizes of about 1600, 2350, 3700, 4900, and 6900 nucleotides (nt). (Note: These sizes are somewhat different from those presented in ref. 20. The values presented herein should be more accurate than the previous one since these are averaged from three separate determinations and more size standards were included in the determination of sizes in this study.) The lowest M_r species was about 50% of the hPC-specific transcripts detected, and was assumed to be the mature mRNA. This mRNA may be the only one which is translated to rhPC. However, it is possible that the higher M_r species are also translated to rhPC (if they are transported to the cytoplasm), since the coding region for hPC is from a cDNA and is "mature" (See ref. 48 for examples of transcripts with abnormally long 3' ends but are translated normally. However, those in ref. 48 do not have introns present.); thus the higher M_r species must represent pre-mRNA transcripts corresponding to hPC plus WAP RNA and various splices of the WAP RNA (as shown in a blot probed with WAP cDNA in ref. 20).

Human liver RNA (Lane 4, "hu L") had a mature band at about 1650 nt, with no apparent higher molecular weight forms.

Other lanes are labelled as follows for the indicated tissues on gels 1-5: H=Heart, K=kidney, L=liver, M=mammary gland, S=salivary gland, I=small intestine, U=uterus, Sk=skin, Lu=lung, Sp=spleen.

Lactating Control Mouse RNA

In Figure 14, a band of about 1700 nt was seen in the control mouse liver RNA (L). It was assumed that this band corresponds to the endogenous mouse Protein C mRNA. A broad band of 600-800 nt was present in the RNA from mammary (M) and salivary (S) glands of a control mouse. The lower edge of the broad band corresponds approximately to the size published (620 nt, ref. 49) for WAP mRNA. The broadness of the band could be due to heterogeneity of the poly A tail or to artifactual diffusion of the small RNA in the gel. Also, higher molecular weight forms (~2300 and 3400 nt) were seen in the mammary gland RNA. The band at 3400 nt likely corresponds to the primary WAP transcript with poly A, as the coding region (exons and introns) of the WAP gene is about 3300 bp. The band at 2300 likely corresponds to a partially spliced transcript. No control mouse RNA in the range of 1000-2000 nt hybridized with the probe DNA. No transcripts were detected in any of the other tissues tested.

Lactating p238.18 Mouse RNA

The p238.18 mouse presented in Figure 14 expressed rhPC at about 5000 μg/ml. Very abundant transcripts hybridizing with the hPC cDNA were detected in the mammary glands of the p238.18 mouse. These were easily seen even in the short exposure (Figure 14 A) and in fact the signal overwhelmed the entire lane and the adjacent lanes in a long exposure. The mature message, about 25% of the total hPC-specific transcripts detected in the mammary gland RNA, was centered at about 1710 nt, with other bands of about 2400, 4800, and 6000 nt. Protein C transcripts were also detected in the kidney (about 1750 and 2250 nt) and very light bands were seen in the heart RNA and the intestine RNA (both $\sim$1750 nt). In the salivary gland RNA, any bands that might have been present were obscured by the adjacent mammary gland lane.

Lactating WAPPC3 Mouse RNA

The in-gel analyses of four WAPPC3 mice (three F_1s and one F_2) are presented in Figures 15 A & B and Figures 16 A & B (RNA Gels 2 and 3). These mice expressed rhPC at levels between 135 and 299 μg/ml on day 10 or 11 of lactation (one day prior to tissue collection). Four other mice from the same lines expressed rhPC at levels between 73 and 255 μg/ml for 3 individual samples each between days 3 and 15 of lactation (averages, μg/ml for these mice: day 3-5 = 119, day 10-11 = 199, day 13-15 = 98). However, upon short autoradiographic

exposures (as above), there were no bands observed that corresponded to hPC transcript in any of the WAPPC3 mice. After long exposure to the x-ray film (as above), some hPC specific bands could be seen in some of the WAPPC3 mice. Specifically, for mouse 21-14 (RNA Gel 2, Figure 15 B, lanes 13-20) transcripts of about 1600-1700 nt were seen in the kidney, mammary gland, salivary gland, and uterus RNA. Again, no transcripts of this size range were detected in control mouse RNA, indicating that these transcripts were from the WAPPC3 transgene.

The only other tissues from WAPPC3 mice with hPC specific bands (RNA Gel 3, Figures 16 A and B) were the kidney of mouse R46-2-2 (RNA Gel 3, Lane 14, ~1700 nt) and possibly the kidney of mouse 23-15 (RNA Gel 3, Lane 6, ~1700 nt). No hPC specific transcripts in the mammary gland RNA of WAPPC3 mice (except mouse 21-14) were detected by the in-gel hybridizations. However, all these mice expressed rhPC in their milk (see Table 3), and mouse 21-14 (highest level of hPC RNA detected) expressed at the next-to-lowest level of the mice for which RNA was analyzed. Liver RNA from all WAPPC3 mice contained a hybridizing transcript at ~1700 nt, as seen in the control mouse.

As expected from the WAPPC3 transgene design, the only hPC-specific transcript observed was the mature message, with no pre-mRNA or splices of pre-mRNA. Also, although the three WAP specific transcripts mentioned above were detected in the mammary gland RNA of WAPPC3 mice, all three bands were absent in the salivary gland RNA of all WAPPC3 mice.

WAPPC1 RNA in Pigs, Lactating and Non-lactating

In RNA Gel 4 (Figures 17 A and B), results from the hybridization to RNA from lactating and non-lactating WAPPC1 pigs are presented. As described before, lane 3 ("Pig M") is mammary gland RNA from the first lactation of pig 29-2. The biopsy was taken on day 55 of lactation, when the expression level of rhPC in the milk was about 500 μg/ml. See the fourth paragraph of "RNA ANALYSES" for a detailed description of the RNA from a lactating mammary gland of pig 29-2.

Lanes 5-13 are RNA from several tissues of pig 29-2 about 10 days post-lactation. A single hPC specific band ($\sim$1650 nt) was detected in the kidney RNA of pig 29-2. A band was also seen in the liver RNA, but this was assumed to be from hybridization to the endogenous pig Protein C RNA, since it was also present in the other pig liver (Lane 15). (Note: A control pig was not used in this study, but it was previously shown in ref. 20 that there were no hPC specific transcripts present in the mammary gland RNA of a control pig).

Protein C transcripts were not present in the non-lactating mammary gland of pig 29-2 (Lane 8), even though hPC transcripts were very abundant during the first lactation of this animal (Lane 3). No tissue, other than kidney, in the non-lactating pig exhibited transcripts specific to hPC.

RNA from tissues of pig 31-5, a daughter of pig 29-2, is presented in lanes 14-20 of RNA gel 4, Figure 17 A & B. Abundant transcripts were detected in the mammary gland RNA (Lanes 16 & 17) of pig 31-5, and the pattern of the bands

closely resembled that of pig 29-2 (Lane 3). There were at least 5 hPC specific bands, corresponding to sizes of about 1550, 2250, 3850, 6200, and 7300 nt. These sizes are similar to those observed for pig 29-2 lactating mammary gland, described under **RNA ANALYSES**, paragraph 4. The total amount of hPC specific RNA for pig 31-5 was visually estimated to be about 1/20th of that for pig 29-2.

Pig 31-5 exhibited a diffuse band in the liver RNA that was also assumed to be the endogenous pig Protein C mRNA.

WAPPC1 RNA in Mice

In-gel hybridization of several tissues of two WAPPC1 F_2 mice from the same founder animal are shown in RNA Gel 5 (Figures 18 A & B, Lanes 5-13). Again, the mammary gland (Lane 8) RNA had the three WAP-specific bands previously associated with the control mouse; but in contrast to WAPPC3 mice, the salivary gland (Lanes 7 & 11) also exhibited these endogenous WAP bands. The only potential hPC-specific band was seen in RNA from heart tissue of mouse 18-6-4 (Lane 5, $\sim$ 1500 nt), although the band was not well-defined and may be a gel artifact. Figure 19, lanes 5-11 (RNA Gel 6), is the in-gel hybridization of RNA from an F_6 generation WAPPC1 mouse. This gel was probed with the "cleaner" hPC cDNA, which did not contain any WAP sequence. For this WAPPC1 mouse, no distinct bands were detected.

Enrichment for Poly-A$^+$ RNA, In-gel Hybridizations

The results for the poly-A enriched mammary gland RNA and human liver RNA samples are presented in Figure 19 (Gel 6, Lanes 12-20). One mg of total RNA was processed and then 1/10th of the final sample (equivalent to 100 μg of total RNA, assuming 100% recovery) was loaded onto the gel. The gel was probed with the hPC cDNA containing no residual WAP sequence. Again, for the control mouse (Lane 12) no hPC-specific transcript was detected. However, the hPC specific transcript from WAPPC3 mice was very little enriched, if at all (Lanes 13-17). Lanes 13-16 were from the same mice (7-3-9, 21-14, 23-15, and R46-2-2) as presented in RNA Gels 2 and 3 (Figures 15 and 16). Lane 17 was from a WAPPC3 mouse on her third lactation, in which no rHPC was detected in the milk. The band previously seen for WAPPC3 mouse 21-14 (Lane 14, RNA Gel 6, Figure 19) was easily seen in the enriched sample, even for the relatively short autoradiograph presented (24 h at -70°C), but this hPC-specific band was not detected in the RNA of any other WAPPC3 mouse.

However, for the other samples enriched for poly-A$^+$ RNA there was significant improvement in the signal obtained by hybridization. Especially notable is the sample in lane 18, poly-A$^+$ enriched RNA from mammary gland total RNA of the same F_6 WAPPC1 mouse as presented in lane 8 (Gel 6, Figure 19). Though the hPC-specific transcript was not readily seen in 10 μg total RNA, it is clearly seen in the poly-A$^+$ enriched sample. In fact, this sample showed a stronger signal than that

for pig 31-5, Lane 19, which had a relatively good signal even in the total RNA. Also, the human liver RNA was enriched in the hPC-specific transcript (Lane 20). The polyA enrichment was not repeated due to limited amounts of total RNA available.

Mammary Gland RNA Dot Blots

A dot blot was done using serial dilutions of mammary gland total RNA which had been purified by the LiCl step of the protocol in **MATERIALS AND METHODS** and (80) and subsequent RNAse-free DNAseI digestion. (The LiCl step yielded RNA preparations with significantly less contaminating DNA than was obtained using the ammonium acetate step.) The hybridization was performed using the hPC cDNA containing no WAP sequence. For the WAPPC1, WAPPC3, and control RNA samples, the application amounts were 100, 10, 1, and 0.1 μg. For the DNA background and the p238.18 samples, the application amounts were 10, 1, 0.1, and 0.01 μg (amounts are μg of starting RNA for the DNA background sample). The blot, shown in Figure 20, had a high non-specific background due to probe binding to the membrane. The autoradiograph was exposed for about 40 hours at -70°C. The dot blot was not repeated due to the limited amounts of total RNA available.

Lanes 1, 3, and 10 were the control pig, control mouse, and background DNA control samples, respectively, The lack of any signal in these lanes indicates that there was no background due to endogenous pig RNA or to mouse RNA or DNA.

Lane 2 was mammary gland RNA from WAPPC1 pig 29-2. The signal is clearly seen down to the third application amount, 1 μg. Lane 4 was RNA from a WAPPC1 mouse, which shows almost no signal. Lanes 5-9 are RNA from WAPPC3 mice 23-16 (third lactation), R46-2-2, 23-15, 21-14, and 7-3-9, respectively with mouse 21-14 showing the strongest RNA signal of the WAPPC3 mice. Lane 12 contained 10, 1, 0.1, and 0.01 μg of total RNA from a p238.18 mouse and exhibited very strong signals.

The individual dots were scintillation counted to quantitate the relative amounts of hPC-specific RNA. However, due to the high background signal ($\sim$80 cpm; normal background is $\sim$30 cpm), the more faint signals were not able to be quantitated above background levels. Only the WAPPC1 pig and p238.28 mouse RNA yielded countable signals. These two dilution series yielded signals corresponding to 6130 cpm per 100 μg RNA for the p238.18 mouse and 3100 cpm per 100 μg RNA for the WAPPC1 pig 29-2. The signals for the WAPPC1 mouse (Lane 4) and the strongest WAPPC3 mouse (Lane 8) were visually compared to the signals for which values were known (Lanes 2 and 12), and were estimated to be 1 and 2 cpm per 100 μg RNA, respectively.

Primer Extension Analysis -- Determination of Transcription Start Site

Primer extension analysis was done to determine the transcription start site of the different transgenes and to compare this site to that of the endogenous WAP.

Antisense primers were chosen in the first exon of the sequences of the hPC and WAP gene. Upon reverse transcriptase extension of these primers to the 5' end of the specific RNA transcripts, the presence of a defined band(s) indicated the apparent transcriptional start site. That is, the reverse transcription continues until the end of the template (mRNA) is reached, and the end is assumed to be the first nucleotide of the transcription. This technique must assume no degradation at the 5' end of the RNA and no interference with the reverse transcriptase from the 5' cap or mRNA secondary structure.

Ten micrograms of mammary gland total RNA from a WAPPC1 mouse, pig 29-2 (WAPPC1), mouse 21-14 (WAPPC3, in which an RNA hybridization signal was obtained), a p238.18 mouse, and a control mouse were used for primer extension. The transgene primer and the WAP primer were specific for the first exon of their respective genes. WAPPC3 was also sequenced for running in an adjacent lane for the determination of the nucleotide start site of WAPPC3 mRNA and as a size marker for the other samples.

The results, using total RNA, are presented in Figure 21. Strong signals were obtained for the WAPPC1 pig mRNA (Lane 1), p238.18 mouse mRNA (Lane 2), and the control mouse WAP mRNA (Lane 3). However, no specific signal was obtained for either WAPPC1 in a mouse or WAPPC3 in a mouse (data not shown). This experiment was repeated several times, using from 1 to 100 μg total RNA or poly-A^+ enriched WAPPC1 or WAPPC3 RNA as template, with the same results. That is, no

primer extension signal for WAPPC1 in a mouse or WAPPC3 in a mouse was ever obtained.

Primer extension of the WAPPC1 pig 29-2 RNA and the control mouse WAP RNA both yielded doublets one nucleotide apart. The p238.18 mouse RNA yielded a single band. The sizes for these bands were determined by comparison to an adjacent sequencing lane (not shown). Comparing the sizes of these primer extension products, it appears that the transcription start site is the same for the three genes, relative to the WAP promoter. Specifically, transcription appears to begin with either the C or the A, at nucleotide -28 or -27, respectively, (only at the -27 "A" for p238.18) of the WAP promoter, with +1 being A in the AUG start codon of WAP. Upon sequencing this region in the DNA, two discrepancies were found compared to the GENBANK sequence for MMWAP, the genomic sequence. First, our WAP promoter contains an A insert at position -9, resulting in the sequence GACTGAGGCCATGG ATG. This insertion effectively lengthens the leader sequence by 1 nt relative to the published sequence. Second, there is a substitution of an A for C at -32, resulting in the sequence CAGACAGCCATC. This substitution is just upstream of the transcription start site. The transcription start site positions reported here reflect the DNA sequence in our transgenes.

RNAse H Analysis

RNAse H is a nuclease that degrades only the RNA strand of an RNA:DNA hybrid (85). Thus, by hybridizing an RNA transcript with a DNA oligonucleotide (specific for a particular region of the transcript) and subsequent treatment with RNAse H, one can generate site-specific cleavage of RNA. RNAs from a WAPPC1 mouse (poly-A^+ enriched mammary gland RNA), a WAPPC1 pig (pig 29-2, 5 μg total RNA), a WAPPC3 mouse (mouse 21-14, 10 μg mammary gland total RNA), a p238.18 mouse (1 μg mammary gland total RNA), and human liver (10 μg total RNA) were used for RNAse H analysis. The sequence-specific primer was chosen to yield a 5' RNA fragment larger than the 3' fragment. $Oligo(dT)_{25}$ was used to degrade the polyA tail of the mRNAs.

The pig 29-2 RNA was the only sample to yield a reasonable result (Figure 22). The other RNAs either appeared partially degraded or did not give a strong enough signal to see the bands clearly. Undigested mammary gland total RNA from pig 29-2 is shown in Lane 1, with the lowest band being the mRNA. The higher M_r bands were retained at the top of the gel. After RNAse H digestion, an upper band that corresponds to the 5' end of the WAPPC1 transcript was seen (Lane 2). The lower band corresponds to the 3' end of the transcript including the polyA tail (Lane 2) or after removal of the polyA tail (Lane 3). Accurate size estimates were not able to be obtained under the electrophoresis conditions used (migrations were not semi-log in the 2.5% agarose, thus the standard "curve" was not linear and yielded large errors

in size estimation).

Contaminating genomic DNA probably resulted in some of the RNA degradation observed; however, the experiment was repeated, using RNAs that had been further purified using RNAse-free DNAse I, with no better results. The RNAse H technique appears to be limited to analyzing a very abundant RNA of a discrete size.

DISCUSSION

Our studies have shown that the expression of a foreign cDNA under the regulation of WAP can be greatly affected by the nature of the 3' downstream sequences. In contrast to previous studies for a WAP-tPA fusion transgene, in which expression levels varied over three orders of magnitude (62), hPC expression levels from the WAPPC3 transgene varied only about 1 order of magnitude. This variation of expression for a particular transgene is normally attributed to chromosomal location of the integrated transgene(s) or to the number of transgene copies integrated into the host genome. However, we have seen ranges of expression between WAP-hPC transgenes with dissimilar downstream sequences (i.e., from a different origin than the promoter) which differ by greater than an order of magnitude. The "downstream 3' UTR elements" in consideration chiefly differ in the 3' UTR elements which may exist between the intact coding regions and the polyadenylation signal, and 3' flanking "elements" after the polyadenylation signal. Other studies have suggested that these 3' regions of the WAP gene may contain essential elements for WAP regulation (50, 52, 64, 93, and 94). A salient feature of the WAPPC3 gene is the use of the "matching" 5' and 3' UTRs from the WAP gene bounding the hPC cDNA coding region exactly as they border the coding region of WAP. Previous genomic WAP transgene studies used the same 5' and 3' elements (as in WAPPC3) but with an insertional marker within a WAP intron (59). The maximal observed expression level

of hPC from WAPPC3 in mice was about one-third the level of WAP expression from the endogenous gene and was similar to the genomic WAP transgenes having the same 5' and 3' elements (59, mRNA "expression" levels were reported as % of endogenous). In addition, other studies in transgenic mice with the same WAP 5' flanking sequence driving expression of an hPC cDNA but with an SV40 3' polyadenylation element gave expression levels of only none detected to 1 μg/ml (96). These results suggest that cDNA transgenes can be expressed in the milk at appreciable levels if 5' and 3' UTRs of WAP are kept intact and juxtaposed to the cDNA without interrupting linker or residual cDNA UTR sequences. However, the presence of linker regions in the 5' upstream and/or residual hPC 3'UTR in some of our WAP-hPC transgenes make the exact identification of key regulatory elements and mechanisms difficult.

Given below is a more detailed comparison of hPC expression in terms of tissue specificity, nature and levels of the messages, as well as the protein expression level to several other previous studies of the WAP gene and WAP transgenes.

Effect of 5' Linker Sequences Upon Regulation of WAP Transgenes

The presence of linker DNA in transgenes which intervenes between 5' protein coding sequences and 5' WAP sequences can exert a negative impact on expression by interfering with transcription (59). The presence of linker sequences could also affect recognition of the resultant open reading frame by the ribosome complex.

Kozak (1 and 46) has shown that the sequence around the translational start site can be critical for recognition of the correct AUG and initiation of translation. Figure 23 compares the translation start region of the WAPPC series (WAPPC1, WAPPC2, and WAPPC3) and p238.18 transgenes with the consensus nucleotides determined by Kozak. In both WAPPC1 and p238.18, the first 5 nt upstream of the ATG are retained from the cDNA of hPC. The -6 nt of WAPPC1 and the -9 nt of p238.18 match the consensus, but it appears from examination of the consensus that the most critical bases are at positions -1, -3, and -4, (C, A, and C, respectively), which are all greater than 50% conserved. Of these three nt, both WAPPC1 and p238.18 match positions -3 and -4. The leader sequence of the WAP gene was present in this position for the WAPPC2 and WAPPC3 transgenes, and the WAP gene sequence matches the consensus at positions -1 and -3. Further, Kozak experimentally determined that positions -3 and +4 of the consensus were the most critical for efficient translation initiation (46); it was noted that, for practical purposes, the initiator codon could be designated "strong" or "weak" solely on the basis of positions -3 and +4. All the transgenes in question match the consensus at position -3. The +4 position is determined by the hPC sequence and does not match the consensus; however, all the transgenes have the same +4 nt (T) from the hPC sequence. The studies of Kozak also implied that the +4 nt only came into play in the absence of a purine at -3.

The known leader sequences (5' UTRs) of most eukaryotic mRNAs are in the

range of 20-100 nt. Whereas a very short (less than 20 nt) leader sequence might be inefficient in translation initiation at the first AUG codon, there is no evidence that very long leader sequences are incompatible with efficient translation, provided that secondary structure and upstream AUG codons are not present (46). Thus, based of their length and lack of ATG "codons", the 5' linkers of WAPPC1 and p238.18 probably do not have an inhibitory effect on translation considering the information from Kozak (1 and 46).

Studies using an intact WAP genomic transgene with 5' linker sequences have shown that linkers at the translational start site can attenuate transcription of these WAP transgenes in transgenic mice relative to intact genomic WAP transgenes having only insertional changes (in otherwise intact genomic sequences) located in the introns of the WAP gene (59). This observation may provide sufficient reason for the lower expression of the WAPPC1 transgene as the WAPPC1 5' linker size is similar to and the linker position is the same as that ("KH" allele) in the report of Burdon, et al (59).

Tissue Specificity

WAPPC3 transcripts were detected at similar, low levels in the kidney, mammary gland, salivary gland, and uterus. In contrast, the relative levels of p238.18 transcripts in kidney, salivary gland, and uterus were much lower than in the mammary gland. Previous transgenic mouse studies of a WAP-tissue Plasminogen

Activator (tPA) cDNA transgene (62) having the same 5' WAP UTR sequences as the WAPPC series transgenes showed tissue specificity (relative levels) of message similar to the p238.18 transgene. In the WAP-tPA study (62) and in a study of rat genomic WAP in transgenic mice (57), low levels of transgene RNA were detected in the kidneys. Also, in (52) endogenous mouse WAP mRNA was detected in the kidneys at low levels. However, unlike the WAP-tPA and endogenous WAP studies, the hPC transcripts in the kidneys of WAPPC3 mice were at levels equal to or higher than the hPC-specific RNA levels detected in the mammary gland. In one WAPPC3 mouse (21-14) the hPC-specific RNA was about equal in kidney, mammary gland, and salivary gland. Thus, the WAPPC3 format may have a different constitutive expression than other WAP-directed transgenes in non-mammary tissue.

Effects of 3' Transcript Processing: Length of PolyA Tail

The WAPPC series transgenes differ in 3' UTR and thus in the expected polyadenylation which would result from 3' processing of all transcripts from the transgene. In particular, the 3' of WAPPC1 has two potential sites of polyadenylation: an already existing sequence of polyA of 76 nt in length, which is residual from the hPC cDNA, and the complete polyadenylation signal contributed by the genomic WAP 3' sequence far downstream. Only a residual, partial polyadenylation signal (AATAAA consensus, ATTAAA in hPC gene) is present from the hPC cDNA in the WAPPC1 transgene. The WAPPC2 transgene does not have

the residual 76 nt of polyA, but does retain the ATTAAA partial polyadenylation signal; however, no RNA analysis was done for WAPPC2. The WAPPC3 transgene has only the polyadenylation sequence from the WAP gene. The p238.18 genomic hPC transgene has two potential polyadenylation sites contributed by the hPC 3' genomic sequence (91). The predicted and observed (mature) mRNA sizes for each construct are presented in Table 4. Values given for the predicted sizes are calculated from the predicted transcriptional start site to the predicted polyA addition site. The difference in the two values for a particular construct therefore gives an estimate of the length of polyA present in the mRNA from that construct.

For WAPPC1, the apparent length of the polyA tail is about 100 nt for that transgene in both mice and pigs. We cannot determine if the polyA in the mRNA is present as a result of residual polyA contributed by the transgene or from addition of polyA by polyA Polymerase. It is apparent, from the higher M_r hPC-specific transcripts from WAPPC1 in pig and mouse, that transcription can proceed across the polyA region of WAPPC1, but the actual processing after that point is not clear. Previous studies (34, 35, 38, and 39) suggest that the residual polyadenylation signal (ATTAAA) from the hPC cDNA would not be sufficient to allow efficient cleavage and polyadenylation at the correct site relative to the hPC coding region. No mechanism is known to enable the WAP 3' elements of WAPPC1 to direct proper cleavage and polyadenylation at the far upstream hPC (partial) polyadenylation site in WAPPC1. The more likely polyadenylation of WAPPC1 appears to be that

contributed by the residual polyA of the transgene to yield a mature (polyA$^+$) transcript of about 1600 nt, although it is not clear how the transcript is terminated at this point.

WAPPC3 mRNA has an apparent polyA length of about 150 nt, which is greater than that for WAPPC1 and similar to the range calculated for the polyA of endogenous WAP messages which have been reported. The endogenous WAP mRNA length (minus polyA) is predicted to be 561 nt. Using the literature value of 620 nt (49) for the polyadenylated WAP mRNA yields a polyA length of 59 nt, but the observed mRNA size in this study yields a polyA length in the range of about 50 to 250 nt. Furthermore, the WAPPC1 mRNA appears to have less polyA than that for WAPPC3 according to the mobility measurements, but it was much more enriched through selection by oligo(dT) than was WAPPC3 mRNA. The reason for this apparent discrepancy is unclear.

The predicted size of mRNA from the p238.18 transgene is 1511 nt. This size includes the hPC 3' end down to the same polyA addition site as is in WAPPC1. However, Foster et al (91) report two possible polyadenylation sites located at 68 nt and 294 nt downstream of the stop codon. Therefore, if the proximal polyA site is chosen the predicted size (minus polyA) is 1511 nt and if the distal site is chosen the predicted size is 1737 nt. The observed size for the mRNA is about 1700 nt; therefore, either the proximal site was used and there were about 200 bases of polyA, or the distal site was chosen and there was virtually no polyA. Selection by oligo(dT)

was not done for the p238.18 RNA sample.

POTENTIAL MECHANISMS FOR DIFFERENCES IN PROTEIN LEVELS

Comparisons of the low levels of mRNA and protein produced from WAPPC1 with the low protein levels in WAPPC2 in transgenic mice suggest that either transcription is inefficient and/or the stability of the transcripts is low. Further, translation of the transcripts may be inefficient. In contrast, the WAPPC3 transgene in mice produced significantly higher levels of protein but with low levels of mRNA. Thus, transcription rate and/or mRNA stability may be low for all WAPPC series constructs. Since WAPPC3 has the same 3' UTR as the native WAP mRNA, we would expect it to have stability which is similar to WAP mRNA, though this discounts the intrinsic stability of the coding region RNA (57); it appears from the absolute specific RNA levels and lack of obvious degradation products that WAP mRNA is not unstable in vivo.

Comparing WAPPC1, WAPPC2, and WAPPC3 expression in mice, it appears that the low levels of WAPPC1 and WAPPC2 expression must be due to poorly functioning and incomplete 3' ends next to the hPC coding region. But the issue may not simply be differences in mRNA stability, since the WAPPC1 transgene in pigs did not appear to be unstable (judged by the absolute amount and the lack of obvious degradation products). Also, in the mouse, intact WAPPC1 mRNA was greatly enriched by polyA selection, which indicates that it was probably not degraded, even

though the levels were low. However, it is difficult to delineate the factors involved, since, judging from the expression levels, it seems that the pig regulates the WAPPC1 transgene differently than does the mouse. Another complicating factor, mentioned above, is that it is not known which of the specific sizes of WAPPC1 transcripts are actually transported to the cytoplasm and translated. Comparison of cytoplasmic versus nuclear RNA could have helped resolve this question, but this was not done.

Comparing WAPPC2 and WAPPC3, we see that the two transgenes are the same from the 5' end through the hPC coding region. Yet, the expression levels from WAPPC3 were about 50-fold greater than WAPPC2. Also, in WAPPC3 the WAP stop codon was substituted for the hPC stop codon. In light of the universality of the three stop codons, this substitution is not expected to have any effect on the expression from these transgenes. The only difference at the 5' end between WAPPC1 and these two transgenes (WAPPC2 and WAPPC3) is the presence of 27 bp extra (linker + hPC 5' UTR) DNA in WAPPC1 just upstream of the hPC start codon. As discussed above, the presence of the 5' linker in WAPPC1 may suppress transgene expression at the transcriptional level. Since the presence of introns has been linked to examples of high expression levels, it is somewhat surprising that WAPPC3, lacking introns, was expressed at the high levels reported here. Further, it appears that the format of the WAPPC1 and WAPPC2 transgenes does not enable the WAP introns to exert a dominant positive effect upon expression.

The p238.18 transgene had about 1.6 kbp more upstream WAP gene sequence

in its 5' flanking DNA, and this extra WAP sequence could potentially increase (or decrease) the expression from this transgene. However, in another construct (ARC p230.12, unpublished observations) the longer WAP promoter was used on the WAPPC1 transgene. Though only a few mice were evaluated with this transgene, the expression levels appeared to be about the same as that from the WAPPC1 transgene (data not shown). Therefore, it appears that the extra sequence in the promoter of p238.18 had little effect. From the literature (15-18), it would seem likely that the genomic hPC sequences of the p238.18 transgene would be the dominant factor in bringing about the extremely high expression levels observed. However, there could also be an effect from the 3' end of the transgene, which is 400 bp of the hPC gene 3' end. This 3' end could potentially increase (or decrease) the stability of the mRNA in the mammary gland, or could possibly even have a direct effect on the transcription or translation of the transgene. However, we have found no evidence suggesting an effect of 3' UTR on transcriptional efficiency. Introns are believed to increase transcription by catalyzing changes in the chromatin structure in and around the transgene.

The total amount of p238.18 mRNA indicates that the transgene may not be regulated in the same manner as endogenous WAP in the mammary gland. While it is difficult to determine the expected relative amounts of heteronuclear RNA (hnRNA) to mRNA, human liver RNA shows similar (hnRNA) band intensity to p238.18 mouse mammary gland RNA when applied in amounts yielding similar intensity of

mature hPC message (data not shown). Thus, no rate limitations in spliceosome processing seem to occur at the high levels of transcription seen with p238.18 in transgenic mice.

It is noteworthy to compare the WAPPC3 expression levels with those of other transgenes containing cDNAs driven by WAP in transgenic mice. In WAPPC1, the maximal expression level of hPC was about 1% of that of endogenous WAP. WAPPC2 expressed hPC at maximal levels less than 0.5% of endogenous WAP. For the expression of a WAP-tPA transgene in mice (62), the highest expression level observed was estimated to be only 5% of endogenous WAP expression and the expression between different transgenic mouse lines ranged over 3 orders of magnitude. In striking contrast, the WAPPC3 maximal expression levels were ~30% of endogenous WAP and average levels were ~16% of endogenous WAP. Also, the range of expression from WAPPC3 was less than one order of magnitude. Thus, the expression levels from the WAPPC3 cDNA format were markedly higher and more consistent than for previous cDNA transgenes using the WAP promoter. Therefore, the positional effects on WAPPC3 were apparently not very strong. However, transgene copy number was not correlated with expression level for WAPPC3. Further, all WAPPC3 female mice have expressed hPC on their first lactations. This result is in contrast to other studies using the WAP promoter where only a fraction of the mice expressed the transgene (62 and 92). It is not known why the transgene expression levels of WAPPC3 mice were dramatically decreased on subsequent

lactations. The effect could be due to developmental regulation which attenuates or inactivates the WAPPC3 transgene due to its integration site or its gene structure.

In vitro studies (using both immortalized and primary mammary cells) of WAP gene regulation indicate that the WAP gene requires a trans-acting factor(s) (other than prolactin, insulin, and hydrocortisone) which may not be synthesized in mammary tissue (50 and 56). Alternatively, the required factor may be made in mammary tissue in vivo but the cells may lose the ability to make this factor upon culturing in vitro. In (55), it was determined that the hormone prolactin exhibited both transcriptional and post-transcriptional effects on the regulation of the casein gene in question. The observed effects of prolactin on expression were: 1) a 2-4 fold increase in the rate of transcription of the gene, and 2) a 17-25 fold increase in the stability of the mRNA. It was not shown, but is likely that the stabilizing effect due to prolactin is exerted at the 3' UTR of the mRNA (see **Post-Transcriptional: mRNA Stability** in **LITERATURE REVIEW**). In a related study, Eisenstein and Rosen (50) showed that glucocorticoids influence casein gene expression solely at the post-transcriptional level; but this effect was seen on casein mRNA (i.e. not casein protein), so the hormonal effect on casein gene expression is apparently a stabilizing effect of the mRNA. From the same study, the WAP gene in cell culture was not responsive to the hormones prolactin, insulin, and hydrocortisone as the ß-casein gene was. In fact, WAP transcripts were detectable by kinase-treated RNA run-off transcription, but not by blotting. This result indicated that the WAP transcripts that

were present were degraded, signifying that the WAP mRNA was unstable. Thus, in (50) the WAP gene in cell culture appeared to be controlled, at least in part, at the post-transcriptional level by altered mRNA stability.

Several other studies have implied or speculated that the 3' end of the WAP gene is important in the regulation of the gene (50, 52, 64, 93, and 94). Most of these studies also implied or hypothesized that a major level of WAP gene regulation was post-transcriptional. Dale, et al (60) implicated the 3' UTR in proper expression of the transgene, though the mechanism(s) was not clear. Further, the authors proposed that the WAP 3' elements should be used in preference to heterologous 3' elements for transgenes with the WAP gene promoter. This recommendation was given on the basis of the circumstantial evidence of low-level expression from WAP-driven heterologous transgenes with 3' elements other than from the WAP gene. In (93), Hennighausen, et al determined that the 3' UTRs of rat and mouse WAP genes had 91% sequence identity, whereas the coding regions of the genes were 82% conserved. They speculated that the highly conserved nature of the 3' region might be indicative of a functional significance at the DNA or RNA level.

From the in-gel RNA hybridization and the dot blot measurements, we cannot conclusively make comparisons between absolute RNA levels from the various transgene constructs. However, from both the dot blot measurements and the RNA in-gel hybridizations (estimated from multiple exposures of the gels for different lengths of time) we observe that the p238.18 mouse produced about three orders of

magnitude (10^3-fold) more hPC-specific RNA than did WAPPC3, yet the p238.18 mouse in the present study only secreted about one order of magnitude more rhPC than did WAPPC3 mice. Another p238.18 mouse also produced about three orders of magnitude more hPC-specific RNA than WAPPC3 mice, yet only secreted about 4-fold more rhPC than did the WAPPC3 mice (data not shown). Again, based on the limited number of animals in this study, we cannot make the generalization that these observations hold true for all animals made with these transgenes.

In comparing the fairly high expression of rhPC from the extremely low levels of WAPPC3 RNA to the values for the other constructs, we believe that the mRNA from the WAPPC3 transgene may be much more efficient at being translated than the other transgenes, although the conclusion is tentative for the following reasons: 1) only one WAPPC1 mouse, two WAPPC1 pigs, and one p238.18 mouse were used for the present study; thus, the observations comparing these samples may not be valid for generalization due to differences in position effects and statistically insignificant samples; 2) the high expression levels of WAPPC1 in pig and p238.18 in mouse may have reached a rate limitation (or saturation) at the translational level or at the transport (secretion) level; and, 3) the high expression levels from WAPPC1 in pig and p238.18 in mouse may have led to an intracellular degradation of rhPC, thus skewing the apparent translational efficiency to a lower value. However, the possibility of buildup of (intact) rhPC in the mammary gland is unlikely, since intracellular hPC buildup has been shown to result in cell death (95) and the animals

in question exhibited no mammary gland pathology (data not shown). Also, in comparing WAPPC1 in pig 29-2 and p238.18 in mouse, we see that the p238.18 mouse produced only twice as much hPC specific RNA as did WAPPC1 in pig 29-2, but the p238.18 mouse produced ten times more rhPC than WAPPC1 in pig 29-2. This observation does not preclude species-specific effects, but suggests that the factor of translational efficiency is involved.

Zaret and Sherman (48) were able to conclusively show that differences in translational efficiency were occurring in the mutant yeast systems they analyzed. This conclusion was possible due to the common genetic background shared by the yeast systems. Though we suggest that translational efficiency may be involved in our transgenic systems and Zaret and Sherman proved that it was a factor in a mutant yeast system, we have found no reports with evidence for differential expression regulation through altered translational efficiency of natural genes. Zaret and Sherman showed a linear relationship between the mRNA and protein levels for the normal yeast gene and for mutant genes that retained the normal 3' end. While this linearity may not be a universal observation for all natural genes, it is apparent from the present study and from the study by Zaret and Sherman that the question of mRNA to protein linearity and altered translational efficiency must be addressed to make an accurate assessment of expression from a particular transgene. The results from the WAPPC3 transgene are important to suggest that high levels of protein expression can be obtained from even very low levels of RNA.

From review of the present data and comparison to the literature, we believe that the vital element of the WAPPC3 transgene is the presence of the WAP 3' flanking region. Due to the extremely low levels of RNA from the WAPPC series transgenes, it is unclear if differences in the relative rates of transcription to transcript degradation exist. We further assert that the mode of action is likely at the post-transcriptional stage and may be at the translational stage in the form of increased translational efficiency of WAPPC3 mRNA. It is not known whether the mechanism is a general regulatory mode for the WAP gene. The effect could be caused by: 1) a specific interaction of a trans-acting factor with the WAP 3' UTR of WAPPC3; 2) a specific interaction of the WAP 5' and 3' UTRs of WAPPC3, possibly mediated by a trans-acting factor; or, 3) an interaction of the WAP 3' UTR and the internal sequence of hPC, possibly mediated by a trans-acting factor. On the basis of the present study, it is not without question to implicate the 3' UTR solely, but in view of this evidence and the evidence from the literature, the 3' UTR seems a likely candidate for a positive regulatory role in WAPPC3. It is not known if this mechanism may be a general one for the WAP gene or if the observation is restricted to the WAPPC3 transgenes. Studies using other cDNAs inserted into the WAP gene with the format of WAPPC3 are currently under way and should give more insight into these mechanisms.

CONCLUSIONS

The WAPPC3 construct is an improvement over WAPPC1 and WAPPC2 constructs in terms of expression levels. Recombinant Protein C was expressed in the milk of WAPPC3 transgenic mice at up to 300 μg/ml (average 160 μg/ml), or about 30% (average 16%) of endogenous WAP expression -- higher than any WAP fusion transgene for which we have found reports. These results provide an example of an alternative regulatory format to that of using introns to obtain high expression levels. It appears that the important feature of WAPPC3 to give high expression levels is the presence of the WAP 3' UTR directly adjacent (downstream) to the Protein C coding region. The mechanism of action to give the higher expression levels appears to be at the post-transcriptional level. We speculate that the relative differences in levels of transcripts (between WAPPC3 and p238.18 in mice and WAPPC1 in pigs) and the levels of protein obtained from them may also indicate that differences in translational efficiency may exist between transcripts from different transgenes. It is not known whether the mechanism of action may be a general regulatory feature or whether it is specific to the WAP gene or the WAPPC3 transgene. Prior reports of WAP expression implicate the WAP 3' end in post-transcriptional regulation, though it is not clear if the reported regulatory mechanisms are the same as the mechanism for WAPPC3.

SUMMARY

We have compared the expression of rhPC from different WAPPC3 transgenic mouse lines (n=4) and the extremely low levels of WAPPC3 RNA in mammary tissue samples to those values obtained for other WAP-hPC transgenes. We have seen that the WAPPC3 construct was expressed at relatively high levels (up to 300 μg/ml in milk, or about 30% of endogenous WAP expression), and we attribute the effect to the presence of the WAP 3' UTR and flanking region immediately adjacent to the hPC cDNA. Also, in contrast to previous studies, all WAPPC3 females that were tested expressed hPC on their first lactations. Whereas expression from WAP-cDNA fusion transgenes in previous studies varied over several orders of magnitude, the range of expression level from WAPPC3 varied less than one order of magnitude. However, hPC expression was dramatically lower in subsequent lactations for WAPPC3 mice.

For the animals used in this study, we have shown that the ratio of secreted protein to specific transcripts is higher for WAPPC3 than for the other transgenes in question. Based on this observation, we suggest that the translational efficiency of the WAPPC3 mRNA may be higher than for the other transgenes. Further, we suggest that altered translational efficiency may exert a regulatory effect for fusion transgenes in general. A generalization of this hypothesis from the present study is limited by the following: 1) the mRNA from only one WAPPC1 mouse, two WAPPC1 pigs, and two p238.18 mice were evaluated for the present study; 2) the high expression

levels of hPC from WAPPC1 in F_0 pigs (n=2) and p238.18 (n=8) in F_0 mice may have resulted in a rate limitation at the translational level or at the transport (secretion) level, resulting in depressed estimates of hPC synthesis.

To determine if the WAPPC3 format may be useful as a general expression vector, we are currently evaluating other cDNA transgenes bounded by the WAP 5' and 3' elements as in WAPPC3. These studies should lend insight into mechanisms of WAP transgene regulation and possibly transgene regulation in general.

LITERATURE CITED

1 M. Kozak (1991) *J. Cell Biol.* **115**:887-903

2 Benjamin Lewin (1990) *GENES IV* (Oxford University Press, Oxford, England)

3 Meade, H., L. Gates, E. Lacy, and N. Lonberg (1990) *Bio/Technology* **8**:443-446

4 Reid, L. H., E. G. Shesely, H. -S. Kim, and O. Smithies (1991) *Mol. Cell. Biol.* **11**:2769-2777

5 Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S Kucherlapati (1985) *Nature* **317**:230-234

6 Behringer, R. R., R.E. Hammer, R. L. Brinster, R. D. Palmiter, and T. M. Townes (1987) *Proc. Natl. Acad. Sci. USA* **84**:7056-7060

7 Bode, J., Y. Kohwi, L. Dickinson, T. Joh, D. Klehr, C. Mielke, and T. Kohwi-Shigematsu (1992) *Science* **255**:195-197

8 Albitar, M., M. Katsumata, and S. A. Liebhaber (1991) *Mol. Cell. Biol.* **11**:3786-3794

9 Sternberg, N. (1990) *Proc. Natl. Acad. Sci. USA* **87**:103-107

10 Forget, B. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**:7909-7911

11 Hohn, B. and J. Collins (1980) *Gene* **11**:291-298

12 Grosveld, F. G., T. Lund, E. J. Murray, A. L. Mellor, H. H. M. Dahl, and R. A. Flavell (1982) *Nuc. Acids. Res.* **10**:6715-6732

13 Brinster, R. L., R. E. Braun, D. Lo, M. R. Avarbock, F. Oram, and R. D. Palmiter (1989) *Proc. Natl. Acad. Sci. USA* **86**:7087-7091

14 Pieper, F. R., I. C. M. deWit, A. C. J. Pronk, P. M. Kooiman, R. Strijker, P. J. A. Krimpenfort, J. H. Nuyens, and H. A. de Boer (1992) *Nuc. Acids. Res.* **20**:1259-1264

15 Wall, R. J., V. G. Pursel, A. Shamay, R. A. McKnight, C. W. Pittius, and L. Hennighausen (1991) *Proc. Natl. Acad. Sci. USA* **88**:1696-1700

16 Wright, G., A. Carver, D. Cottom, D. Reeves, A. Scott, P. Simons, I. Wilmut, I. Garner, and A. Colman (1991) *Bio/Technology* **9**:830-834

17 Palmiter, R. D. and R. L. Brinster (1986) *Ann. Rev. Gen.* **20**:465-499

18 Brinster, R. L., J. M. Allen, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter (1988) *Proc. Natl. Acad. Sci. USA* **85**:836-840

19 Palmiter, R. D., E. P. Sandgren, M. R. Avarbock, D. D. Allen, and R. L. Brinster (1991) *Proc. Natl. Acad. Sci. USA* **88**:478-482

20 Velander, W. H., J. L. Johnson, R. L. Page, C. G. Russell, A. Subramanian, T. D. Wilkins, F. C. Gwazdauskas, C. Pittius, and W. N. Drohan (1992) *Proc. Natl. Acad. Sci. USA* **89**:12003-12007

21 Chee, P. P., R. C. Klassy, and J. L. Slightom (1986) *Gene* **41**:47-57

22 Gross, M. K., M. S. Kainz, and G. F. Merrill (1987) *Mol. Cell. Biol* **7**:4576-4581

23 Clark, A. J., H. Bessos, J. O. Bishop, P. Brown, S. Harris, R. Lathe, M. McClenaghan, C. Prowse, J. P. Simons, C. B. A. Whitelaw, and I. Wilmut (1989) *Bio/Technology* **7**:487-492

24 Ebert, K. M., J. P. Selgrath, P. DiTullio, J. Denman, T. E. Smith, M. A. Memon, J. E. Schindler, G. M. Monastersky, J. A. Vitale, and K. Gordon (1991) *Bio/Technology* **9**:835-838

25 Raghow, R. (1987) *Trends Biochem. Sci.* **12**:358-360

26 Singer, R. and S. Penman (1973) *J. Mol. Biol.* **78**:321-334

27 Berger, S. L. and H. L. Cooper (1975) *Proc. Natl. Acad. Sci. USA* **72**:3873-3877

28 Greaves, D. R., F. D. Wilson, G. Lang, and D. Kioussis (1989) *Cell* **56**:979-986

29 Edelson, E. (1990) *Mosaic* **21**:2-9

30 Singh, H., R. Sen, D. Baltimore, and P. A. Sharp (1986) *Nature* **319**:154-158

31 Grosschedl, R. and D. Baltimore (1985) *Cell* **41**:885-897

32 Bornstein, P., J. McKay, D. J. Liska, S. Apone, and S. Devarayalu (1988) *Mol. Cell. Biol.* **8**:4851-4857

33 Nevins, J. R. and J. E. Darnell (1978) *Cell* **15**:1477-1493

34 Birnstiel, M. L., M. Busslinger, and K, Strub (1985) *Cell* **41**:349-359

35 Mason, P. J., J. A. Elkington, M. M. Lloyd, M. B. Jones, and J. G. Williams (1986) *Cell* **46**:263-270

36 Brawerman, G. (1981) *CRC Crit. Rev. in Biochem.* **10**:1-38

37 Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp (1986) *Ann. Rev. Biochem.* **55**:1119-1150

38 Green, M. R. (1989) *Current Opinion in Cell Biol.* **1**:519-525

39 Merrill, G. F., S. D. Hauschka, and S. L. McKnight (1984) *Mol. Cell. Biol.* **4**:1777-1784

40 Shaw, G. and R. Kamen (1986) *Cell* **46**:659-657

41 Wilson, T. and R. Treisman (1988) *Nature* **336**:396-399

42 Kaufman, R. J. and P. A. Sharp (1983) *Mol. Cell. Biol.* **3**:1598-1608

43 Brewer, G. and J. Ross (1988) *Mol. Cell. Biol.* **8**:1697-1708

44 Stoeckle, M. Y, and H. Hanafusa (1989) *Mol. Cell. Biol.* **9**:4738-4745

45 Marzluff, W. F., and N. B. Pandey (1988) *Trends Biochem. Sci.* **13**:49-52

46 Kozak, M. (1987) *Nuc. Acids. Res.* **15**:8125-8148

47 Kozak, M. (1992) *Crit. Rev. in Biochem. and Mol. Biol.* **27**:385-402

48 Zaret, K. S. and F. Sherman (1984) *J. Mol. Biol.* **176**:107-135

49 Hennighausen, L. and A. E. Sippel (1982) *Eur. J. Biochem.* **125**:131-141

50 Eisenstein, R. S. and J. M. Rosen (1988) *Mol. Cell. Biol.* **8**:3183-3190

51 Gordon, K., J. Vitale, B. Roberts, G. Monastersky, P. DiTullio, and G. Moore (1990) in *Transgenic Models in Medicine and Agriculture*, 55-59 (Wiley-Liss, Inc.)

52 Bayna, E. M. and J. M. Rosen (1990) *Nuc. Acids. Res.* **18**:2977-2984

53 Greenberg, N. M., J. W. Anderson, A. J. W. Hsueh, K. Nishimori, J. J. Reeves, D. M. DeAvila, D. N. Ward, and J. M. Rosen (1991) *Proc. Natl. Acad. Sci. USA* **88**:8327-8331

54 Krimpenfort, P., A. Rademakers, W. Eyestone, A. van der Schans, S. van den Broek, P. Kooiman, E. Kootwijk, G. Platenburg, F. Pieper, R. Strijker, and H. de Boer (1991) *Bio/Technology* **9**:844-847

55 Guyette, W. A., R. J. Matusik, and J. M. Rosen (1979) *Cell* **17**:1013-1023

56 Schoenenberger, C.-A., A. Zuk, B. Groner, W. Jones, and A.-C. Andres (1990) *Dev. Biol.* **139**:327-337

57 Rosen, J. M., E. Bayna, and K.-F. Lee (1989) *Mol. Biol. Med.* **6**:501-509

58 Quirk, S. J., J. A. Slattery, and J. W. Funder (1989) *Mol. Cell. Endocrin.* **66**:135-141

59 Burdon, T., L. Sankaran, R. J. Wall, M. Spencer, and L. Hennighausen (1991) *J. Biol. Chem.* **266**:6909-6914

60 Dale, T. C., M. J. Krnacik, C. Schmidhauser, C. L.-Q. Yang, M. J. Bissell, and J. M. Rosen (1992) *Mol. Cell. Biol.* **12**:905-914

61 Velander, W. H., R. L. Page, T. Morcol, C. G. Russell, R. Canseco, J. M. Young, W. N. Drohan, F. C. Gwazdauskas, T. D. Wilkins, and J. L. Johnson (1992) *Ann. New York Acad. Sci.* **665**:391-403

62 Pittius, C. W., L. Hennighausen, E. Lee, H. Westphal, E. Nicols, J. Vitale,

and K. Gordon (1988) *Proc. Natl. Acad. Sci. USA* **85**:5874-5878

63 Gordon, K., E. Lee, J. A. Vitale, A. E. Smith, H. Westphal, and L. Hennighausen (1987) *Bio/Technology* **5**:1183-1186

64 Pittius, C. W., L. Sankaran, and L. Hennighausen (1990) in *Transgenic Models in Medicine and Agriculture*, 47-54 (Wiley-Liss, Inc.)

65 Bischoff, R., E. Degryse, F. Perraud, W. Delmans, D. Ali-Hadji, D. Thepot, E. Devinoy, L.-M. Houdebine, and A. Pavirani (1992) *Fed. of Eur. Biochem. Soc.* **305**:265-268

66 Esmon, C.T. (1987) *Science* **235**:1348-1352

67 W.J. Dower, J.F. Miller, and C.W. Ragsdale (1988) *Nuc. Acids. Res.* **16**:6127-6145

68 W.J. Dower (1988) *Bio-Rad Laboratories Molecular Biology Reports* No. 6

69 Sambrook, J., E. R. Fritsch, and T. Maniatis (1989) *Molecular Cloning: A laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.)

70 Morelle, G. (1989) *Bethesda Research Laboratories Focus* **11**:1

71 Birnboim, H. C. and J. Doly (1979) *Nuc. Acids Res.* **7**:1513-1523

72 Fink, P. S. (1991) *BioTechniques* **10**:446-450

73 Hogan, B., G. Costantini, E. Lacy (1986) *Manipulation of the Mouse Embryo* (Cold Spring Harbor Laboratory, New York)

74 Brinster, R. L., H. Y. Chen, N. E. Drumbeater, M. K. Yagle, and R. D. Palmiter (1985) *Proc. Natl. Acad. Sci. USA* **82**:4438-4442

75 Johnson, J. L. (1991) in *Nucleic Acid Techniques in Bacterial Systematics*, 6-8, E. Stackebrandt and M. Goodfellow, editors (John Wiley & Sons, Ltd., West Sussex, England)

76 J. Marmur (1961) *J. Mol. Biol.* **3**:208-218

77 John, S. W. M., G. Weitzner, R. Rozen, and C. R. Scriver (1990) *Nuc. Acids Res.* **19**:408

78 Saiki, R. K., P. S. Walsh, C. H. Levenson, and H. A. Ehrlich (1989) *Proc. Natl. Acad. Sci.* **86**:6230-6234

79 Ehtesham, N. Z. and S.E. Hasnain (1991) *BioTechniques* **11**:718-721

80 Puissant, C. and L. -M. Houdebine (1990) *Biotechniques* **8**:148-149

81 Celano, P., P. M. Vertino, and R. A. Casero, Jr. (1993) *BioTechniques* **15**: 27-28

82 Kroczek, R. A. and E. Siebert (1990) *Anal. Biochem.* **184**:90-95

83 Ahmad, I., J. A. Finkelstein, and A. W. Steggles (1990) *BioTechniques* **8**:162-165

84 Chaudhari, N. (1991) *BioTechniques* **10**:601-605

85 *Promega Notes* No.6, (November 1986)

86 Laemmli, U. K. (1970) *Nature* **227**:680-685

87 Towbin, H., T. Staehelin, and J. Gordon (1979) *Proc. Natl. Acad. Sci. USA* **76**:4350-4354

88 Raymond L. Page, (1993) Ph.D. Dissertation, Virginia Polytechnic Institute and State University

89 Hennighausen, L. (1990) *Protein Expression and Purification* **1**:1-6

90 Weiss, E. A., G. M. Gilmartin, and J. R. Nevins (1991) *EMBO J.* **10**:215-219

91 Foster, D. C., S. Yoshitake, and E. W. Davie (1985) *Proc. Natl. Acad. Sci. USA* **82**:4673-4677

92 Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias (1987) *Cell* **51**:975-985

93 Hennighausen, L. G., A. E. Sippel, A. A. Hobbs, and J. M. Rosen (1982)

Nuc. Acids Res. **10**:3733-3744

94 Chen, L.-H., and M. J. Bissell (1989) *Cell Regul.* **1**:45-54

95 Grinnell, B. W., J. D. Walls, B. Gerlitz, D. T. Berg, D. B. McClure, H. Ehrlich, N. U. Bang, and S. B. Yan (1990) in *Protein C and Related Anticoagulants*, 29-63, D. F. Bruley and W. N. Drohan, editors (The Portfolio, The Woodlands, TX)

96 Wydro, R. M. (1990) "Transgenic Production of Protein C", in *Abstracts* from *Protein C and Related Anticoagulants*, ref 95.

PRESENT AND FUTURE WORK

Present Work

Since the WAPPC3 transgene appeared to express at fairly high levels, it was used to develop a general "cassette" vector for expression from the WAP gene promoter. Three bases of the six nucleotide region (GCAGCC) immediately downstream of the WAP stop codon (TGA) were mutated to form a Kpn I restriction site (GGTACC). The new WAP vector, designated WAP4, was then cloned into pUC18 and designated pUCWAP4 for the molecule including the cloning vector (Figure 24). This WAP vector allows the insertion of virtually any sequence bounded by Kpn I sites. Thus, the strategy is to amplify a coding region of interest by PCR, clone it into pUC and sequence to check for errors, then digest the pUC/insert with Kpn I (partial digestion if the fragment of interest contains internal Kpn I sites) and ligate the insert into the pUCWAP4 vector. At this point, one needs only to verify the orientation of the insert and check the sequence at the Kpn I junctions. The transgene can then be removed from the cloning vector by restriction digestion with Eco RI enzyme and subsequent purification.

A further modification to the WAP4 vector was the addition of Not I restriction sites in place of the Eco RI sites at the ends of the WAP4 vector. The new vector was designated WAP5 and is shown in Figure 24 (pUCWAP5 with the cloning vector). This modification allows more flexibility in inserting coding regions that

contain an Eco RI site(s). Thus, one can insert a cDNA or genomic coding region that contains an Eco RI site(s) (assuming Not I sites are not in the insert) and then remove the transgene from the cloning vector by restriction digestion with Not I enzyme (and subsequent purification).

Both of the above expression vectors, WAP4 and WAP5, have been used for cDNA inserts. Transgenic mice have been generated with these transgenes; however, the initial results are as yet inconclusive. Transgenic pigs are also being made with these constructs.

Future Work

The "motif" of the WAPPC3 transgene can be described as a coding region inserted between matching and complete 5' (promoter) and 3' flanking regions. Although it is not clear whether the transgene motif used for WAPPC3 might be one which is useful for many target proteins and many promoters, this study and the literature reviewed herein indicate that the WAPPC3 concept may be valuable for optimizing expression from a transgene. Due to the possibility of specific interactions of introns with a promoter or other intragenic region, the preferred transgene construct would also have the introns of the "driving" gene (e.g. WAP introns for WAPPC3) incorporated into the transgene. Further, due to potential position and orientation effects of introns, the introns should be inserted in the most natural position possible within the coding region of the gene to be expressed. The optimum

locations for the introns (in a fusion gene) are not known for any particular driving gene. Also, it is not known whether all of the introns are necessary for a particular driving gene. The goals of the research outlined herein are to delineate the cis- and trans- regulatory interactions of a particular gene and to apply those concepts to the development of transgenes that give desired regulation and expression levels.

TABLE 1

EXPRESSION LEVELS FROM WAP-PROTEIN C TRANSGENES IN MICE

	Number of Independent Lines Tested	Range of Expression Level μg/ml	Average Expression Level μg/ml
WAPPC1	10	0.01-10	3
WAPPC2	5	0.01-5	1
WAPPC3, G_0	11	30-80	39
WAPPC3, F_1, F_2	4	75-299	162
p238.18 (genomic hPC)	9	300-10,000	1000

TABLE 2

EXPRESSION LEVELS OF RECOMBINANT HUMAN PROTEIN C IN THE MILK OF WAPPC3 FOUNDER MICE

FOUNDER MICE		FIRST LACTATION	SECOND LACTATION
Mouse #	Day of Lactation	rhPC Expression Level, μg/ml	rhPC Expression Level, μg/ml
1	6	37	4.4
	12	40	
	16	44	
2	6	62	3.4
	16	18	
3	6	49	2.0
	12	58	
	16	45	
12	7	33	5.8
	12	28	
22	6	33	17
	12	39	
	16	32	

TABLE 3

EXPRESSION LEVELS OF RECOMBINANT HUMAN PROTEIN C IN THE MILK OF WAPPC3 F_1 AND F_2 MICE

F_1 AND F_2 MICE		FIRST LACTATION	SECOND LACTATION
Mouse #	Day of Lactation	rhPC Expression Level, μg/ml	rhPC Expression Level, μg/ml
7-3-2	3	181	None Detected
	9	255	
	13	86	
21-13	4	73	Not Tested
	10	142	
	14	114	
23-16	5	105	None Detected
	11	169	
R46-2-4	3	115	0.3
	11	230	
	15	94	
7-3-9	10	135	NOT TESTED: KILLED FOR RNA ON 1ST LACTATION
21-14	11	174	
23-15	10	299	
R46-2-2	10	250	

TABLE 4

Apparent Sizes of PolyA Tails
on WAP-Protein C Transgene mRNAs

	Predicted * mRNA size, nt (polyA-)	Observed mRNA size, nt	Deduced PolyA size, nt
WAPPC1	1508	~1600	~100
WAPPC3	1545	~1700	~150
p238.18	1511 or 1737	~1700	~200 or 0
WAP	561	600 to 800	~50 to 250

* Size was predicted by the sum of:
length of published or observed 5' UTR
length of linker regions (if present)
length of coding region
length of published or predicted 3' UTR

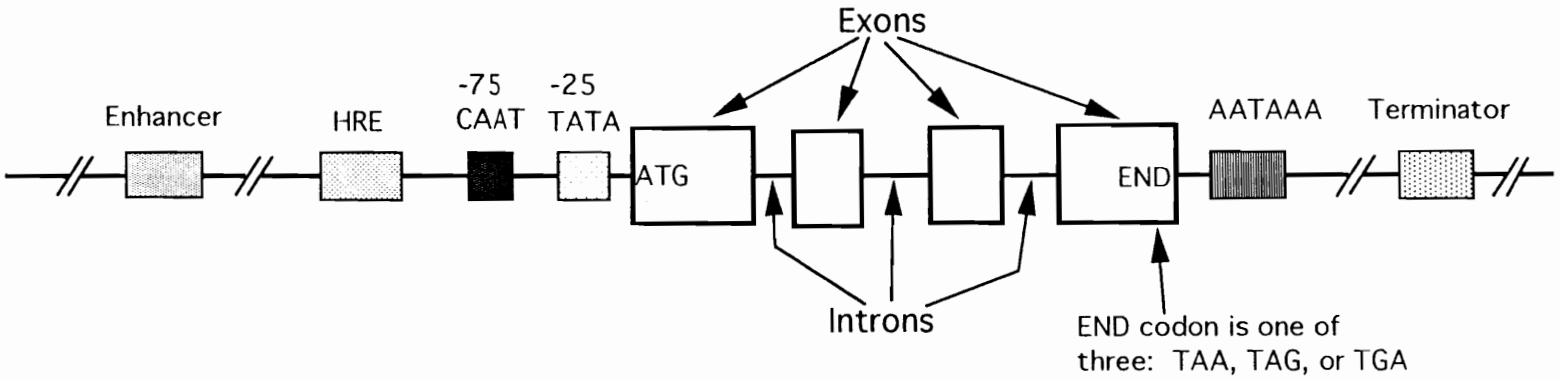


Figure 1
Eukaryotic Gene

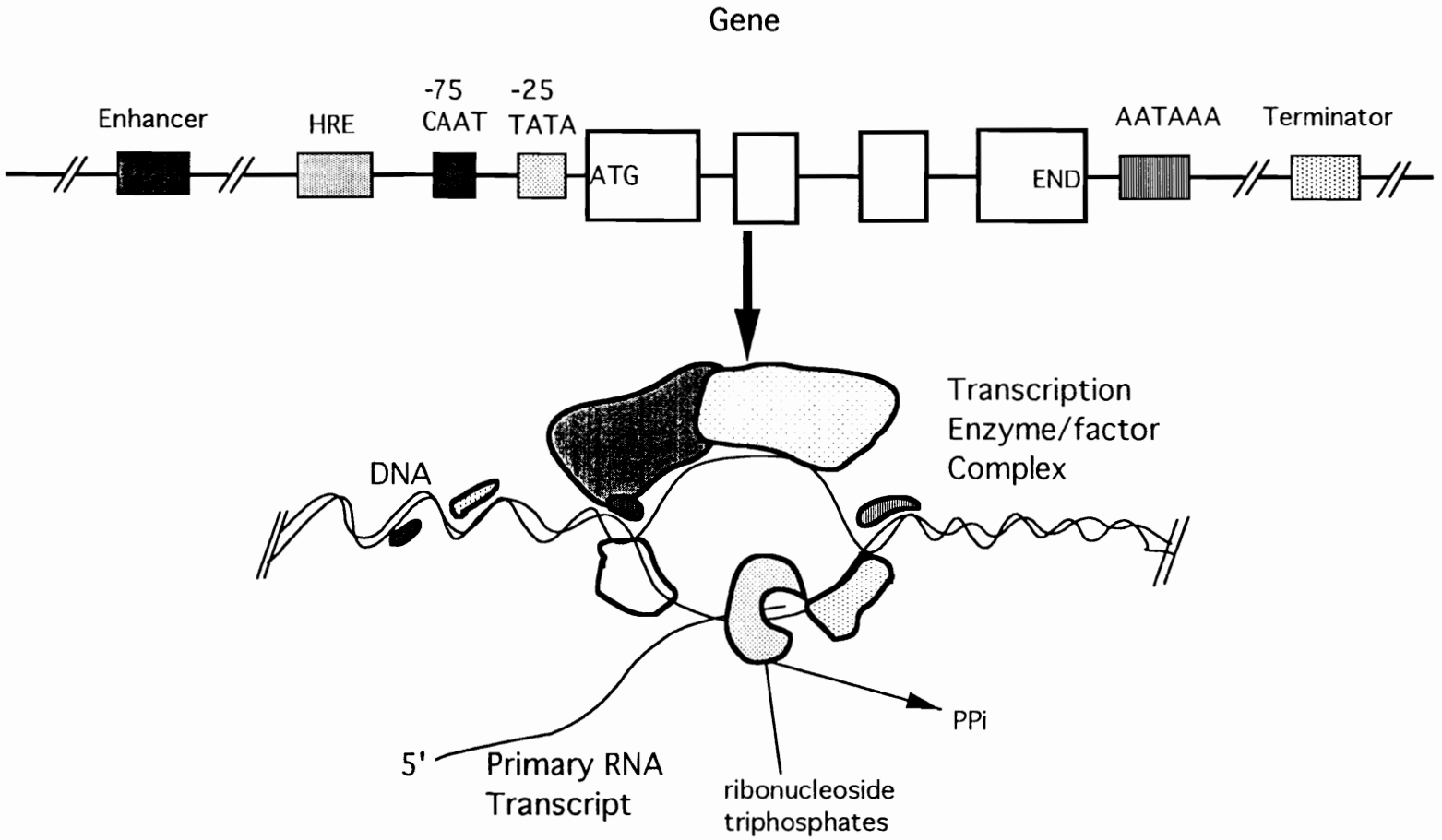


Figure 2
Transcription Reaction with Gene

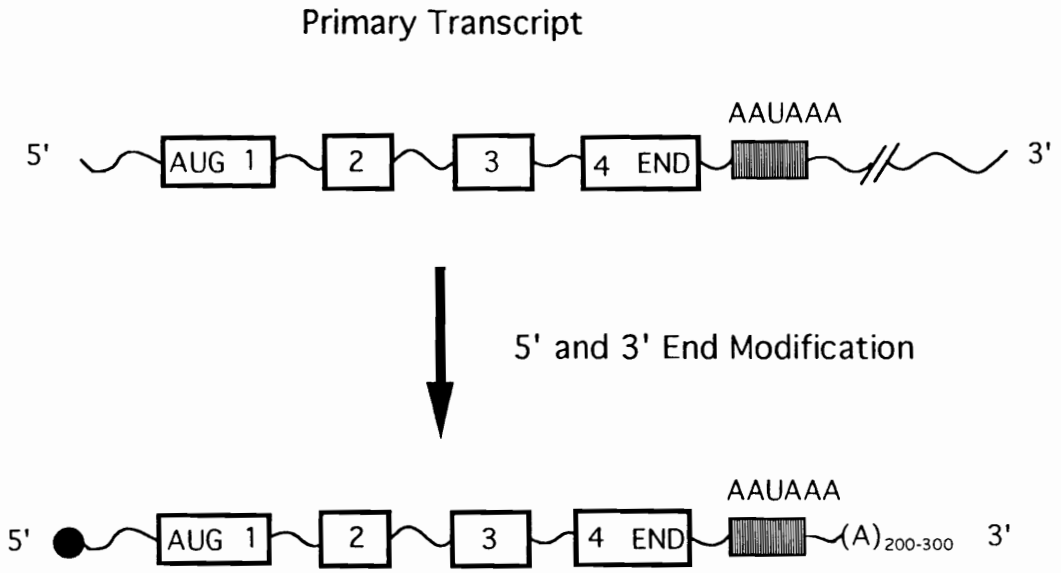


Figure 3
5' and 3' End Modification of RNA

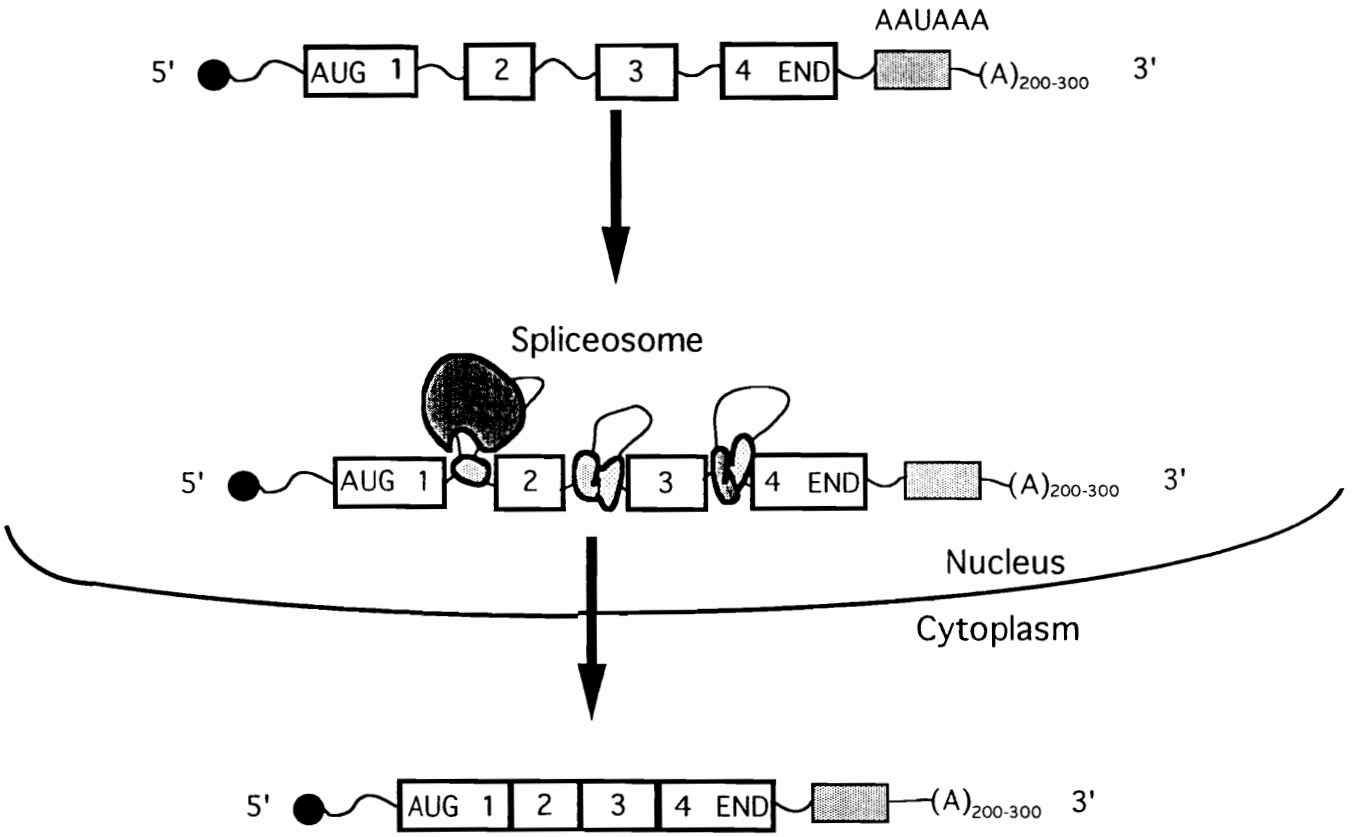


Figure 4
Splicing Reaction and mRNA Transport

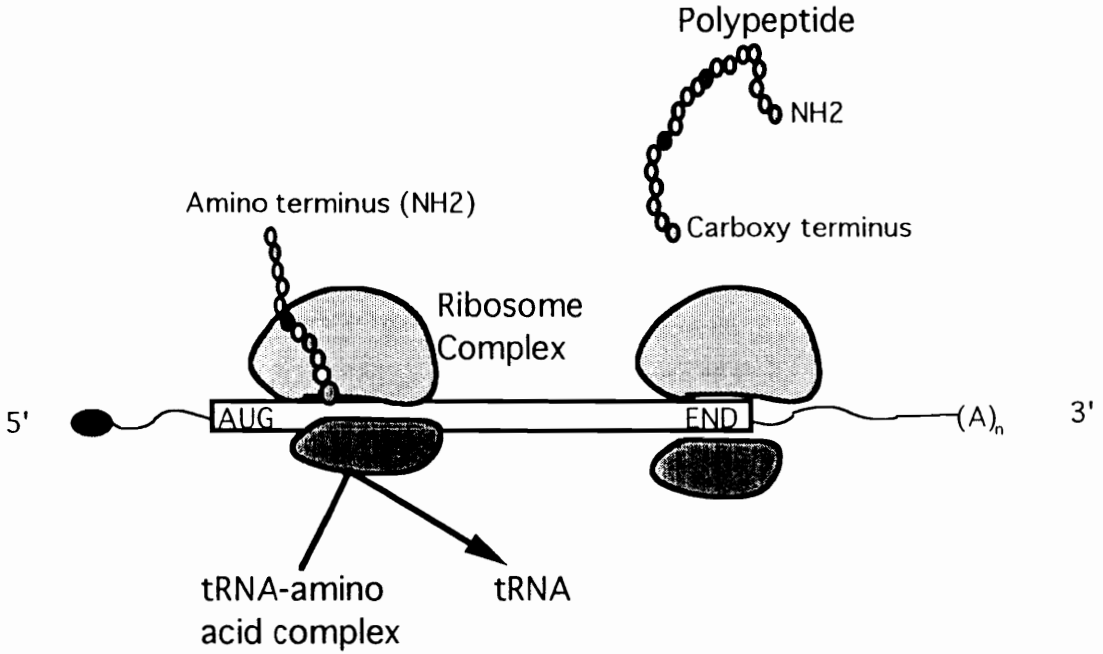


Figure 5
Translation Reactions with mRNA
into Polypeptide Product

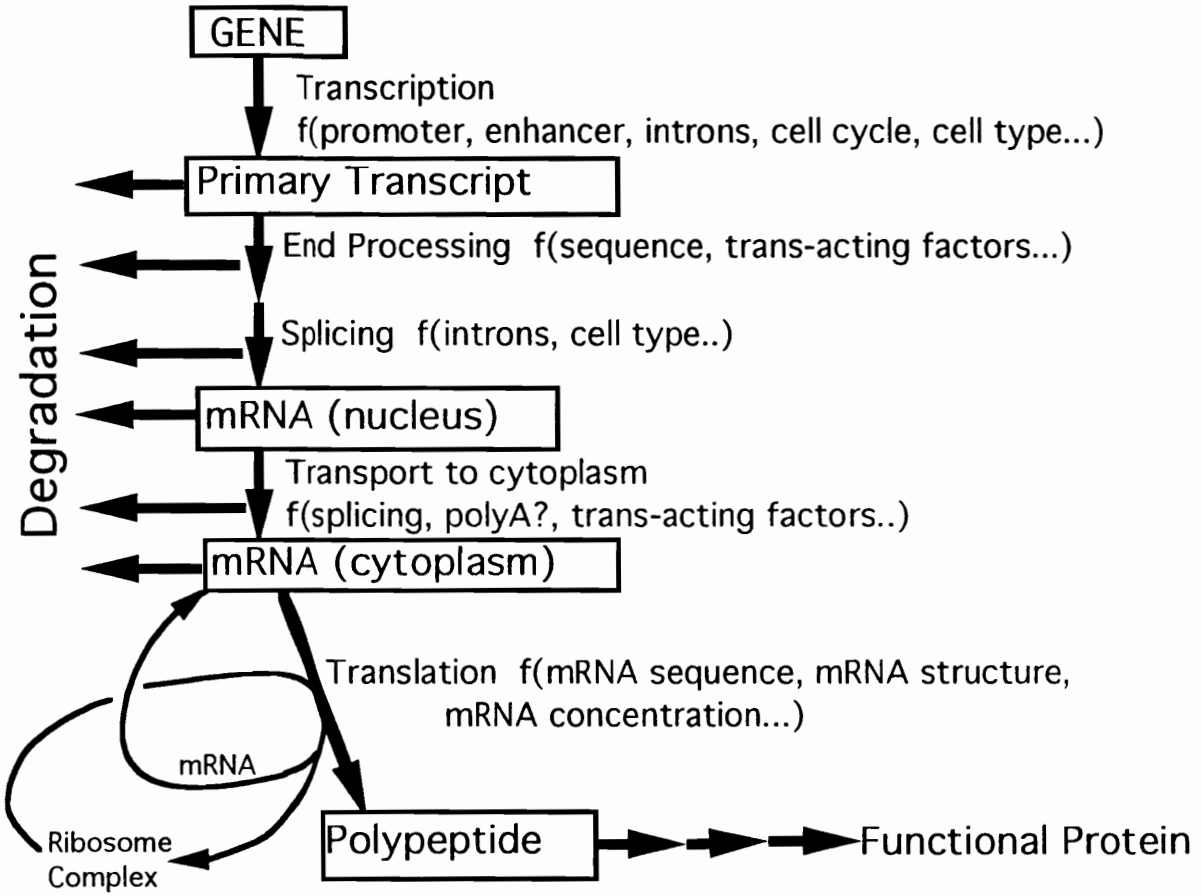
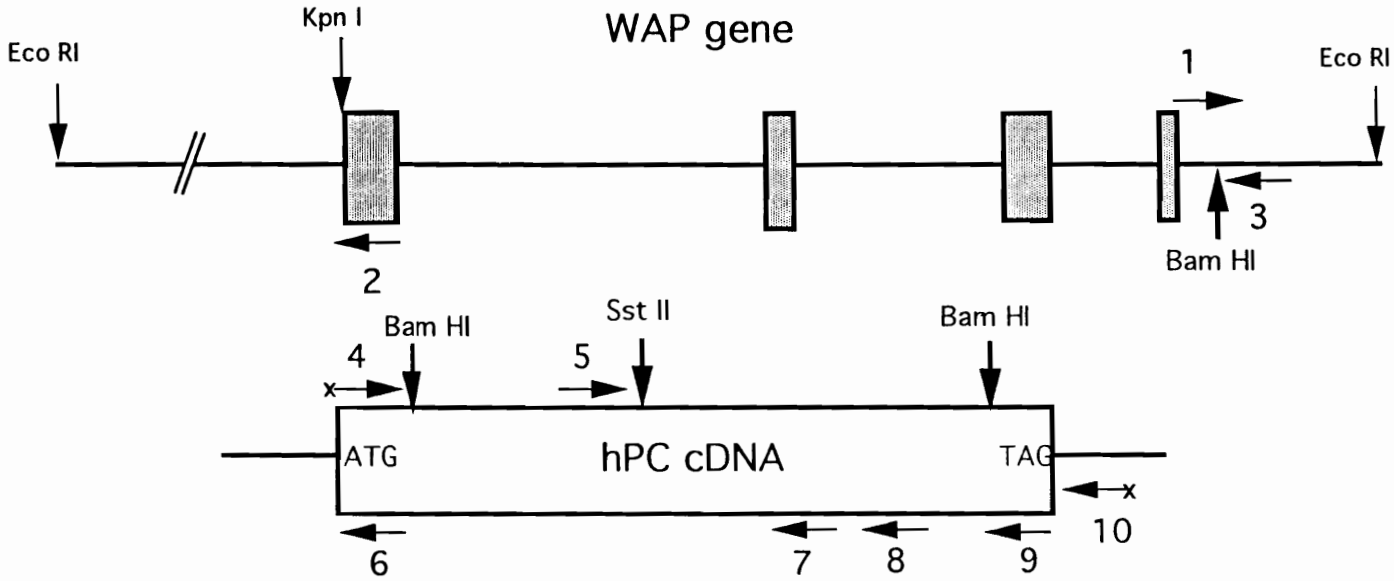


Figure 6
Gene Expression Pathway



Primers -- Arrow on top of gene (left to right) = sense
Arrow on bottom (right to left) = antisense

1 -- WAP 3' A1	6 -- ProCA8
2 -- WAPA1	7 -- ProCA5
3 -- WAP 3' S1	8 -- ProCA2
4 -- hPC5' KpnI	9 -- ProCA4
5 -- ProCS2	10 -- hPC3' KpnI

Figure 7
Locations of Oligonucleotide Primers

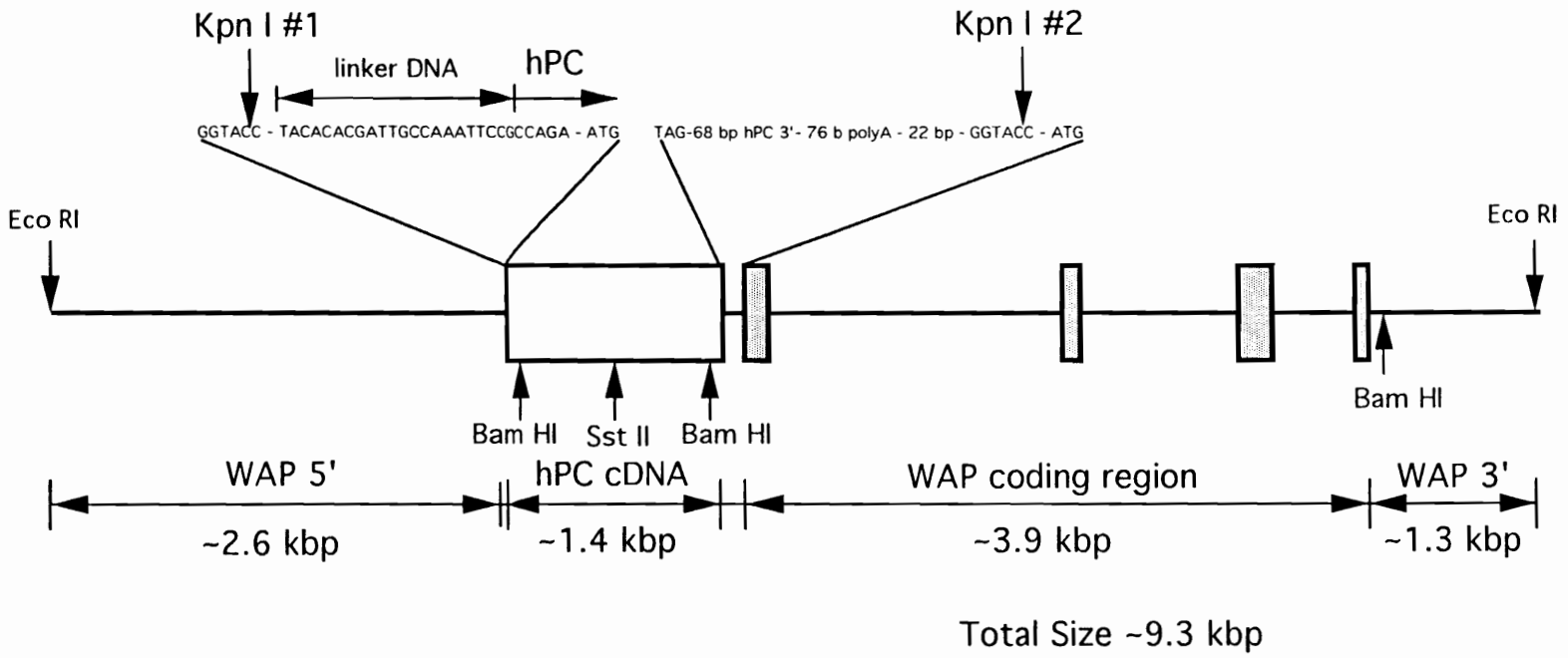


Figure 8
WAPPC1 Transgene Construct

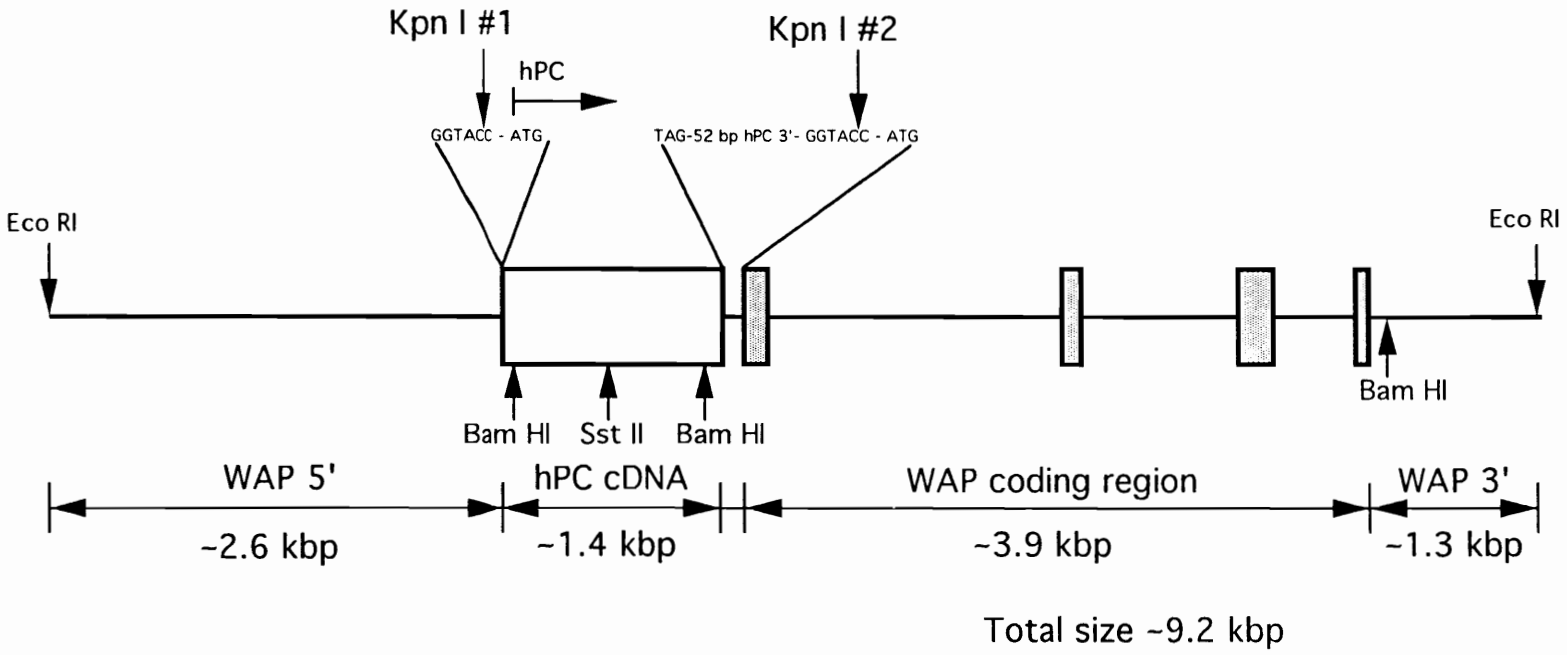


Figure 9
WAPPC2 Transgene Construct

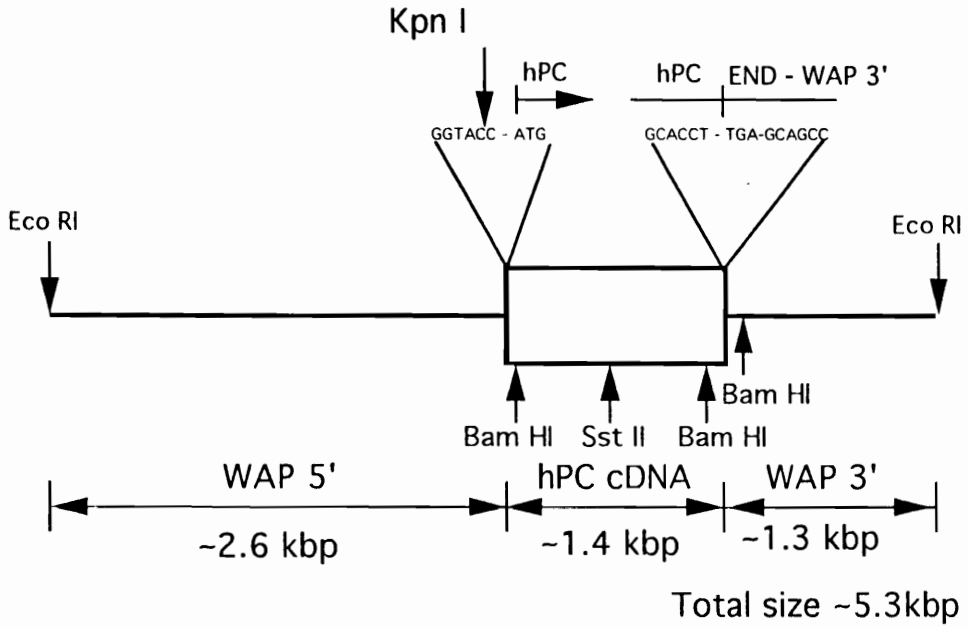


Figure 10
WAPPC3 Transgene Construct

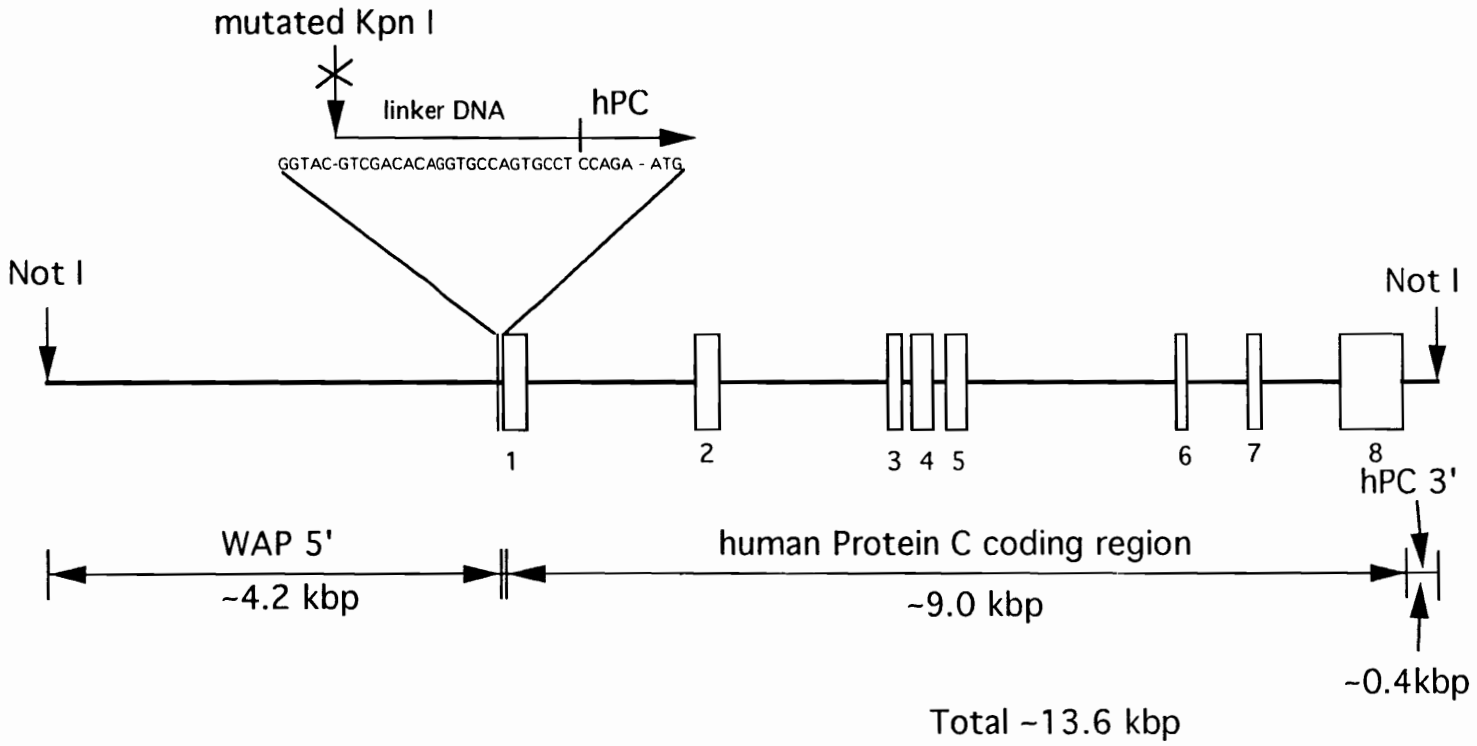


Figure 11
p238.18 WAP-genomic hPC Transgene Construct

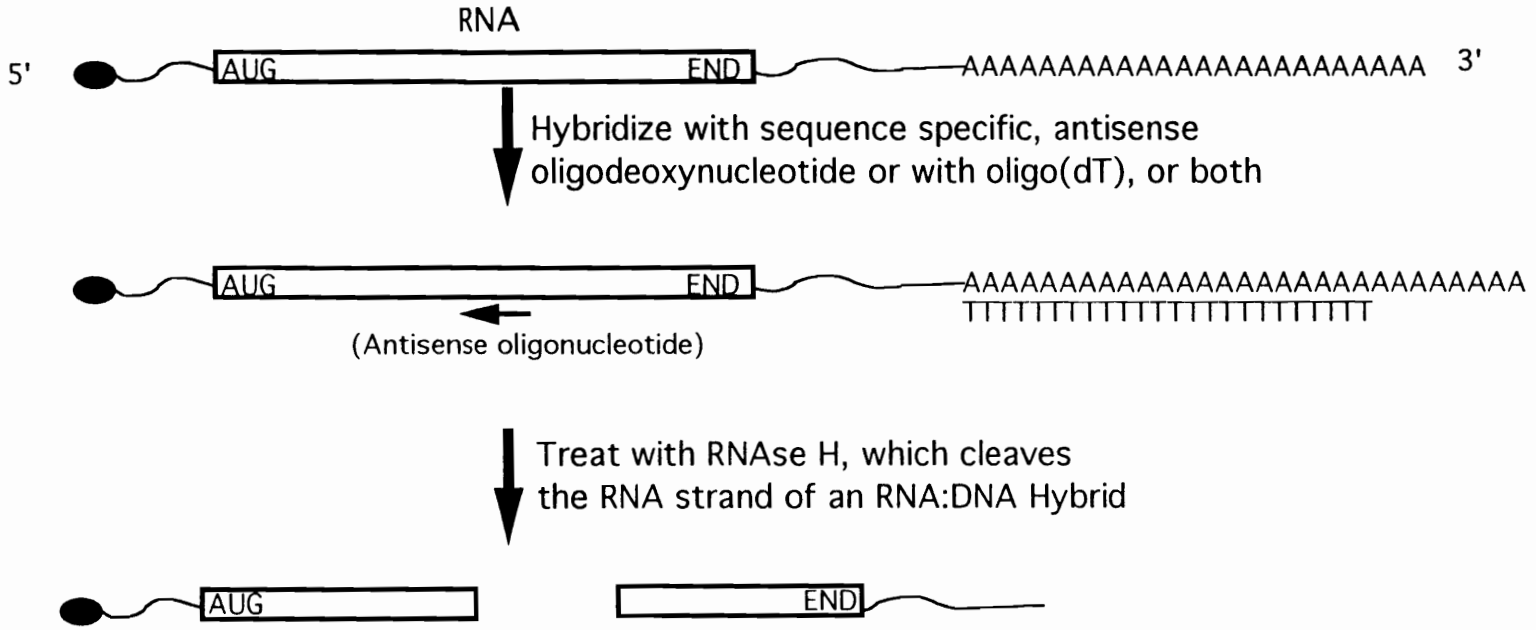


Figure 12
RNAse H Strategy

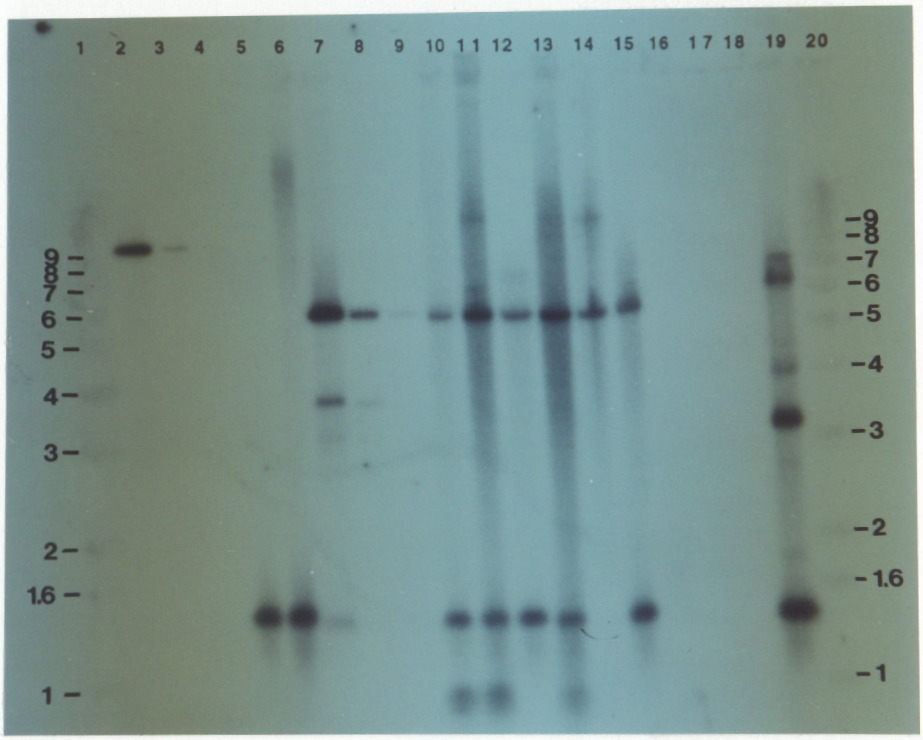


Figure 13
DNA In-Gel Hybridization

10 μg each mouse Mammary Gland DNA hybridized overnight as in **Materials and Methods**, with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 20 h at room temperature. Lanes 1 and 20, 1 kb ladder; Lanes 2-5, 250, 25, and 5 pg of WAPPC1 DNA, respectively; Lane 5, control mouse; Lane 6, WAPPC1 mouse F_6 #1; Lanes 7-9, 250, 25, and 5 pg of WAPPC3 DNA, respectively; Lanes 10-15, WAPPC3 mice 7-3-9, 21-14, 23-15, R46-2-2, 2-10, and 23-16, respectively; Lane 19, p238.18 mouse #1.

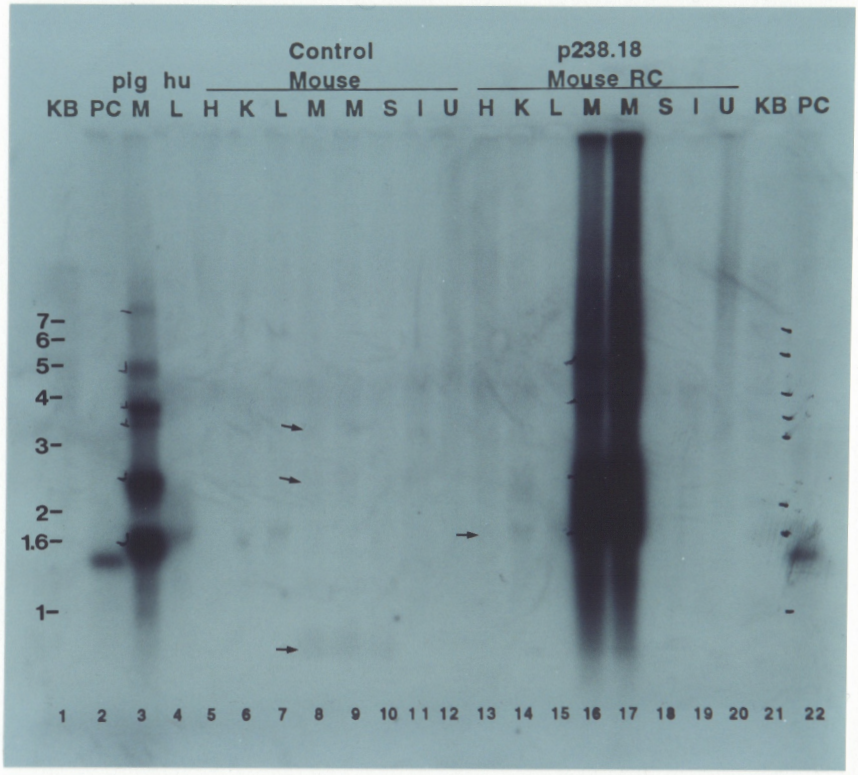


Figure 14 A
RNA In-Gel Hybridization -- RNA Gel 1

Control Mouse and p238.18 Transgenic Mouse

10 μg total RNA was used for each RNA lane. Gel was hybridized overnight as in **Materials and Methods** with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 20 h at room temperature. Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, Control Mouse RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8 and 9, Mammary Gland; 10, Salivary Gland; 11, Small Intestine; 12, Uterus; Lanes 13-20, p238.18 Mouse #1 RNA, mid-lactation, from the following tissues: 13, Heart; 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Small Intestine; 20, Uterus.

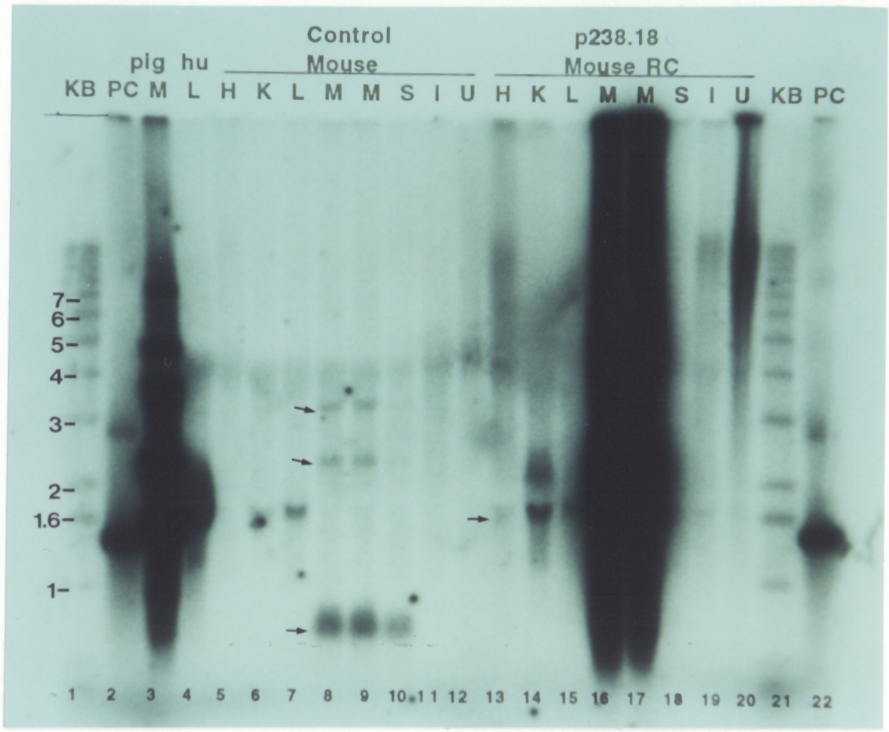


Figure 14 B
RNA In-Gel Hybridization -- RNA Gel 1

Control Mouse and p238.18 Transgenic Mouse

Same gel and conditions as in Figure 14 A, except the exposure of x-ray film was about 84 h at -70°C with an intensifying screen (~30 times more exposure than in A). Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, Control Mouse RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8 and 9, Mammary Gland; 10, Salivary Gland; 11, Small Intestine; 12, Uterus. Lanes 13-20, p238.18 Mouse #1 RNA, mid-lactation, from the following tissues: 13, Heart; 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Small Intestine; 20, Uterus.

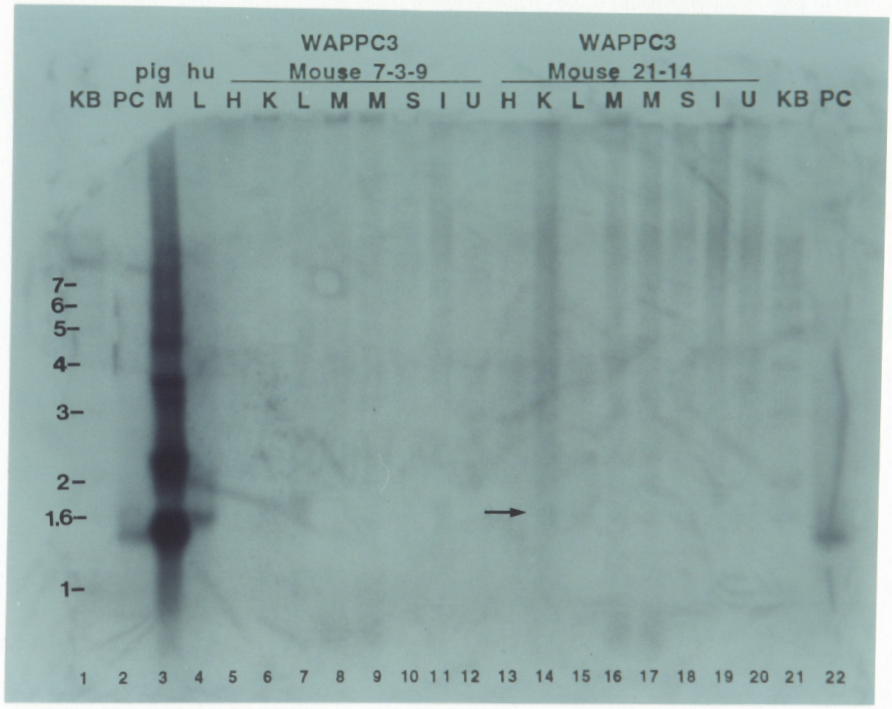


Figure 15 A
RNA In-Gel Hybridization -- RNA Gel 2

WAPPC3 Mice, 7-3-9 and 21-14

10 μg total RNA was used for each RNA lane. Gel was hybridized overnight as in **Materials and Methods** with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 20 h at room temperature. Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, WAPPC3 Mouse 7-3-9 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8 and 9, Mammary Gland; 10, Salivary Gland; 11, Small Intestine; 12, Uterus. Lanes 13-20, WAPPC3 Mouse 21-14 RNA, mid-lactation, from the following tissues: 13, Heart; 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Small Intestine; 20, Uterus.

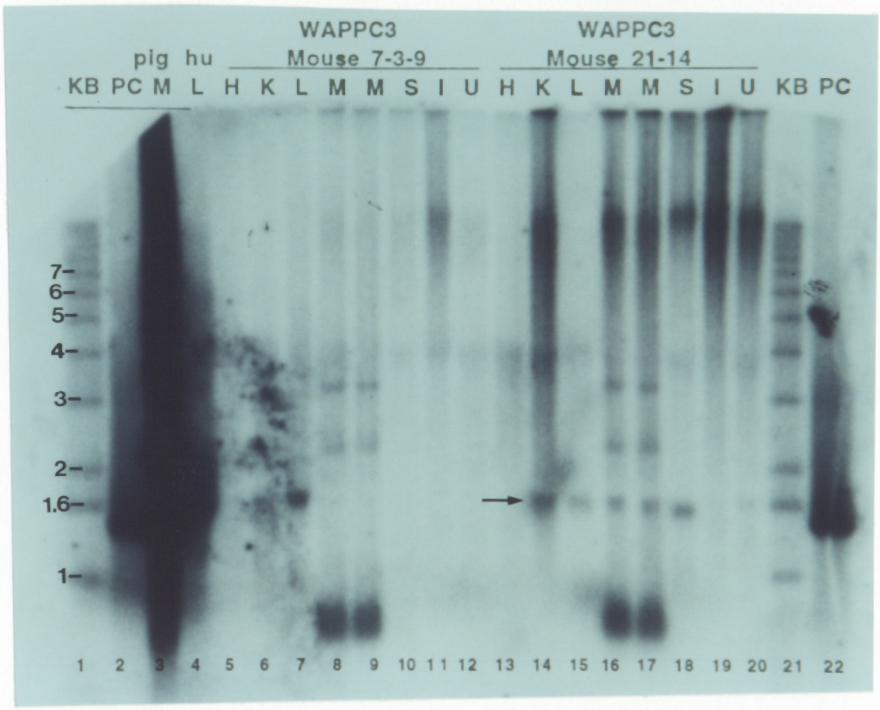


Figure 15 B
RNA In-Gel Hybridization -- RNA Gel 2

WAPPC3 Mice, 7-3-9 and 21-14

Same gel and conditions as in Figure 15 A, except the exposure of x-ray film was about 84 h at -70°C with an intensifying screen (~30 times more exposure than in A). Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, WAPPC3 Mouse 7-3-9 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8 and 9, Mammary Gland; 10, Salivary Gland; 11, Small Intestine; 12, Uterus. Lanes 13-20, WAPPC3 Mouse 21-14 RNA, mid-lactation, from the following tissues: 13, Heart; 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Small Intestine; 20, Uterus.

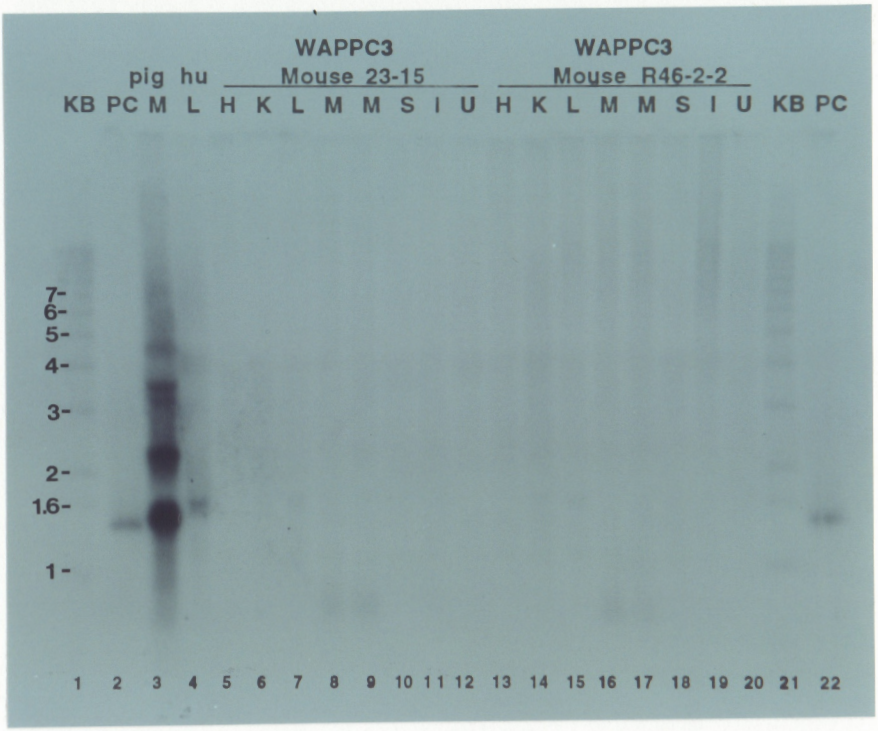


Figure 16 A
RNA In-Gel Hybridization -- RNA Gel 3

WAPPC3 Mice, 23-15 and R46-2-2

10 μg total RNA was used for each RNA lane. Gel was hybridized overnight as in **Materials and Methods** with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 20 h at room temperature. Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA. Lanes 5-12, WAPPC3 Mouse 23-15 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8 and 9, Mammary Gland; 10, Salivary Gland; 11, Small Intestine; 12, Uterus; Lanes 13-20, WAPPC3 Mouse R46-2-2 RNA, mid-lactation, from the following tissues: 13, Heart; 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Small Intestine; 20, Uterus.

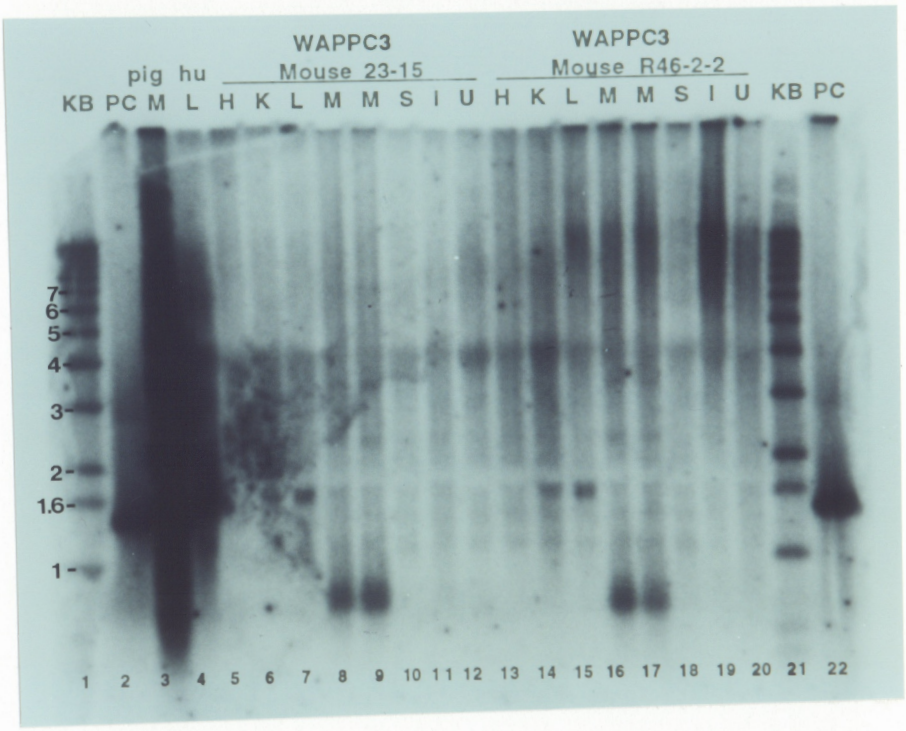


Figure 16 B
RNA In-Gel Hybridization -- RNA Gel 3

WAPPC3 Mice, 23-15 and R46-2-2

Same gel and conditions as in Figure 16 A, except the exposure of x-ray film was about 84 h at -70°C with an intensifying screen (~30 times more exposure than in A). Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, WAPPC3 Mouse 23-15 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8 and 9, Mammary Gland; 10, Salivary Gland; 11, Small Intestine; 12, Uterus. Lanes 13-20, WAPPC3 Mouse R46-2-2 RNA, mid-lactation, from the following tissues: 13, Heart; 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Small Intestine; 20, Uterus.

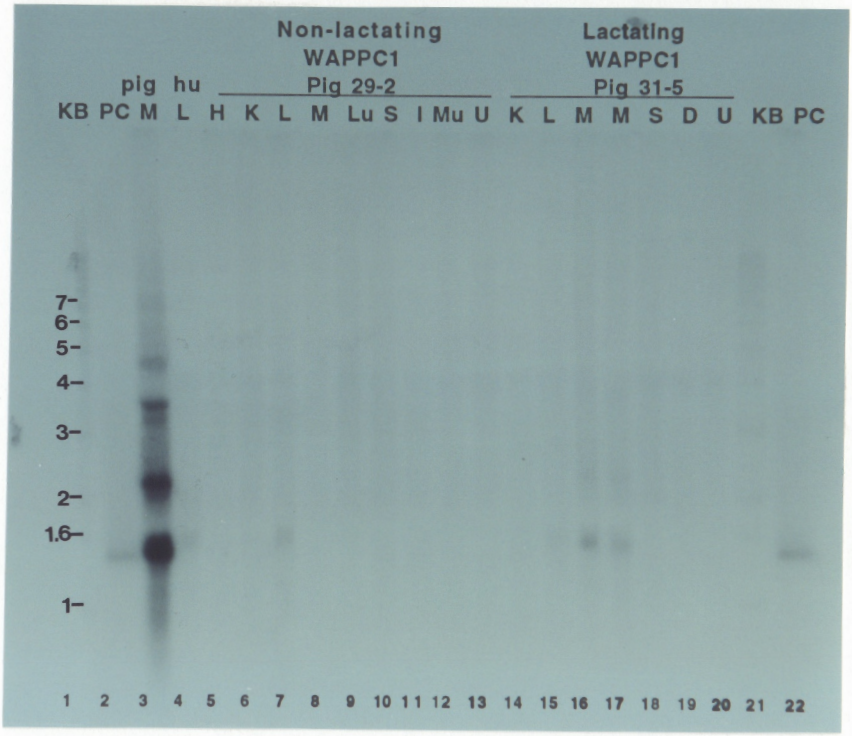


Figure 17 A
RNA In-Gel Hybridization -- RNA Gel 4

WAPPC1 Pigs, 29-2 and 31-5

10 μg total RNA was used for each RNA lane. Gel was hybridized overnight as in **Materials and Methods** with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 20 h at room temperature. Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, WAPPC1 Pig 29-2 RNA, 10 days post-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8, Mammary Gland; 9, ovary; 10, Salivary Gland; 11, Small Intestine; 12, Quadricep; 13, Uterus. Lanes 14-20, WAPPC1 Pig 31-5 RNA, day 15 of third lactation, from the following tissues: 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Diaphragm; 20, Uterus.

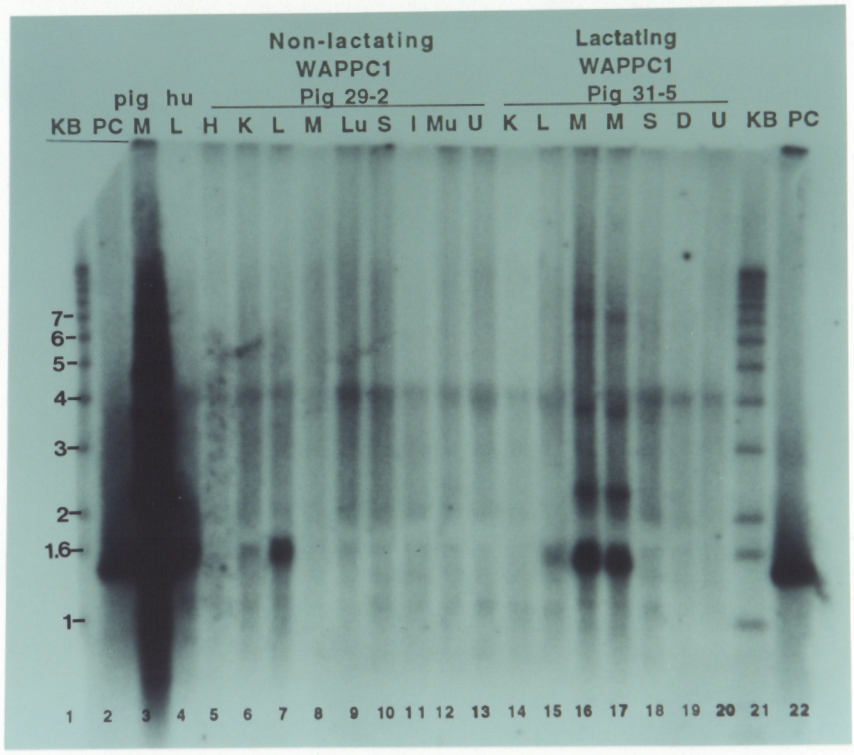


Figure 17 B
RNA In-Gel Hybridization -- RNA Gel 4

WAPPC1 Pigs, 29-2 and 31-5

Same gel and conditions as in Figure 17 A, except the exposure of x-ray film was about 84 h at -70°C with an intensifying screen (~30 times more exposure than in A). Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-12, WAPPC1 Pig 29-2 RNA, 10 days post-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8, Mammary Gland; 9, ovary; 10, Salivary Gland; 11, Small Intestine; 12, Quadricep; 13, Uterus. Lanes 14-20, WAPPC1 Pig 31-5 RNA, day 15 of third lactation, from the following tissues: 14, Kidney; 15, Liver; 16 and 17, Mammary Gland; 18, Salivary Gland; 19, Diaphragm; 20, Uterus.

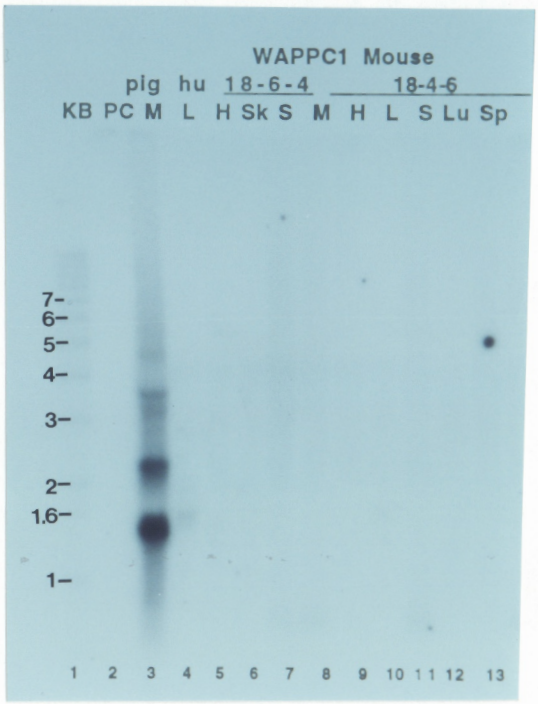


Figure 18 A
RNA In-Gel Hybridization -- RNA Gel 5

WAPPC1 Mice, 18-6-4 and 18-4-6

10 μg total RNA was used for each RNA lane. Gel was hybridized overnight as in **Materials and Methods** with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 20 h at room temperature. Lane 1, 1 kb ladder; Lane 2, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA. Lanes 5-7, WAPPC1 Mouse 18-6-4 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Skin; 7, Salivary Gland. Lanes 8-13, WAPPC1 Mouse 18-4-6, mid-lactation, from the following tissues: 8, Mammary Gland; 9, Heart; 10, Liver; 11, Salivary Gland; 12, Lung; 13, Spleen.

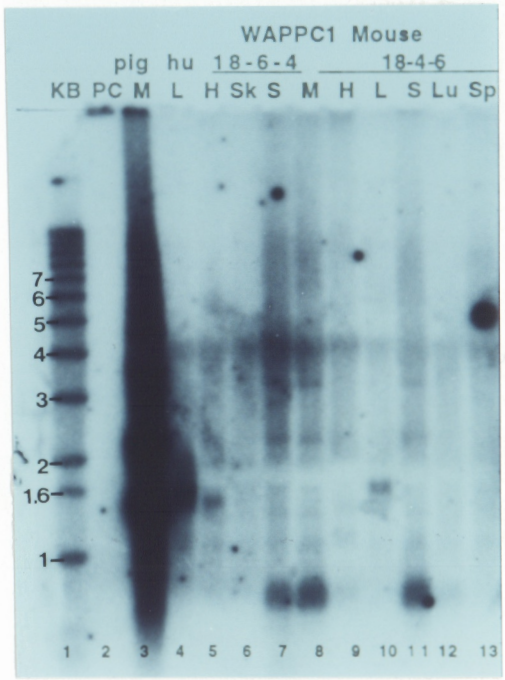


Figure 18 B
RNA In-Gel Hybridization -- RNA Gel 5

WAPPC1 Mice, 18-6-4 and 18-4-6

Same gel and conditions as in Figure 18 A, except the exposure of x-ray film was about 84 h at -70°C with an intensifying screen (~30 times more exposure than in A). Lane 1, 1 kb ladder; Lane 2, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA. Lanes 5-7, WAPPC1 Mouse 18-6-4 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Skin; 7, Salivary Gland. Lanes 8-13, WAPPC1 Mouse 18-4-6, mid-lactation, from the following tissues: 8, Mammary Gland; 9, Heart; 10, Liver; 11, Salivary Gland; 12, Lung; 13, Spleen.

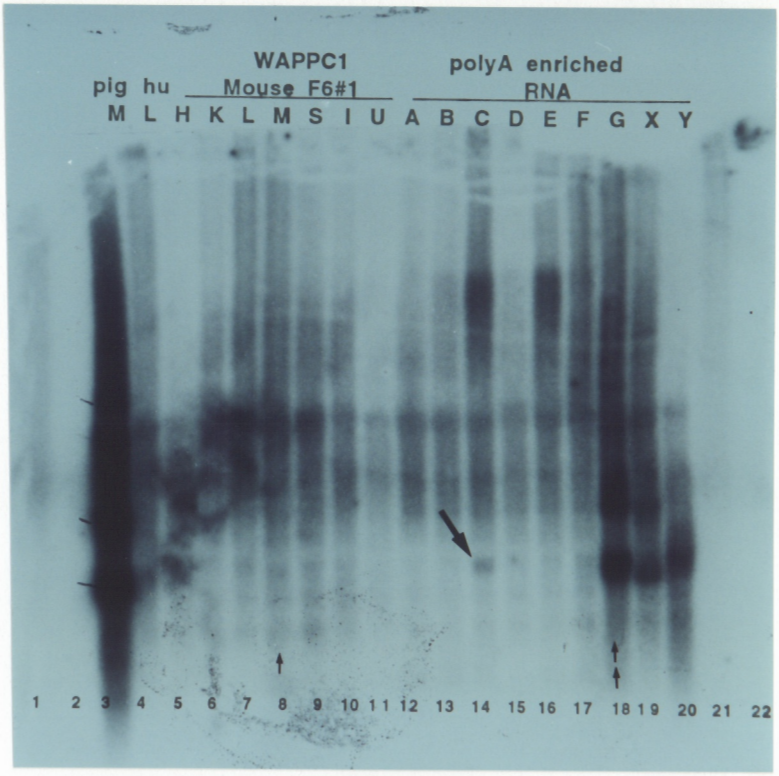


Figure 19
RNA In-Gel Hybridization -- RNA Gel 6

WAPPC1 Mouse F_6 #1 and polyA enriched RNAs

10 μg total RNA was used for each RNA in lanes 3-11. PolyA-enriched RNAs (100 μg total RNA equivalent, assuming 100% recovery) were used for each sample in lanes 12-20. Gel was hybridized overnight as in **Materials and Methods** with ^{32}P-labelled hPC cDNA and ^{32}P-labelled 1 kb ladder as probes. Exposure of x-ray film was 24 h at -70°C with intensifying screen. Lanes 1 and 21, 1 kb ladder; Lanes 2 and 22, hPC cDNA (1456 bp); Lane 3, WAPPC1 Pig 29-2 Mammary Gland RNA, day 55 of first lactation; Lane 4, Human Liver RNA; Lanes 5-7, WAPPC1 Mouse F_6 #1 RNA, mid-lactation, from the following tissues: 5, Heart; 6, Kidney; 7, Liver; 8, Mammary Gland; 9, Salivary Gland; 10, Small Intestine; 11, Uterus; 12-20, polyA enriched RNAs from the following (all mouse Mammary Gland samples are from mid-lactation): 12, Control Mouse Mammary Gland; 13, WAPPC3 Mouse 7-3-9 Mammary Gland; 14, WAPPC3 Mouse 21-14 Mammary Gland; 15, WAPPC3 Mouse 23-15 Mammary Gland; 16, WAPPC3 Mouse R46-2-2 Mammary Gland; 17, WAPPC3 Mouse 23-16 Mammary Gland, third lactation; 18, WAPPC1 Mouse F_6 #1 Mammary Gland; 19, WAPPC1 Pig 31-5 Mammary Gland, day 15 of third lactation; 20, Human Liver.

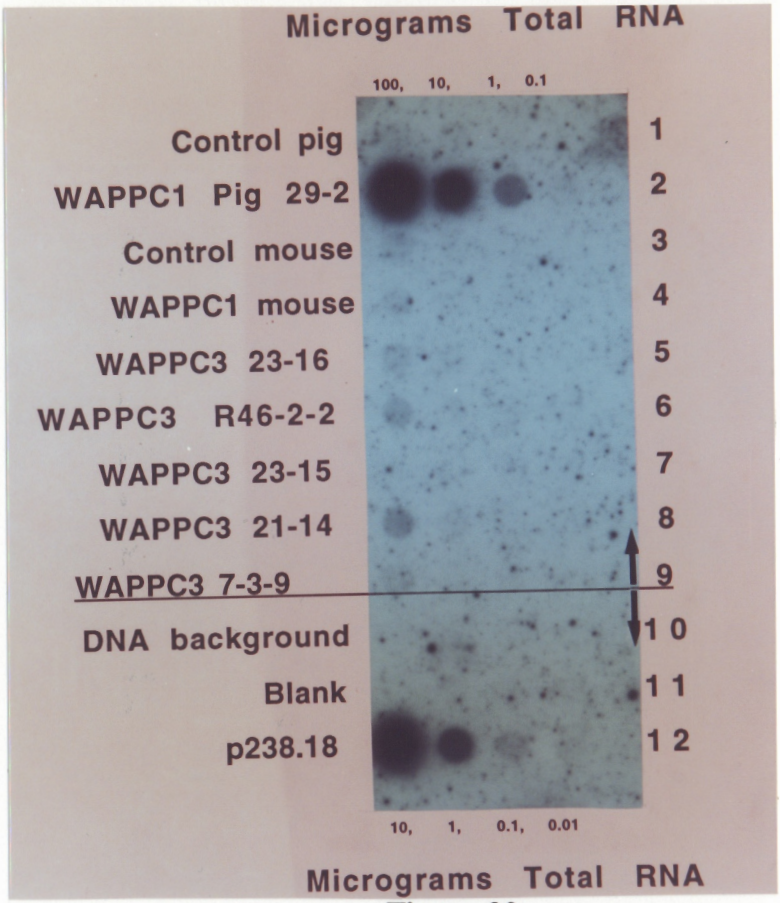


Figure 20
Mammary Gland RNA Dot Blot

Mammary gland total RNA was applied to a nitrocellulose membrane as in **Materials and Methods**. The blot was then crosslinked, baked, and hybridized with ^{32}P-labelled hPC cDNA as the probe (free of contaminating WAP gene sequences). Lanes 1-9 contained a dilution series of 100, 10, 1, and 0.1 μg total mammary gland RNA. Lanes 10 and 12 contained a dilution series of 10, 1, 0.1, and 0.01 μg total RNA. Lane 10 was a DNA background control of the indicated amounts of WAPPC3 mouse 7-3-9 mammary gland total RNA which had been subjected to alkaline hydrolysis (as in **Materials and Methods**) to remove all the RNA, thus leaving only background DNA in the sample. The lanes are from lactating mammary gland total RNA samples as follows: 1, control pig; 2, WAPPC1 Pig 29-2 (first lactation); 3, control mouse; 4, WAPPC1 mouse 74 F_6 #1 (first lactation); 5, WAPPC3 mouse 23-16 (third lactation); 6, WAPPC3 mouse R46-2-2 (first lactation); 7, WAPPC3 mouse 23-15 (first lactation); 8, WAPPC3 mouse 21-14 (first lactation); 9, WAPPC3 mouse 7-3-9 (first lactation); 10, WAPPC3 mouse 7-3-9 DNA background as indicated above; 11, blank; 12, p238.18 mouse #1 (fourth lactation).

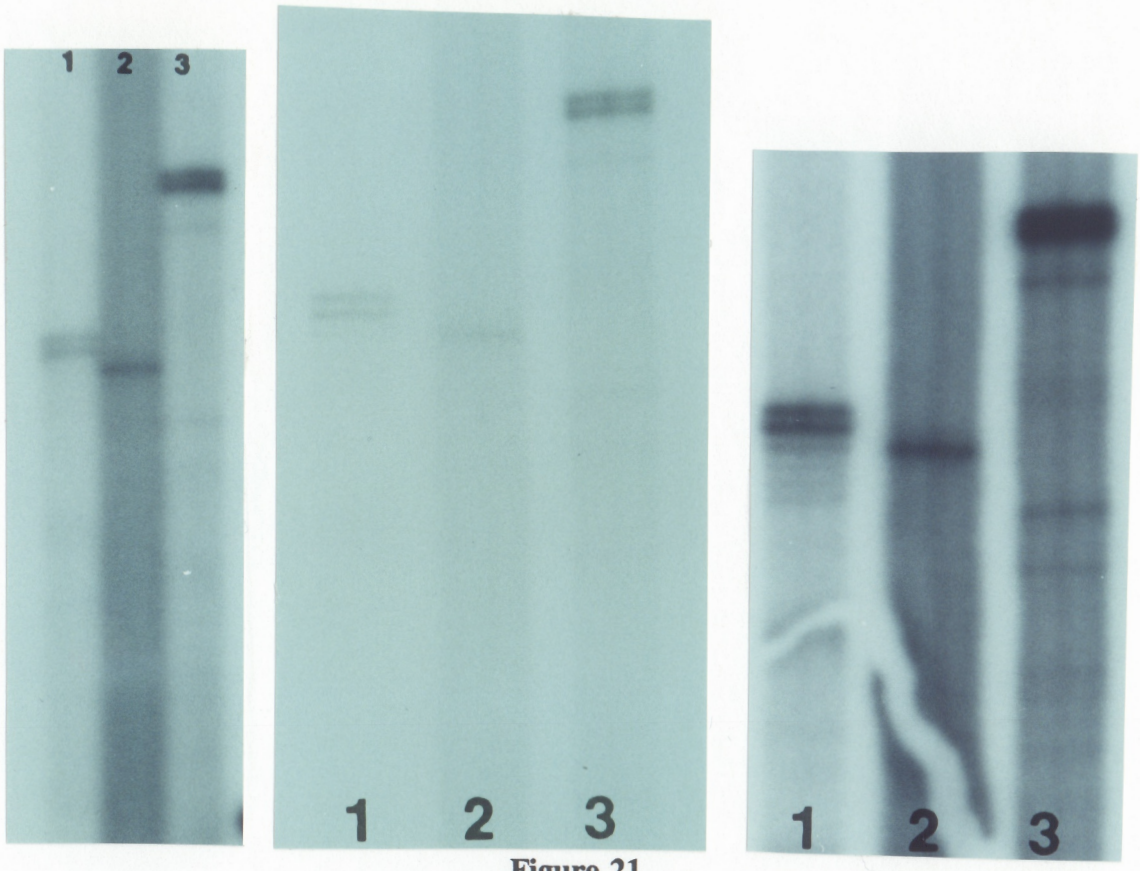


Figure 21
Primer Extension Analysis

Primer extension analysis was performed as in **Materials and Methods** to determine the transcription start sites of the transgenes and the endogenous WAP gene in lactating mammary gland RNA. No bands were obtained for WAPPC1 and WAPPC3 mouse RNA (not shown). All samples above contained 10 μg total RNA from lactating mammary gland. Lane 1, WAPPC1 Pig 29-2 with primer ProCA8; Lane 2, p238.18 Mouse #1 with primer ProCA8; Lane 3, Control Mouse with primer WAPA1. An adjacent sequencing lane (not shown) was used to determine the lengths of the products. The WAPPC1 pig and control mouse RNA both yielded a doublet one nucleotide apart, whereas the p238.18 mouse gave a single band. The size determinations indicated that all the genes in question had the same start site relative to the WAP promoter. Specifically, the transcription start site appears to be the C or the A, at nucleotide -28 or -27, respectively, for the endogenous WAP gene and the WAPPC1 transgene, and only at the A (-27) for the p238.18 transgene, with +1 being A in the AUG start codon of WAP mRNA.

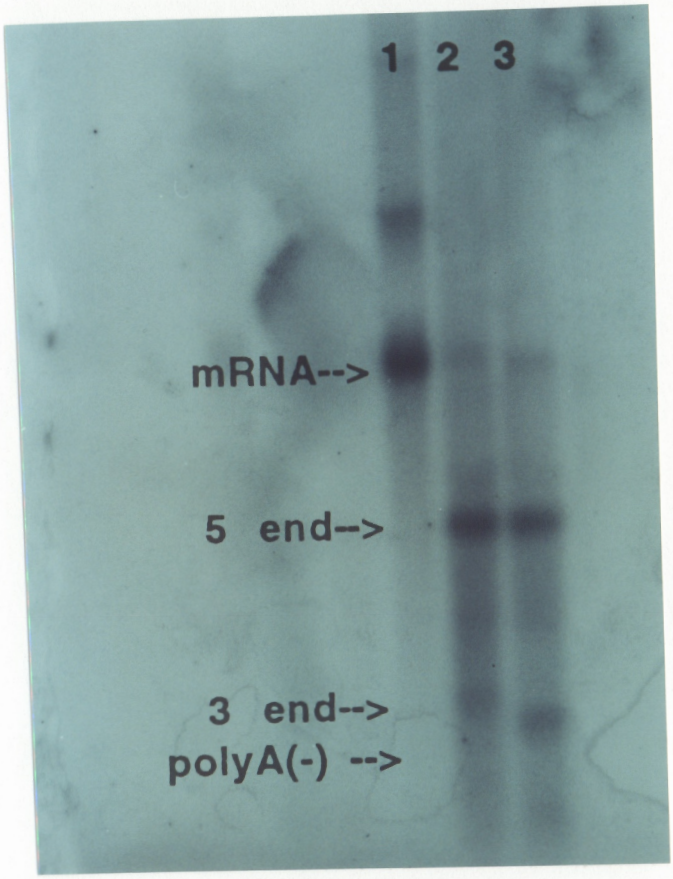


Figure 22
RNAse H Analysis

RNAse H analysis was done for transgenic animal mammary gland RNA samples. The only sample yielding clear bands was the WAPPC1 Pig 29-2 total RNA. 5 μg total RNA from Pig 29-2 was hybridized with the oligonucleotide ProCA5 and/or oligo$(dT)_{25}$ and treated with RNAse H, electrophoresed on a 2.5% agarose gel and dried and hybridized with ^{32}P-labelled hPC cDNA as in **Materials and Methods**. The RNAse H methodology allows site specific cleavage of an RNA. Lane 1 is the starting RNA, untreated; Lane 2, RNAse H with ProCA5 only; Lane 3, RNAse H with ProCA5 plus oligo$(dT)_{25}$. Bands corresponding to the full-length mRNA (Lane 1), 5' end (to ProCA5 site, Lanes 2 and 3), 3' end (up to ProCA5 site, Lane 2), and deadenylated 3' end (Lane 3) are indicated on the figure.

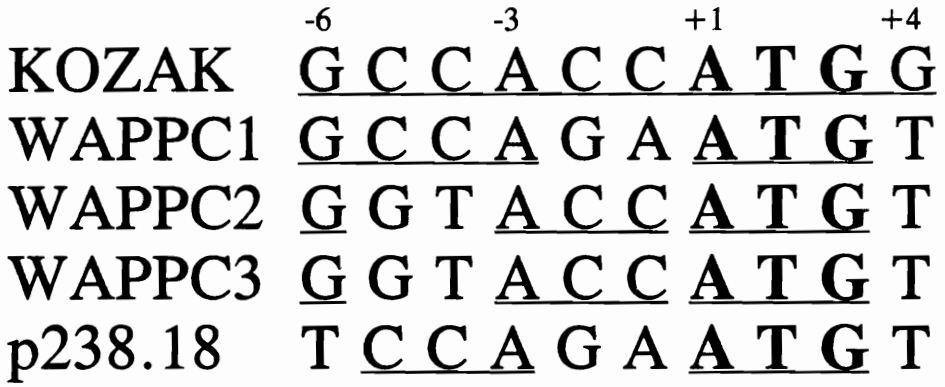


FIGURE 23
INITIATOR CODON EFFICIENCY

Nucleotides that match the consensus "Kozak" sequence are underlined. Experimentally, the -3 and +4 have the greatest effect on translational efficiency, therefore an initiator codon can be designated "strong" or "weak" by considering only these two positions.

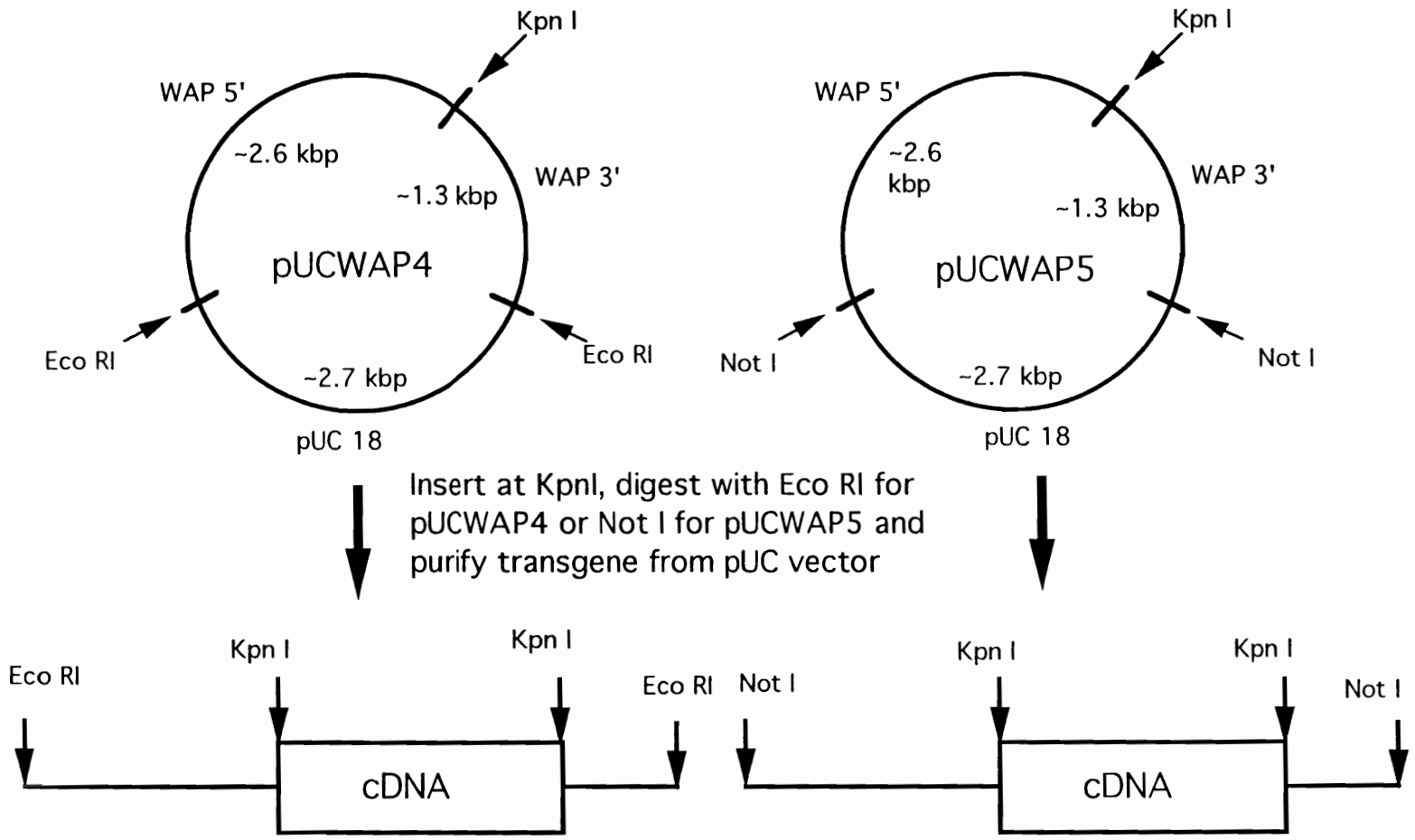


Figure 24
WAP "Cassette" Vectors

CURRICULUM VITAE

CHRISTOPHER GLYN RUSSELL

PERMANENT ADDRESS:
Route 2 Box 293
Hartselle, AL 35640
(205) 773-7657

OBJECTIVE: Responsible position in biotechnology research and development, with emphasis on medical applications of innovative science.

EDUCATION:

Ph.D. Chemical Engineering
Virginia Polytechnic Institute and State University, 1993
G.P.A 3.72/4.00
Southeastern Regional Fellowship, Pratt Presidential Scholarship.
Responsible for molecular biology in transgenic animal research, including: design, preparation, and evaluation of transgene constructs; molecular analysis and characterization of transgenic animals; analysis of transgene regulation and expression.
Optimized detection methods used for nucleic acid characterization.
Assisted in animal surgeries and maintenance. Some experience in embryology, protein biochemistry, and immunology.
Coursework emphasis on molecular biology, gene regulation, and biochemistry.

B.S. Chemical Engineering
Auburn University, 1989
G.P.A. 3.41/4.00 overall, 3.73/4.00 in major.
Graduated with honor.
University Honors Program, National Merit Scholarship, Auburn University Alumni Academic Scholarship, Outstanding College Students of America, Phi Lambda Upsilon (National Chemistry and Chemical Engineering Honorary), Omega Chi Epsilon (National Chemical Engineering Honorary), Intramural Sports.
Coursework emphasis on biochemical engineering.

EXPERIENCE:

June 1992 to present	TransPharm, Inc., Blacksburg, VA Consultant. Involved in research of company to produce transgenic animals. Molecular biology analysis and support.
Summer 1989	Amoco Performance Products, Inc., Augusta, GA Summer Intern. Analyzed faulty polymer reactor and recommended changes to improve performance.
Summer 1986,'87,'88	Saginaw Division, General Motors, Athens, AL Metallurgical and Chemical Laboratory Technician. Performed chemical, rubber, and metallographical analyses. Responsible for collecting samples, preparing and analyzing them, and effecting any corrective actions as necessary.
December 1985,'87,'88	United Parcel Service, Hartselle, AL Driver Helper. Assisted driver in delivering packages.

PUBLICATIONS IN REFEREED JOURNALS

W. H. Velander, R. L. Page, T. Morcol, **C. G. Russell**, R. S. Canseco, J. M. Young, W. N. Drohan, F. C. Gwazdauskas, T. D. Wilkins, and J. L. Johnson. 1992. "Production of Biologically Active Human Protein C in the Milk of Transgenic Mice". Annals of the New York Academy of Sciences 665:391-403.

W. H. Velander, R. L. Page, T. Morcol, **C. G. Russell**, J. L. Johnson, T. D. Wilkins, R. S. Canseco, B. L. Williams, F. C. Gwazdauskas, J. W. Knight, J. M. Young, and W. N. Drohan. 1992. "High Level Expression in the Milk of Transgenic Swine Using the cDNA Encoding Human Protein C". Proceedings of the National Academy of Sciences (USA) 89:12003-12007.

R. S. Canseco, A. E. T. Sparks, **C. G. Russell**, R. L. Page, J. L. Johnson, W. H. Velander, W. N. Drohan, R. E. Pearson, and F. C. Gwazdauskas. 1992. "Influence of DNA Microinjection on Embryo Development". Proceedings of the 12th International Congress of Animal Reproduction. 4:1764-1766.

A. E. T. Sparks, R. S. Canseco, **C. G. Russell**, J. L. Johnson, R. L. Page, W. H. Velander, M. L. McGilliard, and F. C. Gwazdauskas. 1993. "Evaluation of Bovine Morula Bisection and Biopsy as Cell Acquisition Techniques for DNA Amplification by the Polymerase Chain Reaction". Animal Reproductive Science (in press).

A. E. T. Sparks, R. S. Canseco, **C. G. Russell**, J. L. Johnson, H. D. Moll, W. H. Velander, and F. C. Gwazdauskas. 1993. "Evaluation of In Vitro Co-Culture, Microsurgery, and the Polymerase Chain Reaction Technique for Selection of Transgenic Bovine Preimplantation Embryos". Journal of Dairy Science (in press).

R. L. Krisher, F. C. Gwazdauskas, R. L. Page, **C. G. Russell**, R. S. Canseco, A. E. T. Sparks, W. H. Velander, J. L. Johnson, and R. E. Pearson. 1993. "FSH Induced Superovulatory Response of Dairy Goats with $PGF_{2\alpha}$ and GnRH". Theriogenology (in press).

R. S. Canseco, A. E. T. Sparks, R. L. Page, **C. G. Russell**, J. L. Johnson, W. H. Velander, R. E. Pearson, W. N. Drohan, and F. C. Gwazdauskas. 1993. "Gene Transfer Efficiency during Gestation and the Influence of Co-transfer of Non-manipulated Embryos on Production of Transgenic Mice". Transgenic Research (in press).

ABSTRACTS

R. S. Canseco, A. E. T. Sparks, **C. G. Russell**, R. L. Page, J. L. Johnson, W. H. Velander, W. N. Drohan, R. E. Pearson, and F. C. Gwazdauskas. 1992. "Influence of DNA Microinjection on Embryo Development". 12th International Congress of Animal Reproduction.

F. C. Gwazdauskas, R. L. Page, **C. G. Russell**, R. S. Canseco, A. E. T. Sparks, W. H. Velander, J. L. Johnson, and R. E. Pearson. 1992. "FSH Induced Superovulatory Responses in Dairy Goats with $PGF_{2\alpha}$ and GnRH. Journal of Dairy Science 75:(Supplement 1) 239.

A. E. T. Sparks, R. S. Canseco, **C. G. Russell**, H. D. Moll, W. H. Velander, J. L. Johnson, and F. C. Gwazdauskas. 1992. "Utilization of Pronuclear DNA Microinjection, In Vitro Culture, and the Polymerase Chain Reaction Technique for Production and Selection of Transgenic Bovine Preimplantation Embryos". Biology of Reproduction 46:(Supplement 1) 71.

R. L. Krisher, J. R. Gibbons, F. C. Gwazdauskas, **C. G. Russell**, T. D. Wilkins, and W. H. Velander. 1993. "Influence of Time of Gene Microinjection on Development and DNA Integration of Bovine Embryos". Biology of Reproduction (Supplement 1) 172.

Christopher G Russell
12/10/93