**Title**
Control of cell type specific expression in the rat exocrine pancreas



**Author**
Boulet, Anne Michele

**Publication Date**
1985

Peer reviewed|Thesis/dissertation

CONTROL OF CELL TYPE SPECIFIC EXPRESSION IN THE RAT EXOCRINE PANCREAS

by

ANNE MICHELE BOULET

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

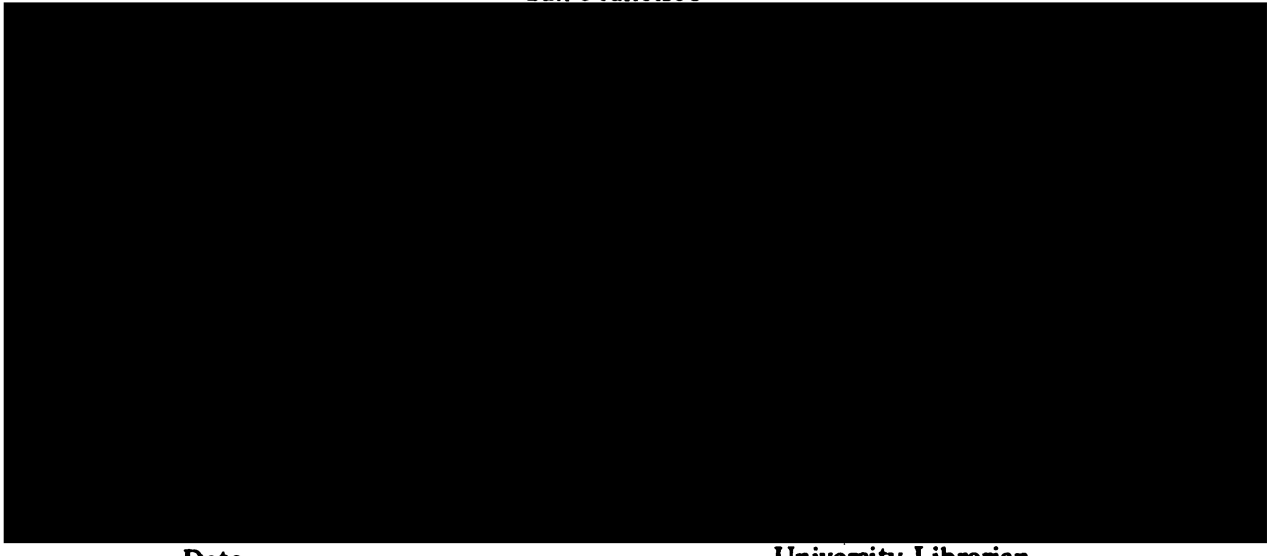
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Date University Librarian

Degree Conferred: SEP 8 1985

ACKNOWLEDGEMENTS

Without the help and support of numerous individuals, the trials of graduate school would have been much more difficult, if not impossible, to bear. I thank the past and present members of the Rutter laboratory / Hormone Research Institute for scientific guidance, for sharing their knowledge and expertise and for companionship. I am especially indebted to:

Mike Walker for teaching me everything I know about CAT, for discussions of my results, for ideas and advice and for appreciating my solutions.

Chris Erwin for invaluable contributions to the amylase and trypsin studies.

Thomas Edlund for advice on tricky techniques and for extremely useful plasmids (even those that give low yields).

Jenny Turner and Claire Griffin for preparing plasmid DNAs.

Homa Moezzi for help with tissue culture.

Leslie Spector for teaching me to use the word processor and for sharing her apartment.

Bill Rutter for making it all possible.

Friends for reminding me that science is not everything.

CONTROL OF CELL TYPE SPECIFIC EXPRESSION IN THE RAT EXOCRINE PANCREAS

BY ANNE MICHELE BOULET

The endocrine and exocrine cells of the pancreas are morphologically and functionally distinct. Endocrine cells synthesize hormones such as insulin and somatostatin while exocrine cells produce an array of digestive enzymes. I have investigated the mechanisms by which the genes encoding exocrine proteins are selectively expressed in the appropriate cells. The results reported in this thesis show that the 5' flanking regions of the amylase, chymotrypsin B and trypsin I genes direct preferential expression of a linked reporter function, the bacterial enzyme chloramphenicol acetyltransferase (CAT), in the pancreatic exocrine cell line AR4-2J. The upstream regions are inactive in fibroblasts and in a pancreatic endocrine cell line, and thus are likely to play a key role in establishment of cell type specific expression of exocrine genes.

The cis-acting elements of the amylase and chymotrypsin genes that confer exocrine cell specificity possess the characteristics of transcriptional enhancers: they act on the heterologous thymidine kinase promoter in a relatively position and orientation independent manner. The DNA sequences required for function of these enhancers have been mapped by deletion analysis. Sequences critical for activity of the amylase and chymotrypsin enhancers span regions of approximately 55 bp located 100 to 225 bp upstream from the transcription initiation sites. Additional sequences closer to the cap site of each gene comprise a second domain which contributes to the magnitude of enhancement obtained. Studies of the effect of mutations within the enhancer

regions suggest that specific nucleotide sequences are essential for enhancer function.

Comparison of the 5' flanking regions of nine genes expressed in pancreatic exocrine cells does not reveal any overall homology. However, a family of short related sequences has been identified. The homologous regions of the amylase and chymotrypsin genes overlap with the regions essential for cell type specific enhancer function, and the related sequences in the other exocrine genes are found at similar positions relative to the transcription start sites. The short homologous sequence may play a role in establishing high level expression of several genes in pancreatic exocrine cells: related sequences could, for instance, provide a recognition site for a single pancreas specific factor or a set of related factors.

TABLE OF CONTENTS

# INTRODUCTION

During the process of differentiation, each cell type in a complex multicellular organism is endowed with the ability to express a specific set of genes. Some of the genes comprising the repertoire of a particular cell are expressed in all cells of the organism. The products encoded by these "housekeeping genes" are required for cell metabolism and cell structure. At the opposite extreme are genes expressed at high levels in only one cell type and at undetectable levels in all other cell types.

The study of cell type specific regulation concerns the manner in which strict cell type specificity of gene expression is achieved and the mechanisms by which genes expressed at high levels in the same cell type are coordinately regulated. Formulation of tenable hypotheses addressing these aspects of cell type specific regulation requires an understanding of the basic concepts of gene regulation derived from investigation in prokaryotic and eukaryotic systems. Since cell type specific expression appears to be regulated mainly at the level of transcription (Darnell, 1982), the following discussion will be limited to research pertaining to this aspect of gene control.

Genetic and molecular analysis in bacteria has led to identification of DNA sequences required for transcription and the components of many gene regulatory circuits. Promoter elements are defined as the DNA sequences required for recognition, binding, and initiation of RNA synthesis by RNA polymerase. Promoters in E. Coli consist of two regions located upstream from the start site of transcription. The -10 region or Pribnow box and the -35 region are both required for transcription initiation, and the sequences of these

elements are conserved among various bacterial genes. The spacing between the two regions is critical: changes of only 1 or 2bp drastically affect promoter function (reviewed in Hawley & McClure, 1983).

Bacteria utilize both positive and negative regulatory mechanisms to modulate the frequency of transcription initiation and thus the level of gene expression. Many regulatory factors are DNA binding proteins which recognize specific sequences at or near the promoters of the genes under their control (reviewed in Raibaud & Schwartz, 1984; Ptashne et al., 1980). Some negatively acting molecules or repressors inhibit transcription initiation by blocking the binding of RNA polymerase to the promoter. Regulation of lac promoter function by the lac repressor-operator interaction (Majors, 1975) and regulation of the lambda promoter $P_R$ by the lambda repressor (Ptashne et al., 1980) are examples of this type of mechanism. Positive regulatory molecules can stimulate transcription by acting as RNA polymerase accessory factors which bind DNA sequences at or near the RNA polymerase binding site, by replacing an RNA polymerase subunit to alter the promoter specificity of the enzyme, or indirectly by altering the DNA configuration at the promoter. The mechanism of transcriptional activation at a specific promoter might involve enhanced binding of RNA polymerase to that site, an increased rate of isomerization to an open transcription complex, or decreased binding at competing RNA polymerase recognition sites (deCrombrugghe et al., 1984). The lambda repressor is thought to exert its positive effect on $P_{RM}$ transcription by direct contact with RNA polymerase bound at an adjacent site (Hochschild et al., 1983). Hawley and McClure (1982) determined that repressor increases the rate of

isomerization of closed to active open complexes at $P_{RM}$ while having no significant effect on RNA polymerase binding. At the lac promoter, the cAMP-CRP complex does not have an effect on the rate of isomerization, but stimulates the formation of closed complexes between RNA polymerase and the "correct" recognition site and inhibits binding to competing sites (deCrombrugghe et al., 1984).

Although most of the sites at which regulatory proteins interact in bacterial genes overlap or lie adjacent to the binding site for RNA polymerase, this is not true in all instances. The gal operon contains two repressor binding sites located approximately 110bp apart. One of these sites lies within the galE structural gene (Irani et al., 1983). Binding of the repressor to both sites appears to be required for repression of transcription from the gal promoters. Dunn et al. (1984) identified a binding site for the araC protein 280bp upstream from the transcription start site. Binding of araC to this region and to the araO region close to the transcription start site is required for repression of transcription. It is suggested that araC bound at the upstream site affects transcription by interacting with the initiation complex via formation of a DNA loop. In both cases, since one of the two repressor binding sites lies at or near the promoter, the actual mechanism for repression may not differ from that of the lac or lambda repressor. Interaction of the repressor at a second site may, for instance, somehow stimulate repressor binding at the active operator site.

E.Coli utilizes a mechanism involving replacement of an RNA polymerase subunit to activate transcription of specific genes during the heat shock response. The 32kd protein encoded by the HtpR gene replaces the 70kd sigma subunit of RNA polymerase and enhances

transcription initiation at heat shock promoters (Grossman et al., 1984). Sequence comparisons of heat shock and non-heat shock promoter regions and in vitro transcription experiments indicate that the holoenzyme containing sigma 32 (the HtpR product) recognizes different promoter sequences than does RNA polymerase containing sigma 70 (Grossman et al., 1984; Cowing et al., unpublished).

The schemes of positive and negative regulation mentioned thus far have served as a basis for the development of complex regulatory circuits in bacteria. A regulatory molecule such as the cAMP-CRP complex (deCrombrugghe et al., 1984) or the araC gene product (Lee et al., 1981) can function as either a repressor or an activator of transcription. The expression of a gene or operon may be controlled by more than one regulatory molecule: transcription of the lac operon is inhibited by the lac repressor and stimulated by cAMP-CRP (reviewed in Beckwith, 1978). Further levels of control can be achieved when bacterial genes possess more than one promoter as in the case of the gal operon (Musso et al., 1977). Other forms of transcriptional control, which will not be discussed in detail, also contribute to the complexity of gene regulation observed (reviewed in Platt, 1981).

Coordinate gene regulation in bacteria is accomplished by clustering of genes in polycistronic operons or by evolution of global regulatory networks in which physically separated operons are controlled by a common regulatory molecule. Coordinate induction of a set of genes during the heat shock response involves a factor, the HtpR gene product, which activates expression of several genes (Neidhardt et al., 1981).

The regulatory mechanisms utilized by lower eukaryotes such as the

yeast, Saccharomyces cerevisiae, show both similarities and differences to those described in bacterial systems. Yeast transcription typically depends on a TATA box, a conserved initiation element found upstream from eukaryotic RNA polymerase II genes, and one or more elements located between 30 and 300bp upstream from the transcription start site (Struhl, 1982a; Guarente et al., 1982; Beier & Young, 1982; Guarente & Mason, 1983). Upstream elements may act as regulatory sites, activating transcription in response to particular physiological signals, or may be responsible only for basal, unregulated levels of transcription (Guarente, 1984; Struhl, 1982a).

Yeast genetic analysis frequently has allowed the identification of both the gene coding for a regulatory protein and the site through which this protein exerts its effect. Transcription of the Gal1, Gal7, and Gal10 genes, encoding products involved in galactose utilization, requires positive activation by the product of the Gal4 gene (Johnston & Hopper, 1982; Laughon & Gesteland, 1982). The Gal1 and Gal10 genes, separated by about 600bp of DNA, are transcribed in opposite directions. Transcriptional induction by galactose is mediated by a common site located about 250 bp upstream from each of the respective transcription start sites (Johnston & Davis, 1984; West et al., 1984; Yocum et al., 1984). The product of the Gal4 gene is presumed to interact at the Gal1/Gal10 upstream activation site, termed $UAS_G$. When the region containing $UAS_G$ is placed upstream from the TATA box and initiation site of a heterologous gene, transcription of that gene becomes galactose-inducible (Guarente et al., 1982). $UAS_G$ was able to function in the absence of the Gal4 product when moved to within 40bp of the Gal1 TATA box, though at only about 5% of the induced level (West et al.,

1984). This latter result may indicate that the Gal4 protein stimulates an activity of $UAS_G$ which can also be "unmasked" when $UAS_G$ is situated closer to the TATA box.

The $UAS_G$ and other regulatory regions described for yeast genes differ from their prokaryotic counterparts in that they are located at a much greater distance from the transcription start site (reviewed in Guarente, 1984). In fact, UASs function at a variety of locations upstream of the TATA box and in an inverted orientation, but apparently cannot influence transcription when placed downstream from the initiation site (Guarente & Hoar, 1984; Struhl, 1984). These observations imply that the way in which positive and negative regulatory molecules interact with the transcriptional apparatus may be different in prokaryotes and eukaryotes in at least some cases. However, yeast and bacteria appear to achieve regulatory specificity through sequence-specific protein-DNA interactions: the Gal4 protein specifically recognizes four related 17bp sequences in $UAS_G$ (Giniger et al., 1985).

Establishment of mating type in yeast, which may be analogous in some ways to the generation of different cell types in higher eukaryotes, utilizes both positive and negative regulatory mechanisms. S. cerevisiae exhibits three cell types: a, alpha, and the a/alpha diploid. The a and alpha cell types are determined by the presence of the MATa or MATalpha allele at the mating type locus. Positive and negative regulatory proteins encoded by MATa and MATalpha control the expression of genes whose products distinguish a and alpha cells (MacKay & Manney, 1974; Strathern et al., 1981). Based on physiological and genetic evidence, the current model (Strathern et al., 1981) proposes

that two regulators encoded by MATalpha are produced in alpha-cells. Alpha1, a positive regulator, stimulates expression of alpha-specific genes while alpha2, a negative regulator, inhibits expression of a-specific genes. The characteristic pattern of a-cell gene expression is obtained because alpha-specific genes are not expressed in the absence of alpha1 and a-specific genes are expressed constitutively in the absence of alpha2. Sprague et al. (1983) isolated the alpha-specific STE3 gene and assayed RNA production from this gene in the presence or absence of the proposed regulatory molecule, alpha1. The results demonstrate that STE3 transcription requires the alpha1 product of the MATalpha locus.

In a similar manner, Wilson and Herskowitz (1984) showed that the alpha2 product represses transcription from an a-specific gene, STE6. Johnson and Herskowitz (1985) demonstrated that the alpha2 product acts by binding specifically to a 31bp site located 135bp upstream from the STE6 transcription initiation site. The binding site lies within a 130bp region identified by Wilson and Herskowitz (submitted) as necessary for alpha2 repression. A synthetic alpha2 binding site mediates repression of the heterologous Cyc1 promoter when placed between UAS1/UAS2 and the TATA region or upstream of the UAS region, in either orientation. Thus, unlike the lac and lambda repressors, the alpha2 repressor binding site need not overlap with promoter sequences. Sequences related to the 31bp region of the alpha2 recognition site are also found upstream from other genes known to be regulated by alpha2, i.e. a-specific genes.

Due to the lack of powerful genetic techniques applicable to higher eukaryotes, attempts to address the topic of gene regulation has relied

on in vitro systems, biochemical analysis, and, more recently, on introduction of DNA into cells in culture or into whole organisms. These approaches have yielded information on the nature of the transcription apparatus and, in some cases, on factors which mediate transcriptional specificity.

In vivo and in vitro studies of the expression of various eukaryotic protein coding genes have demonstrated that these genes contain a number of distinct transcription initiation elements. Some of these elements are found in close proximity to the transcription start site. The Goldberg-Hogness (TATA) box is generally located approximately 20-30bp upstream from the start site and appears to determine the precise site at which transcription begins and also to contribute to the transcription efficiency of some genes (Breathnach & Chambon, 1981). A second set of transcriptional control elements, located upstream of the TATA box and often referred to as distal signals, is less well-conserved from one gene to another in sequence, precise position with respect to the start site, contribution to attainment of maximal levels of RNA synthesis, and functional characteristics. These sequences generally map within a region -40 to -110bp from the transcription initiation site, and each gene may contain more than one distinct promoter element in this region. Several of these "upstream promoter regions" are characterized by short (6-22bp) repeated sequence elements (McKnight & Kingsbury, 1982; Everett et al., 1983; Dierks et al., 1983). More than a single copy of these short sequences appears to be required for optimal promoter activity. The exact position of the upstream promoter elements with respect to the initiation site is not critical for function in the genes in which this question was addressed. The SV40 21bp repeats

function as promoter elements for transcription from both the early-early and late-early start sites which lie approximately 30bp apart (Baty et al., 1984). The distance between the first and second distal signals and the proximal signal or transcription start site of the HSV tk promoter can be increased by 30-50bp without severely affecting the transcription level (McKnight, 1982). In addition, some of these elements are able to act in either orientation. The second distal signal of the HSV tk gene is active in either orientation at a location 32bp upstream from the cap site (McKnight et al., 1984). The region containing both distal signals is also fully functional when inverted. The hexanucleotide core sequence of the second distal signal is inverted with respect to that of the first signal, consistent with the idea that this sequence functions in a bidirectional manner. Thus, the distal or upstream elements of higher eukaryotic promoters resemble the upstream activating sequences in yeast, but are more sensitive to precise position with respect to the TATA box.

Enhancer elements, which contribute to transcriptional efficiency, can be considered as an additional class of transcription initiation elements. Enhancers have been distinguished from other promoter elements by their ability to act on homologous or heterologous proximal promoter elements (TATA boxes or cryptic TATA boxes) in a relatively position and orientation independent manner. Enhancer sequences are often active at a distance of up to several kilobases from the cap site, in either the 5' or 3' direction, and in both orientations. The effect of enhancers on transcription efficiency is not always dependent on the presence of distal transcription signals (Graves et al., submitted). On the other hand, the activity of the 72bp repeats on the SV40 early promoter is

decreased significantly in the absence of the 21bp repeat elements (Everett et al., 1983). Another feature which distinguishes enhancers from other transcriptional control elements is that DNA sequences required for enhancer function generally span a larger region, on the order of 40-75bp.

Characterization of the DNA sequences required for efficient transcription initiation provides a framework for the study of factors which regulate gene expression in higher eukaryotes. Several groups have identified components of cell extracts which impart varying degrees of promoter specificity to reconstituted in vitro systems. The transcription factor Sp1, isolated from HeLa cell extracts, activates in vitro transcription of a class of polymerase II promoters including the SV40 early promoter but has little or no effect on transcription from other promoters examined (Dynan & Tjian, 1983a). The heat shock transcription factor (HSTF) isolated from extracts of heat shocked Drosophila $K_c$ cells is required for specific transcriptional activation of the hsp70 gene in a reconstituted in vitro system (Parker & Topol, 1984). In addition, each of these factors bind to specific regions of DNA upstream from the transcription start site (Dynan & Tjian, 1983b; Parker & Topol, 1984).

To date, studies in viral systems have provided most of the information on mechanisms of gene regulation in higher eukaryotes. SV40 T antigen stimulates viral replication and inhibits transcription from the early promoter (reviewed in Tjian, 1981). T antigen binds to three distinct, closely spaced recognition sites in the viral genome in a region containing the origin of replication and the early promoter. Studies utilizing SV40 mutants indicate that binding of T antigen to

both sites I and II is required for repression (Rio & Tjian, 1983). Transcriptional repression apparently does not result from a block in elongation: transcription is not inhibited by the binding of T antigen to sites inserted downstream from the start site of a heterologous promoter (Myers et al., 1981). This result suggests that repression of early promoter function is due to a direct inhibition of transcription initiation.

Glucocorticoids stimulate transcription initiation from the mammary tumor virus (MTV) promoter (Ringold et al., 1977; Ucker et al., 1983). This hormone response provides an example of positive regulation in a mammalian system. The glucocorticoid receptor binds to specific sequences both within and upstream of the transcribed region of MTV (Payvar et al., 1983). Chandler et al. (1983) have shown that the MTV LTR contains a glucocorticoid response element (GRE), which confers hormone responsiveness on a heterologous promoter. The location of the GRE correlates with the sites of _in vitro_ receptor binding, indicating that specific receptor:DNA interactions mediate glucocorticoid regulation. In addition, the GRE functions at many positions and in either orientation with respect to the promoter (Chandler et al., 1983; Yamamoto, 1983), suggesting that the GRE is a hormone responsive enhancer element. The significance of the involvement of a sequence-specific DNA binding protein in enhancer function is further discussed in Chapter 3.

The adenovirus E1A gene product is a second example of a positive regulatory molecule active in higher eukaryotes. The E1A protein is required for transcription from early viral promoters at early times after infection (Jones & Shenk, 1979; Berk et al., 1979). Two models

have been proposed for E1A function. Nevins (1981) proposed that the E1A product inactivates a cellular factor which inhibits transcription from early viral promoters. Gaynor & Berk (1983) favor the hypothesis that the E1A protein catalyzes the assembly of stable transcription complexes at early viral promoters.

Just as certain prokaryotic regulatory molecules can function as both positive and negative regulators, the E1A product also has the ability to repress transcription (Borrelli et al., 1984; Velcich & Ziff, 1985). Repression appears to result from an effect on enhancer function, presumably through an interaction between the enhancer and a trans-acting factor, perhaps the E1A protein itself. (However, there is no evidence that the E1A product is a DNA binding protein.) In addition, there is evidence that SV40 T antigen and the glucocorticoid receptor can have positive and negative effects, respectively (Keller & Alwine, 1984; Guertin et al., 1983). These observations also suggest that enhancer regions may contain regulatory sites mediating both positive and negative control.

In spite of the apparent complexity of transcription initiation elements in higher eukaryotes, the available evidence suggests that basic themes established in prokaryotes, such as regulation by both activators and repressors and the importance of sequence-specific protein:DNA interactions, are conserved. However, the characteristics of enhancer elements and their similarity to the UAS elements of yeast imply that eukaryotes have evolved different mechanisms for the action of certain regulatory molecules on RNA polymerase.

Control of gene expression in eukaryotes can also be exerted by

altering chromatin configuration. Chromosomal domains containing active genes are characterized by an overall increase in nuclease sensitivity and by the presence of hypersensitive sites (reviewed in Weisbrod, 1982). Little is known about the mechanisms used to change chromatin structure or to specify regions which are to undergo decondensation in a particular cell or in response to certain external stimuli. Recent evidence suggests that glucocorticoid receptor:DNA interactions at the GRE (Zaret & Yamamoto, 1984) and interaction of a specific factor from erythrocytes with the 5' flanking region of the chicken adult beta-globin gene (Emerson & Felsenfeld, 1984) may be involved in creation of DNase hypersensitive regions.

Early in 1983, when the work described in this thesis was initiated, very little progress had been made toward identification of regulatory mechanisms utilized by higher eukaryotic organisms to control cell type specific expression. Studies in this area were hampered by the complexity of multicellular organisms and difficulties encountered in attempts at experimental manipulation of these systems. However, using a variety of methods for introduction of DNA into cultured cells or whole organisms, it had been possible to mimic to some extent cell type specificity of gene expression in several systems. Expression of an intact chicken delta crystallin gene was higher by at least one order of magnitude after microinjection into mouse lens epithelial cells than after injection into a variety of non-lens tissues (Kondoh et al., 1983). A chicken lysozyme recombinant gene was expressed in chicken oviduct cells but not in chicken macrophages or fibroblasts (Renkawitz et al., 1982). Rice & Baltimore (1982) reported regulated expression of

a transfected kappa light chain immunoglobulin gene in a lymphoid cell line, and Oi et al. (1983) obtained preferential expression of a kappa gene in Ig-producing cells. Human and mouse beta-globin genes are appropriately regulated after introduction into mouse erythroleukemia (MEL) cells (Chao et al., 1983; Wright et al., 1983). In summary, results reported up to that time suggested that protein coding genes may contain information required to direct high level expression in the proper differentiated cell type.

The rat pancreas provides an interesting and potentially instructive model for control of cell type specific expression in mammals. Morphologically and functionally distinct cell types express unique sets of genes. Dr. Rutter's laboratory has focused on pancreatic genes whose expression shows the most dramatic differences between cell types, those expressed at high levels in the pancreas and at undetectable levels in all other tissues. Michael Walker and Thomas Edlund developed methods for studying cell type specific expression of the rat and human insulin genes in beta cells of the endocrine pancreas (Walker et al., 1983). Transient transfection assays in which a reporter function is used to measure expression greatly facilitated the analysis of potential control regions. Once these methods were available, I was able to design experiments to elucidate the control of cell type specific expression of pancreatic exocrine genes.

When I began these experiments, there were no examples of cell type specific enhancers in the literature. Although viral enhancer sequences are active in a wide range of cell types, they do exhibit marked preferences for cells of particular species. For example, in mouse cells

the mouse polyoma virus enhancer is slightly more active than the SV40 enhancer in stimulating transcription of the rabbit B-globin gene. In primate (human) cells, the SV40 enhancer is 4 to 6 fold more active than the polyoma enhancer (deVilliers et al., 1982). Similarly, in transient transfection experiments, the SV40 enhancer-early promoter combination was more active in CV-1 (monkey) cells than an MSV enhancer-SV40 early promoter fusion while the MSV enhancer-SV40 promoter was the more active of the two in mouse L cells (Laimins et al., 1982). The relative activities of different viral enhancers correlates to some degree with the host range of the respective viruses.

The existence of cell type specific enhancers has now been demonstrated in a number of cellular genes. Immunoglobulin K light chain (Queen & Baltimore, 1983; Picard & Schaffner, 1984) and heavy chain (Gillies et al., 1983; Banerji et al., 1983; Neuberger, 1983) genes contain regulatory elements within the J-C introns which are required for high level expression of these genes in Ig-producing cells. These elements are able to act on heterologous promoter elements, function either 5' or 3' of the start site at a distance of up to several kilobases, and function in either orientation with respect to the direction of transcription. However, these enhancer sequences differ from the viral enhancers described above in that they are strictly cell type specific, functioning only in cells of the B lineage. The location of the immunoglobulin enhancers within the J-C introns provides an explanation for the transcriptional activation of rearranged variable (V) regions in cells in which unrearranged V regions remain relatively inactive.

Recently, Gillies et al.(1984) have shown that the $E_B$ gene encoding

the B-chain of a class II molecule of the major histocompatibility complex (MHC) also contains a cell type specific enhancer. The $E_B$ gene enhancer is located upstream from the cap site and functions only in a B lymphoma cell line which expresses class II antigens. In contrast, the heavy chain enhancer element is active both in myeloma cells, which express Ig genes, and B lymphoma cells, which do not express these genes.

The rat insulin gene contains a cell type specific enhancer element within the region from -249 to -103 relative to the transcription start site (Edlund et al., submitted). The enhancer stimulates transcription from the HSV tk promoter and is active when placed 600bp upstream or 1700bp downstream of the cap site in either orientation. The insulin enhancer functions only in pancreatic B cells: it is inactive in all fibroblast lines tested and in a pancreatic exocrine cell line.

Control of beta-globin gene expression involves sequences located much closer to the transcription start site than the cell type specific enhancers described above. Isolated beta-globin genes introduced into MEL cells are appropriately regulated during DMSO-induced differentiation (Chao et al., 1983; Charnay et al., 1984), and induction of beta-globin mRNA levels does not require more than 58bp of sequences upstream from the mRNA start site (Wright et al., 1984). Sequences responsible for induction of the beta-globin gene were localized using gene hybrids containing beta-globin sequences 5' or 3' of the translation initiation codon and the corresponding portions of a gamma-globin (uninducible) gene or a class I MHC gene (Wright et al., 1984). All hybrid genes produced increased levels of mRNA during MEL cell differentiation. Induction was due, at least in part, to an

increase in transcription, even for hybrids containing only beta-globin sequences downstream from the initiator codon. Thus, DNA sequences required for beta-globin gene activation during a process which may mimic normal development are found both within the coding region and upstream of the point at which translation is initiated.

With the goal of identifying the mechanisms of cell specific expression in the exocrine pancreas, the research described in the following chapters addresses three specific questions:

I. Do the 5' flanking regions of exocrine pancreatic genes contain the signals required to direct high level expression in pancreatic acinar cells?

II. Do the 5' flanking regions of the exocrine genes coding for amylase and chymotrypsin B contain cell type specific enhancer elements analogous to the element identified in the insulin gene? If so, which upstream sequences are required for enhancer activity?

III. Does the identification of cell type specific enhancers in exocrine pancreatic genes provide insight into the mechanism of coordinate control of gene expression in the exocrine pancreas?

# CHAPTER 1

## THE 5' FLANKING REGIONS OF THE AMYLASE, CHYMOTRYPSIN, AND TRYPSIN GENES ELICIT CELL TYPE SPECIFIC EXPRESSION IN A RAT EXOCRINE PANCREATIC CELL LINE

Introduction

Pancreatic exocrine or acinar cells are highly specialized to synthesize and secrete several digestive enzymes. Exocrine genes expressed at high levels in the rat pancreas include those for proteases (trypsins, elastases, chymotrypsins, and carboxypeptidases), lipases, ribonuclease, and amylase. In general, mRNAs for the exocrine products are present at 10,000-100,000 fold higher levels in pancreatic acinar cells than in other cell types (J. Han and W.J. Rutter, in preparation).

Mechanisms proposed for regulation of cell type specific expression must explain two aspects of this control; the nearly undetectable levels of mRNAs in "non-expressing" cells and the extremely high levels of mRNAs found in "expressing" cells. Models must also suggest ways in which the patterns of expression established during development are stably maintained.

Evidence from several systems suggests that gene activation requires modification of chromatin structure to produce domains accessible to transcription machinery and activation of transcription in selected subregions of decondensed chromatin (Weintraub & Groudine, 1976; Palmiter et al., 1978; Miller et al., 1978; Wu, 1980; Wu & Gilbert, 1981). As applied to the exocrine pancreatic system, such a two step scheme would predict that during differentiation regions of chromatin containing the various exocrine genes would first undergo decondensation. Extracellular signals acting on acinar cell precursors may determine the proper pattern of active chromatin formation. These signals may act via control genes which translate positional or temporal information into changes in chromosome structure at appropriate regions

of the genome. In the second phase, the expression of exocrine genes within the decondensed chromatin would be activated to produce the high levels of mRNAs seen in the adult rat. Pancreatic exocrine cells may contain specific factors which interact with regulatory sites in the proper genes to stimulate their expression. The production of these factors would in turn be regulated in a cell type specific fashion.

The accumulation profiles of pancreatic products during development suggest a biphasic process. Low level expression of pancreatic genes can first be detected at 11-12 days of gestation in the rat. This "protodifferentiated state" is maintained until 15-17 days at which time synthesis of pancreatic mRNAs is greatly stimulated (reviewed in Rutter et al., 1978; Han & Rutter, in preparation). As the chromatin structure of pancreatic genes has not been examined, it is not known whether the protodifferentiated state results from a low level of transcription of open chromosomal domains prior to specific activation.

Nonpancreatic cells may fail to produce pancreas specific mRNAs simply because the regions containing exocrine genes have not undergone decondensation to form active chromatin and the proper cell specific regulatory factors are lacking. It is also possible that nonpancreatic cells contain factors which repress transcription of pancreas genes. The production of TAT (tyrosine aminotransferase) mRNA is repressed in rat hepatomaXmouse fibroblast hybrids which retain fibroblast chromosome 11 while expression of four other hepatic traits is unaffected (Killary & Fournier, 1984). The genetic locus responsible for specific repression of TAT expression in mouse fibroblasts has been designated Tse-1. These results suggest that negative as well as positive control mechanisms may be utilized to achieve the degree of differential activity seen from one

cell type to another.

As a first step in elucidating mechanisms of cell specific expression in the exocrine pancreas, I designed experiments to determine whether exocrine genes contain elements regulating their expression. Since elements required for efficient transcription and for gene regulation are often located upstream from the mRNA cap site, I initially tested the 5' flanking regions of exocrine pancreatic genes for the ability to direct higher levels of expression in exocrine cells than in fibroblasts or other cell types.

To assay putative control regions in pancreatic exocrine genes it was first necessary to obtain a cell line or tissue culture system expected to possess all factors needed for exocrine gene expression; in other words, a cell line which produces high levels of exocrine mRNAs. Secondly, activity of the test genes in the assay must be easily distinguishable from endogenous gene expression. In the CAT assay system, developed by Gorman et al. (1982a), fusion of the CAT coding region to presumptive regulatory sequences allows one to distinguish the activity of exogenously added DNA from the products of transcription of the endogenous gene.

## Exocrine Pancreatic Cell Lines

Five rat exocrine pancreatic cell lines were obtained from the laboratory of Dr. Y. Kim. The NRP cell line is a putative normal rat pancreatic line, but produces tumors in nude mice and forms colonies in soft agar (L. McIntyre, personal communication). Each of the other cell lines was derived from azaserine induced tumors of the rat exocrine pancreas. In order to determine which, if any, cell lines produce high

levels of the exocrine mRNAs, cytoplasmic RNA was prepared from each cell line, spotted onto nitrocellulose filters, and probed with labeled amylase, chymotrypsin, and ribonuclease cDNAs (figure 1). AR4-2J cells (Jessop & Hay, 1980) produce very high levels of amylase and chymotrypsin mRNAs. By inspection of the autoradiographs, the chymotrypsin mRNA level was crudely estimated to be 5-10% and that of amylase mRNA to be 10-20% of the adult rat pancreas level. In contrast, ribonuclease mRNA was not detectable in AR4-2J cells above the background signal in these experiments, even following a long exposure of X-ray film to the filter (data not shown). NRP and AT3A cells (Rao et al., 1980) produce significant amounts of ribonuclease mRNA, but produce only barely detectable amounts of amylase and chymotrypsin mRNA. AR2-1A (Hay et al., 1977) and AR4-1P (Jessop & Hay, 1980) cells do not show significant levels of any of these three mRNAs. (In particular, amylase mRNA levels in all cell lines except AR4-2J are much less than 0.1% that of the adult pancreas. The signals obtained with liver RNA which contains 0.1% the pancreatic level of amylase and with a 1:1000 dilution of adult pancreatic RNA are easily detectable.)

The AR4-2J Cell Line

AR4-2J cells were chosen for subsequent experiments because they produce high levels of two of the most abundant mRNAs in the rat exocrine pancreas, those coding for amylase and chymotrypsin. C. Erwin and J. Han have carried out a thorough and quantitative examination of the mRNA levels in AR4-2J cells for ten exocrine pancreatic genes. The results of this analysis are shown in Table I. Also shown are the amounts of each of these mRNAs as a percentage of total polyA+ RNA in

the adult rat pancreas calculated by J. Han from results of RNA dot blot analysis. In AR4-2J cells, amylase and chymotrypsin mRNAs are present at 42% and 18%, respectively, of the level found in the adult rat pancreas. Since amylase and chymotrypsin make up 25% and 8%, respectively, of the total polyA+ RNA in the pancreas, we can estimate that amylase and chymotrypsin represent approximately 10% and 1.5%, respectively, of the polyA+ RNA in AR4-2J cells. The estimated levels of the other pancreatic products in AR4-2J cells, expressed as a percentage of polyA+ RNA, are shown in the figure. The percentage of adult pancreatic mRNA levels maintained in AR4-2J cells varies from 0.9% to 42% for the various exocrine genes. Because the levels of expression of all exocrine genes has not decreased by the same amount, relative levels of mRNAs in AR4-2J cells are different from those in adult pancreas. For example, trypsin is the second most prevalent mRNA species in adult pancreas, but is present in AR4-2J cells at a lower level than all but two of the other mRNAs examined. This observation cannot readily be explained by a model in which a single factor activating all genes is present at reduced levels in AR4-2J cells, but suggests that the control of pancreatic exocrine gene expression involves a number of factors.

The response of AR4-2J cells to the hormones insulin and dexamethasone resembles that of normal exocrine pancreatic cells. Glucocorticoids stimulate amylase production in cultured embryonic pancreatic acini both at the transcriptional and post-transcriptional levels (Rall et al., 1977; Harding et al., 1978). Glucocorticoids increase the amylase content of AR4-2J cells 8 fold (Logsden et al., in press). This effect, similar in magnitude to that seen in cultured embryonic rudiments, is due to an increase in amylase synthesis,

partially attributable to an increase in amylase mRNA levels as measured by in vitro translation. Insulin is thought to play a role in regulation of pancreatic exocrine cell function. In particular, there is evidence that insulin is involved in the regulation of amylase mRNA levels (Korc et al., 1981). Moessner, et al. (in press) have demonstrated that insulin stimulates the growth of AR4-2J cells and selectively increases amylase synthesis.

In conclusion, the high level of expression of a number of pancreatic genes and the similarity of the glucocorticoid and insulin responses to those seen in normal pancreatic cells make the AR4-2J cell line a very useful system for study of the regulation of exocrine pancreatic gene expression.

## The CAT System

The transient expression assay developed by Gorman et al. (1982a) was used to determine whether the 5' flanking regions of exocrine pancreatic genes contain sequences involved in the control of cell type specific expression. The CAT system provides a rapid, quantitative, and sensitive assay for the activity of regulatory regions and can be used in a variety of mammalian tissue culture cell lines. The expression of a reporter function, the bacterial enzyme chloramphenicol acetyltransferase (CAT), is used to monitor the activity of eukaryotic transcriptional control regions. Sequences to be tested are fused to the coding region of CAT in a plasmid vector derived from pBR322 (Appendix I). Recombinant plasmids are transfected into various cell lines by the calcium phosphate coprecipitation technique (Appendix II). Approximately 48 hours after the addition of DNA, cells are harvested and crude cell

extracts are prepared. CAT activity is determined by monitoring the conversion of $C^{14}$-chloramphenicol to its acetylated derivatives using thin layer chromatography to separate products from unreacted substrate. The amount of CAT enzyme produced provides a measure of the activity of the transcriptional control region used.

Cell Type Specific Expression Directed By the 5' Flanking Regions of Exocrine Pancreatic Genes

The 5' flanking regions of the amylase, chymotrypsin, and trypsin I genes were fused to the CAT coding region (Appendix I). The 5' flanking region fragments used encompass sequences from -1300 to -17, -711 to -3, and -1172 to -10 relative to the cap sites of the amylase, chymotrypsin, and trypsin I genes, respectively. These constructions were transfected into AR4-2J cells and rat fibroblast cells. To control for differences in transfection efficiency seen with various cultured cells, the CAT activity obtained with the Amy-17.CAT, pBR.Cht.CAT, and pBR.TrpI.CAT constructions is expressed as a percentage of the activity observed with RSV.CAT. In RSV.CAT the Rous sarcoma virus promoter plus upstream sequences is fused to the CAT gene (Gorman et al., 1982b). This promoter shows no apparent cell type specificity in rodent cells: the ratio of activity of the HSV tk promoter (pTE1) to the RSV promoter is similar for all cell types used in these experiments (Table II). In contrast, the amylase, chymotrypsin, and trypsin I 5' flanking regions direct much greater relative levels of CAT expression in AR4-2J cells than in rat fibroblasts (Table II, figure 2). The amylase, chymotrypsin, and trypsin I constructions are 75, 110, and 15 fold more active, respectively, in AR4-2J cells than in fibroblasts.

The chymotrypsin-CAT fusion was also introduced into HIT cells, a transformed line derived from hamster pancreatic endocrine cells that produces insulin at a level 2-5% that of endocrine B-cells. While an insulin-CAT fusion directs preferential expression in HIT cells (Walker et al., 1983), pBR.Cht.CAT shows a lower relative level of CAT expression in HIT cells than seen in rat fibroblasts (Table II). Thus, the ability of the chymotrypsin 5' flanking region to direct high levels of CAT activity is specific to exocrine cells of the pancreas.

To determine whether the correlation between endogenous gene expression and activity of the chymotrypsin 5' flanking region extends to rat pancreatic cells which do not synthesize detectable levels of chymotrypsin mRNA, pBR.Cht.CAT was tested in AT3A cells. The AT3A cell line was derived from an azaserine induced pancreatic acinar cell carcinoma (Rao et al., 1980). Immediately after explantation of tumor fragments, the level of amylase secreted into the medium was very high, but within 3 days, amylase secretion declined to a barely detectable level. By passages 9 and 10 only negligible amounts of chymotrypsin and trypsin were present in the cells, and amylase was not detectable. When cultured AT3A cells were transplanted subcutaneously into nude mice, tumors formed within 3-4 weeks. Tumor cells contained low levels of amylase and about 3 to 6 fold more chymotrypsin and trypsin than detected in cultured AT3A cells after passages 9 and 10. pBR.Cht.CAT showed approximately 2% of the level of CAT expression directed by RSV.CAT in this cell line, a value only 7 fold greater than fibroblasts (compare to the value of 33% of RSV.CAT activity seen in AR4-2J cells).

Localization of the Cell Type Specific Elements in the Chymotrypsin and Amylase Genes

Deletions from the 5' end of the chymotrypsin flanking region were generated to identify sequences required for cell type specific expression. Deletions were produced by digestion at specific restriction enzyme sites or by treatment with nuclease Bal 31 (Appendix I). Each deletion end point was joined to the same plasmid DNA sequences at the ClaI site of pBR322. The activities of the deleted chymotrypsin fragments in AR4-2J cells were compared to that of the complete 5' flanking region fragment (-711 to -3). Deletion of sequences upstream from -225 had no effect on activity (figure 3). However, removal of a further 34bp or more (5' deletions to -192 or -184) causes a dramatic loss of cell type specific activity. In marked contrast, deletion of up to 618bp from the 5' end (to -93) does not decrease the CAT activity of the chymotrypsin 5' flanking region in rat or hamster fibroblasts (data not shown). These results indicate that the 5' border of the DNA sequences required for cell type specific expression is located between positions -225 and -192 relative to the chymotrypsin cap site.

An internal deletion from the chymotrypsin 5' flanking region was produced by removing the 182bp SstI fragment spanning the region from -275 to -93 (Appendix I). The chymotrypsin-CAT recombinant containing this deletion shows little or no cell type specific activity (figure 3). This result provides additional evidence that critical sequences are located in this region and indicates that sequences upstream of -275 cannot substitute for deleted elements to restore cell type specific activity. Two experimental observations indicate that this loss of cell type specific activity in AR4-2J cells is not due to the deletion of

sequences required for transcription initiation. The internal deletion mutant retains the same low level of activity in rat fibroblasts as the complete 5' flanking region. In addition, when the RSV enhancer is inserted upstream of the -93 to -3 chymotrypsin flanking region fragment, CAT activity in fibroblasts is increased by about 14 fold relative to that produced by this fragment alone. Thus, whether or not the region from -93 to -3 is capable of efficiently initiating transcription on its own, it can be activated by enhancer sequences located upstream.

Sequences from the amylase 5' flanking region upstream from a PvuII site at position -125 were deleted by digestion with this restriction enzyme (Appendix I). The resulting fragment, -125 to -17, was unable to direct cell type specific expression in AR4-2J cells (figure 3). Thus, sequences upstream from -125 are required for the activity of the amylase 5' flanking region. As expected, smaller fragments produced by further 5' deletions were also inactive in AR4-2J cells.

## Discussion

The 5' flanking regions of the rat amylase, chymotrypsin, and trypsin I genes elicit preferential expression of a linked reporter function, the CAT gene, in cells which produce amylase, chymotrypsin, and trypsin. The observation that deletions from the 5' end lead to a loss of activity suggests that the cell type specific element is part of a positive regulatory system. Analysis of deletions removing sequences from the 5' flanking region of the chymotrypsin gene indicates that DNA sequences between -225 and -192 are required for efficient cell type specific expression. A less extensive analysis showed that DNA sequences

upstream from -125 are required for the activity of the amylase 5' flanking region. Although the 5' deletion to -125 does not reduce the activity of the amylase fragment in fibroblasts (data not shown), the loss of activity in AR4-2J cells could be due either to deletion of sequences required for cell type specificity or sequences required for promoter function.

Ornitz et al. (1985) have demonstrated that 213bp of rat elastase I sequences, including 10bp of transcribed DNA and 203bp of upstream flanking sequences, are sufficient to direct specific expression of a fused human growth hormone gene in pancreatic exocrine cells of transgenic mice. This result is similar to our finding that 225bp of 5' flanking sequences from the chymotrypsin gene, which is also expressed in the exocrine pancreas, is able to direct cell type specific expression.

The activity of the chymotrypsin 5' flanking region in three cell lines derived from the endocrine or exocrine pancreas is directly related to the level of expression of the endogenous chymotrypsin gene. Episkopou et al. (1984) obtained similar results using the gpt gene (E. Coli xanthine/guanine phosphoribosyltransferase, conferring XAT resistance) linked to the rat insulin II promoter (663bp of 5' flanking sequences) in a retroviral vector. When the recombinant retrovirus was transduced into four cell lines derived from a rat insulinoma, XAT resistant colonies were obtained at a relatively high frequency only in the two cell lines which produce insulin. These results contrast with the activity of the immunoglobulin gene cell type specific enhancers. The Ig heavy chain enhancer is active in both myeloma cells which express the Ig genes and in B lymphoma cells which do not express Ig

genes (Gillies et al., 1984). This indicates that B lymphoma cells contain the factors necessary for activation of the Ig enhancer, yet maintain Ig genes in an inactive state by another mechanism. Apparently, genes introduced into the cell by DNA transfection are not subject to the same constraints on expression.

The cell type specific effects seen with the 5' flanking regions of the three genes studied do not correlate precisely with the relative levels of amylase, chymotrypsin, and trypsin mRNAs in AR4-2J cells. Although the endogenous amylase gene in AR4-2J cells is significantly more active than the endogenous chymotrypsin gene (contributing 10% and 1.5% of the total polyA+ RNA, respectively), the chymotrypsin 5' flanking region is more active in promoting cell type specific expression of the CAT enzyme in these cells. There is a better correlation between relative levels of chymotrypsin and trypsin mRNAs (1.5% versus 0.078% of total polyA+ RNA, i.e. chymotrypsin mRNA about 19 fold more abundant than trypsin mRNA) with the relative abilities of the chymotrypsin and trypsin 5' flanking regions to direct preferential expression in AR4-2J cells (chymotrypsin about 7 fold more active than trypsin).

In general, the cell type specific effects observed for all three genes are much less than expected. Although the amylase, chymotrypsin, and trypsin I genes are expressed at about 500 to 40,000 fold higher levels in AR4-2J cells than in fibroblast cells, the largest effect observed in the experiments described above is 110 fold.

The relative activities of the 5' flanking regions and the magnitude of the cell type specific effects observed indicate that important components of pancreatic gene regulation are not fully

functional in these experiments. Differential effects of RNA processing, transport, or stabilization of pancreatic mRNA species not measured in these experiments may contribute to the apparent discrepancy between endogenous mRNA levels and relative ability to direct cell type specific expression. The normal chromosome environment may be required for maximal activation in pancreatic exocrine cells and/or complete repression in non-expressing cells such as fibroblasts. DNA sequences in addition to those in the 5' flanking region fragments may help to establish the degree of cell type specificity attained in vivo. Further insight into the importance of additional regulatory components may be obtained by testing the activity of regions of pancreatic genes not included in the 5' flanking region fragments in the CAT system and by examination of stable transformants or transgenic animals generated using modified genes.

TABLE I: Pancreatic Exocrine mRNA Levels in Adult Rat and AR4-2J Cells

| GENE | AR4-2J mRNA Levels Relative to Adult Rat Pancreas (%) | Adult mRNA Levels (% Total Pancreas PolyA+ RNA) | AR4-2J mRNA Levels (Estimated % Total PolyA+ RNA) |
|---|---|---|---|
| AMYLASE | 42.1 | 25.0 | 10.5 |
| CARBOXYPEPTIDASE A1 | 33.8 | 4.6 | 1.6 |
| CARBOXYPEPTIDASE A2 | 4.5 | 3.7 | 0.17 |
| CARBOXYPEPTIDASE B | 2.5 | 5.0 | 0.12 |
| CHYMOTRYPSIN B | 18.4 | 8.0 | 1.5 |
| ELASTASE I | 10.9 | 1.5 | 0.16 |
| ELASTASE II | 2.8 | 2.2 | 0.062 |
| LIPASE | 44.4 | 0.7 | 0.31 |
| RNase | 3.0 | 2.5 | 0.075 |
| TRYPSIN | 0.9 | 8.7 | 0.078 |

The mRNA levels for ten pancreatic exocrine genes in the adult rat pancreas and the AR4-2J cell line are compared. Column 1 shows the level of each mRNA in AR4-2J cells expressed as a percentage of the amount seen in adult rat pancreas. The adult mRNA levels, expressed as a percentage of total pancreas polyA+ RNA, are shown in column 2. The third column contains estimates of the percentage of total AR4-2J cell polyA+ RNA represented by the relative levels shown in column 1. These values were derived by taking the column 1 percentages of the adult rat mRNA levels in column 2. The percentages of total AR4-2J cell polyA+ RNA for each mRNA have not been measured directly. The values in column 3 are estimates based on the assumption that the polyA+ RNA content of AR4-2J cells and adult rat pancreas are roughly equivalent (J. Han, in preparation).

TABLE II: Relative Activity of 5' Flanking Regions

| GENE FRAGMENT | CELL LINE | | | | | CELL TYPE SPECIFICITY |
|---|---|---|---|---|---|---|
| | RAT2 | XC | AR4-2J | HIT | AT3A | AR4-2J VS XC |
| RSV (RSV.CAT) | 100 | 100 | 100 | 100 | 100 | -- |
| TK (pTEI) | 5 | 4 | 3 | -- | -- | 0.8 |
| Chymotrypsin (pBR.Cht.CAT) | 0.4 | 0.3 | 33 | 0.2 | 2 | 110 |
| Amylase (pBR.Amy-17.CAT) | 0.2 | 0.2 | 15 | -- | -- | 75 |
| Trypsin I (pBR.TrpI.CAT) | 0.2 | 0.2 | 3 | -- | -- | 15 |

Relative CAT activity directed by the 5' flanking regions of pancreatic exocrine genes. For each cell type CAT activity is expressed as a percentage of that directed by the RSV fragment. AR4-2J and AT3A, rat pancreatic exocrine cell lines. HIT, hamster pancreatic endocrine cell. XC and RAT2, rat fibroblast cell lines.

FIGURE 1: Levels of chymotrypsin, ribonuclease, and amylase mRNAs in five rat exocrine pancreatic cell lines determined by RNA dot blot analysis. Nitrocellulose filters were spotted in triplicate with 10ug of total cytoplasmic RNA isolated from NRP, AT3A, AR2-1A, AR4-1P, and AR4-2J cells, 10ug of total rat liver RNA, and 10ug or 10ng of total rat pancreas RNA. Filters were hybridized with nick translated chymotrypsin ribonuclease, or amylase cDNA probes.

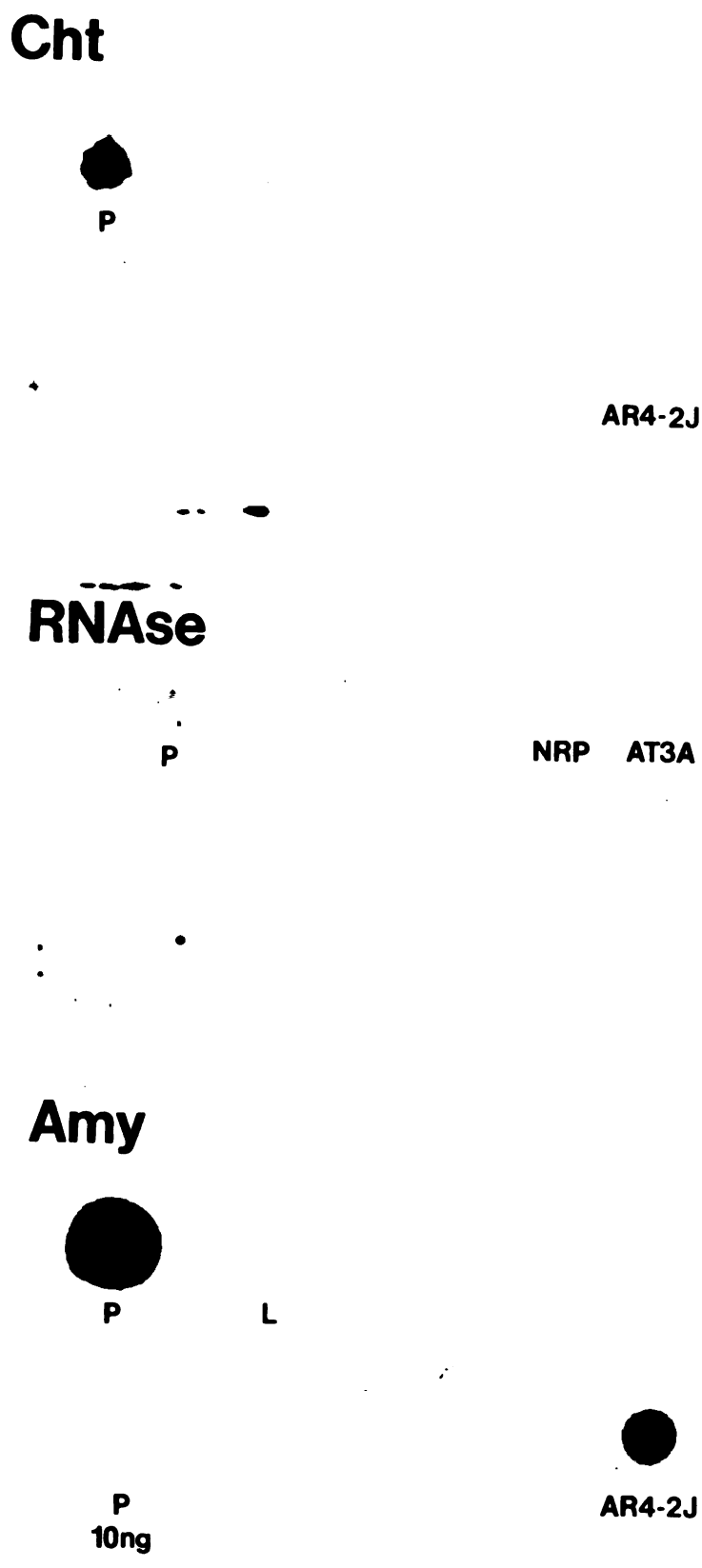



FIGURE 1

FIGURE 2: CAT activity directed by the 5' flanking regions of the rat amylase, chymotrypsin, and trypsin I genes in AR4-2J and XC cells in a representative experiment. Recombinant plasmids containing the 5' flanking regions of RSV, HSV thymidine kinase (tk), amylase (amy), chymotrypsin (cht), and trypsin I (trpI) genes linked to the CAT coding sequences were transfected into AR4-2J and XC cells. Crude cell extracts from transfected AR4-2J and XC cells containing 50ug or 25ug of protein, respectively, were assayed for CAT enzyme activity. AR4-2J and XC CAT assay reaction mixtures were incubated at 37°C for 60 minutes and 30 minutes, respectively. Samples were analyzed by ascending thin layer chromatography after extraction with ethyl acetate.

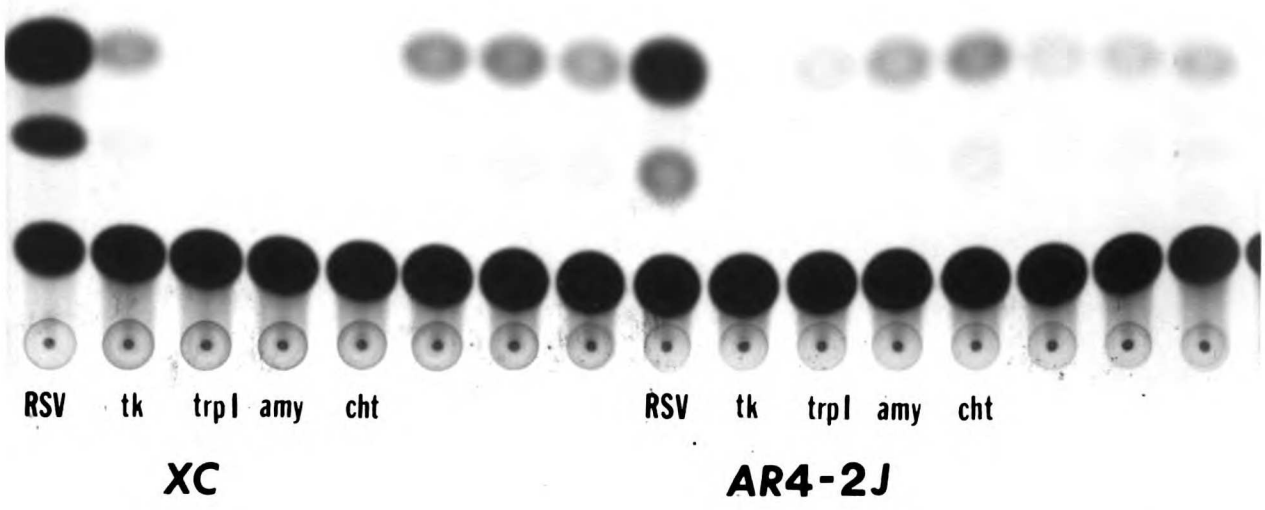



FIGURE 2

FIGURE 3: CAT activity directed by chymotrypsin and amylase 5' flanking region fragments containing 5' deletions or an internal deletion. Activity in AR4-2J cells is expressed as a percentage of that obtained with the chymotrypsin -711 to -3 fragment or the amylase -1300 to -17 fragment. The solid and open bars represent chymotrypsin and amylase sequences, respectively, retained in the deletion mutants. The deletion end points are indicated for each fragment, measured in terms of distance from the cap site of the respective gene.

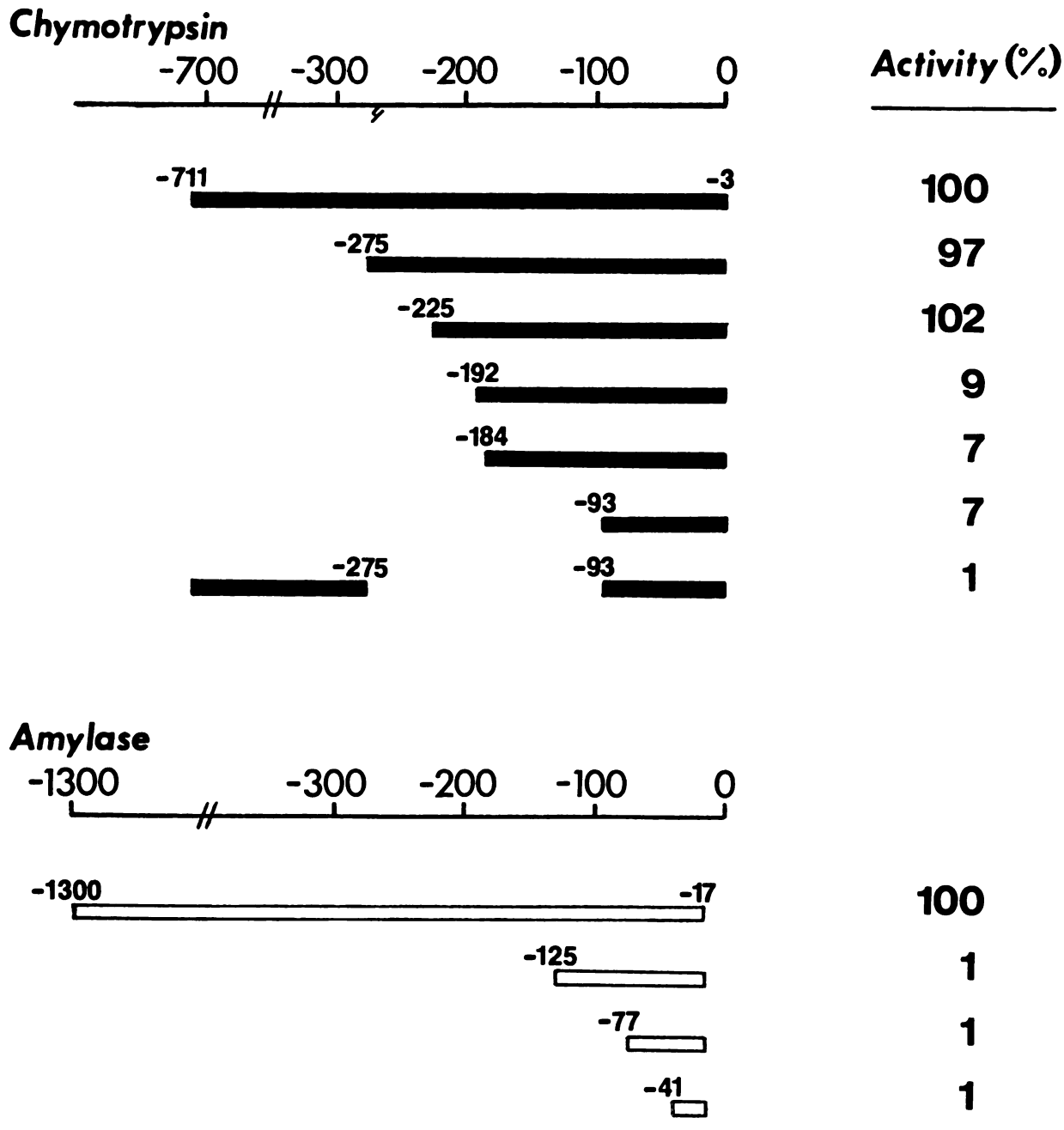

Chymotrypsin
-700
-300
-200
-100
0
Activity (%)
-711
-3
100
-275
97
-225
102
-192
9
-184
7
-93
7
-275
-93
1
Amylase
-1300
-300
-200
-100
0
-1300
-17
100
-125
1
-77
1
-41
1

FIGURE 3

# CHAPTER 2

# THE AMYLASE AND CHYMOTRYPSIN GENES CONTAIN CELL TYPE SPECIFIC ENHANCERS

## Introduction

The 72bp repeated sequences of the DNA tumor virus SV40 were the first example described of a class of transcription elements called enhancers (Banerji et al., 1981). Subsequently, enhancers have been found in several other viral genomes (deVilliers & Schaffner, 1981; Levinson et al., 1982; Lusky et al., 1983; Luciw et al., 1983; Hearing & Shenk, 1983). They are defined by several functional criteria. Enhancers do not contain intrinsic promoter activity, but act on associated transcription initiation elements. When removed from their natural locations, enhancers can also activate transcription from substitute promoters or cryptic transcription initiation elements. Enhancers are relatively position and orientation independent and are often able to activate transcription from promoter elements located several kilobases away.

Cell specific control elements found in the introns of immunoglobulin genes manifest the functional characteristics of viral enhancers (Gillies et al., 1983; Banerji et al., 1983; Neuberger, 1983; Picard & Schaffner, 1984; Queen & Stafford, 1984; Bergman et al., 1984). In addition, the immunoglobulin heavy and light chain enhancers show strict cell type specificity in that they are active only in cells of the B lineage. I have carried out experiments to determine whether the 5' flanking regions of the amylase and chymotrypsin genes contain analogous enhancers specific for pancreatic exocrine cells.

Pancreatic Exocrine Cell Specific Enhancers

An internal deletion of the 182bp SstI fragment (-275 to -93 relative to the mRNA cap site) from the chymotrypsin 5' flanking region resulted in the loss of cell type specific expression directed by the chymotrypsin promoter (Chapter 1). Sequences upstream from -225 do not appear to contribute to cell type specific activity, but the DNA sequence between -225 and -192 is critical for function. Because these results indicate that the Sst I fragment contains at least some of the elements required for cell specific activity, this fragment was used for further experiments.

When a DNA fragment from the 5' flanking region of the amylase gene, including sequences from -125 extending upstream to approximately -1300, was linked to the tk promoter, no enhancement of CAT activity directed by this promoter was seen in AR4-2J cells (data not shown). As described in Chapter 1, the portion of the amylase 5' flanking region downstream from -125 did not direct cell specific expression. These observations could indicate that sequences both upstream and downstream from -125 are essential for the activity of a pancreatic exocrine cell specific enhancer or the amylase promoter or that the sequences responsible for specificity do not affect expression from the tk promoter. To begin to distinguish among these possibilities, amylase 5' flanking region fragments spanning position -125 were tested for enhancer activity. An approximately 450bp AhaIII fragment (about -530 to -77) and a 193bp RsaI fragment (-235 to -41) were used in subsequent experiments.

The chymotrypsin SstI fragment and the two amylase fragments were tested for enhancer activity in the plasmid pTE1 (Edlund et al.,

submitted; Appendix I). In this plasmid, CAT gene expression is under control of the HSV tk promoter (-109 to +55, McKnight et al., 1981). DNA fragments inserted at a polylinker located 600bp upstream from the tk promoter can be tested for the ability to enhance transcription from the tk promoter. The 600bp of pBR322 DNA separating the polylinker from the tk promoter can be deleted to shorten the distance between the promoter and the presumptive enhancer sequence. Fragments can also be inserted at a BamHI site at the 3' end of the CAT/SV40 DNA sequences. The CAT activity directed by the tk promoter in the parent plasmid pTE1 serves as a reference or basal value.

Figure 4 shows the results obtained when various pTE1 derivatives were transfected into AR4-2J cells. CAT activity produced is expressed relative to pTE1 activity set at a value of 1.0. (Since initial experiments demonstrated that the amylase RsaI fragment gave higher levels of CAT activity than the AhaIII fragment (at least 2-3 fold, data not shown), the RsaI fragment was used exclusively for subsequent experiments.) When the chymotrypsin SstI fragment or the amylase RsaI fragment is located directly upstream from the tk promoter (110bp from the transcription start site), CAT activity is increased an average of 41 or 44 fold, respectively, relative to pTE1. When pBR322 DNA sequences separate the chymotrypsin and amylase fragments from the tk promoter, CAT activity, though reduced, is still enhanced relative to pTE1 (3 and 11 fold, respectively). A construction in which the chymotrypsin fragment is located 1700bp downstream from the start site at the 3' end of the CAT gene also showed an increase in CAT activity relative to pTE1 (4 fold). At each position, the chymotrypsin and amylase fragments were active in either the normal or inverted orientation with respect to the

direction of transcription.

To determine whether the increase in CAT activity seen in the presence of the amylase and chymotrypsin fragments is due to an effect on transcription from the tk promoter rather than to intrinsic promoter activity in the DNA fragments used, the tk promoter region was deleted from selected constructions (Appendix I). Figure 5 shows the constructions tested and the CAT activity obtained when these plasmids were transfected into AR4-2J cells. CAT activity is expressed as a percentage of that obtained in parallel transfections with the parent plasmids containing the tk promoter. The average CAT activity of constructions lacking the tk promoter ranges from 3% to 12% that of the respective parent plasmid (approximately 2 to 5 fold greater than pTE1 activity). This indicates that at least about 90% of the CAT activity in constructions containing the chymotrypsin and amylase enhancers can be attributed to activated transcription from the tk promoter. The residual CAT activity produced in the absence of the heterologous promoter is likely to be due to use of cryptic promoter elements within the chymotrypsin and amylase fragments. In particular, the amylase fragment contains AT-rich regions, resembling the Goldberg-Hogness consensus sequence, which could act as substitute "TATA" box promoter elements. An analysis of RNA species produced in the presence of the SV40 enhancer supports this interpretation: the SV40 72bp repeats activated transcription from a number of cryptic promoters, some of which mapped to within the SV40 enhancer region (Wasylyk, et al, 1983).

To test directly whether the increase in CAT activity seen with the amylase and chymotrypsin fragments is due to enhancement of tk promoter function, I examined CAT mRNAs isolated from transfected cells. Since

only a small percentage of the cells (on the order of 1%) take up and express exogenous DNA, levels of CAT mRNA in a plate of transfected cells are very low. The CAT signals obtained with RSV.CAT in XC cells are approximately 10 fold greater than those in AR4-2J cells. (These factors emphasize the sensitivity of the CAT assay system: very low levels of expression of transfected DNAs can readily be measured.) Thus, it was not surprising that CAT mRNA produced in AR4-2J cells transfected with RSV.CAT is only barely detectable by primer extension, and CAT mRNA from AR4-2J cells transfected with pTE1 derivatives is not detectable by this method.

I have carried out several experiments to map CAT mRNAs from AR4-2J cells using radiolabeled Sp6 RNA probes (Appendix II). Transcripts from RSV.CAT in AR4-2J cells are easily mapped using this method. However, detection of probe protected by CAT mRNA isolated from cells transfected with pTE1 derivatives requires long exposures which also increase the intensity of bands presumably resulting from degradation of the probe. As shown in figure 6, protected fragments of the size expected for transcripts starting at the tk promoter cap site can be detected using RNA from AR4-2J cells transfected with pTE1 derivatives containing either the amylase or chymotrypsin enhancer inserted adjacent to the tk promoter.

The plasmids described above were also transfected into XC rat fibroblast cells. The results from a typical experiment of this kind are shown in figure 4. CAT activity is expressed relative to that directed by pTE1, containing the tk promoter alone (set at a value of 1). In general, the CAT activities of each plasmid assayed in XC cells ranges from 0.5 to 2.0, or from 2 fold lower to 2 fold higher than the activity

seen with pTE1. This range of values is probably a consequence of the intrinsic variability seen in the transfection experiments and in the "quality" of plasmid DNA isolated in different preparations. In fact, the average of values obtained in separate transfection experiments using plasmid DNA from different preparations tend to be closer to 1, i.e. equivalent to pTE1. Thus the amylase and chymotrypsin enhancer elements are completely inactive in rat fibroblasts. This cell type specificity is very dramatic: the CAT activity of constructions in which the amylase or chymotrypsin fragment is located directly upstream from the tk promoter in AR4-2J cells ranges up to 100 fold greater than that of the tk promoter alone in single transfection experiments (100 fold enhancement) while no enhancement is seen in XC cells.

I have also tested the CAT activity of several plasmids containing two copies of the chymotrypsin enhancer fragment in tandem or combinations of the amylase and chymotrypsin enhancers. The construction of these plasmids is described in Appendix I and the CAT activities obtained after transfection into AR4-2J cells are shown in figure 7. Activities are expressed relative to that of the tk promoter alone (pTE1) set at a value of 1. The degree of enhancement obtained with two different enhancers is slightly greater than the sum of the enhancements obtained with single copies of the amylase or chymotrypsin fragments.

Identification of Sequences Required For Enhancer Function: Mapping the Chymotrypsin Enhancer

A series of 5' deletions from the chymotrypsin 5' flanking region demonstrated that DNA sequences upstream from -225 do not contribute to the cell type specific effect and presumably also are not required for

enhancer activity of the chymotrypsin SstI fragment (Walker et al., 1983; Chapter 1). In contrast, a deletion of sequences between -225 and -192 effectively eliminated cell type specific expression. These results identify a region between -225 and -192 which is absolutely required for cell type specific activity in our assay.

In order to identify sequences required for activity in the downstream portion of the -275 to -93 enhancer fragment, deletions from the 3' end of this fragment were produced. The construction of these deletions is described in detail in Appendix I. The resulting plasmids contain the chymotrypsin fragment with deletions of 16 to 77bp from the 3' end. Each deletion end point was fused directly to the 5' end of the tk promoter fragment in pTE1. Deletion plasmids were transfected into AR4-2J cells, and CAT activity was compared to that obtained with the entire chymotrypsin SstI fragment in parallel experiments. The results of these transient assays are shown in figure 8. Deletion of 16bp, to -109, produces a fragment with increased enhancer activity relative to the -275 to -93 fragment. Further deletions from the 3' end, in general, caused a progressive decrease in enhancer activity. The increases in CAT activity seen upon deletion of sequences from -109 to -93 and from -150 to -137 could be explained by removal of sequences which block or inhibit enhancer activity. The most extensive deletion, of 77bp (3'd1-170), retained approximately 17% of the CAT activity of the -275 to -93 fragment. This represents about a 7 fold enhancement of activity directed by the tk promoter. These experiments define a second region, between -170 and -109, which is not essential for enhancer function but contributes to full activity.

Experiments carried out on the RSV enhancer (Laimins et al., 1984)

demonstrated that different regions of the "complete" enhancer are able to complement each other in pairs to create an active enhancer. Although a 5' deletion to -225 in the chymotrypsin fragment does not affect enhancer activity and a fragment with a 3' deletion to -170 retains some enhancer activity when these deletions are introduced separately, the region from -225 to -170 may not contain all sequences required for enhancer function. In other words, deletions from the 5' end may be complemented by an intact 3' region and vice versa. To see whether fragments with both 5' and 3' deletions show enhancer activity, the deletion to -225, the most extensive 5' deletion which does not affect enhancer activity, was combined with 3' deletions to -113, -137, -150, and -170bp. The 5' plus 3' deletion recombinants were constructed using an NcoI site at -192 in the chymotrypsin fragment and an NcoI site in the CAT coding region (see Appendix I). The results obtained after transfection of these plasmids into AR4-2J cells are shown in figure 8 (activities are expressed as a percentage of that seen with the -275 to -93 fragment). Except in the case of the -225 to -150 fragment, the 5' plus 3' deletion combinations were slightly more active than the corresponding 3' deletions preserving the -275 to -225 region. The 50bp region between -275 and -225 does not appear to contribute to enhancer activity, nor does it complement the loss of sequences from the 3' end of the enhancer fragment. The increased activity of fragments lacking the 50bp region may be due to deletion of sequences which decrease enhancer activity, inhibitory sequences or cryptic promoters which "divert" enhancer action. The 55bp sequence, -225 to -170, retains 27% of the total enhancer activity of the 182bp chymotrypsin SstI fragment. In transfection assays of the 55bp fragment, this constituted a 7 to 8

fold enhancement of CAT activity driven by the tk promoter. Since further deletions from the 3' end have not been produced and tested, it is not yet known whether the 55bp segment is the shortest sequence retaining enhancer activity.

Mapping the Amylase Enhancer

A 193bp RsaI fragment (-235 to -41) from the amylase 5' flanking region shows cell type specific enhancer activity. A series of 5' and 3' deletion mutants were generated by C. Erwin from pTE1 derivatives in which the amylase fragment is separated from the tk promoter by 600bp of pBR322 sequences. Sequences were deleted from the tk promoter-proximal end of the insert by treatment with exonuclease III and S1 nuclease (Appendix I). Each deletion end point was joined directly to the 5' end of the tk promoter sequences. Using this method, 5' or 3' deletions were generated from the pTE1 derivative with the amylase insert in the inverted or normal orientation, respectively. The deletion plasmids were transfected into AR4-2J cells and CAT activity was compared in each experiment to that obtained with the full 193bp amylase fragment. The results in figure 9 show that deletion of up to 80bp from the 5' end of the amylase fragment (to -154) does not lead to a reduction in enhancer activity. In fact, deletions with end points within this 80bp region show a higher level of CAT activity than the starting fragment. Possible explanations for the increase in activity again may be the removal of sequences which block or repress enhancer function. In sharp contrast, a deletion to -140 abolishes all enhancer function. Amylase DNA sequences between -154 and -140 are thus critical for enhancer activity.

In contrast to the effects on enhancer function seen with 5'

deletions, a 36bp deletion from the 3' end of the amylase fragment (to -77) results in the loss of about 70% of the activity. Fragments produced by deletion of up to 38 additional nucleotides maintain this reduced level of activity. However, further deletion to -159 decreases CAT activity to the background level (tk promoter alone). The results of 5' and 3' deletion analysis indicates that the 193bp amylase fragment contains two discrete enhancer domains. Sequences critical for enhancer function are located between -154 and -115. A second region, -77 to -41, is not critical for activity but contributes to the magnitude of enhancement obtained. When these latter sequences are deleted, enhancer activity is only 30-40% that of the complete RsaI fragment.

The relative activities of particular deletion mutants may reflect the effects of fragment orientation and enhancer/promoter spacing as well as the effects of deletion of specific DNA sequences. For example, the slightly higher activities seen for the 5' deletions to -154 could be the result of positioning critical enhancer sequences closer to the tk promoter or could be due to the deletion of inhibitory sequences. The reduction of enhancer activity observed when 3' sequences to -77 are deleted may result from moving critical regions too close to the tk promoter. In addition, the experiments described thus far do not exclude the possibility that the region between -77 and -41 contains an amylase transcription initiation (promoter) element which potentiates CAT expression. Such an element must either function at a relatively large distance from the tk promoter or be less important for attainment of full activity when the amylase fragment is in the inverted orientation.

To address some of these questions, four of the deleted amylase fragments were inverted so that the non-deleted end would then be joined

directly to the 5' end of the tk promoter region (Appendix I, experiments done in collaboration with C. Erwin). The activity of these constructions was compared to that of the corresponding parent plasmid in transfections carried out in parallel. The results shown in figure 10 indicate that the basic properties of the activity of both 5' and 3' deletion mutants are not significantly altered by fragment orientation: 5' deletion mutants are as active or more active and 3' deletion mutants are less active than the complete 193bp fragment. However, the 3' deletion mutants appear to be less active when the fragment is inverted (still at least 4-5 fold enhancement). This observation could be rationalized by postulating that sequences between -235 and -200 block enhancer action if placed between the enhancer and the promoter and that this effect is not manifested when the -77 to -41 region is present.

As in the case of the chymotrypsin enhancer discussed above, the most extensive deletions retaining enhancer activity made separately from the 5' or 3' end of the amylase fragment do not necessarily identify a minimal enhancer region. To determine whether DNA sequences defined by these deletion mutants show enhancer activity on their own, C. Erwin fused a synthetic 57bp oligonucleotide matching the amylase sequence from -164 to -109 directly to the 5' end of the tk promoter. When the 57bp fragment is inserted in either orientation, CAT activity directed by the tk promoter is significantly greater than that seen with the tk promoter alone. The enhancer activity obtained with the synthetic oligonucleotide is approximately 30% that of the 193bp amylase fragment in the normal orientation and 20% in the inverted orientation. The 57bp sequence is able to direct cell type specific expression: transfection of pTE1 derivatives containing the oligonucleotide into XC cells results

in CAT levels equal to those obtained with pTE1.

Mutational Analysis of the Amylase and Chymotrypsin Enhancer Regions

In order to further define the sequences required for cell type specific enhancer function, we have begun to test the role of specific short nucleotide sequences within the enhancer regions.

A unique NcoI site within the chymotrypsin enhancer fragment at -192 allowed the rapid construction of a 4bp deletion and a 4bp duplication at this position in the chymotrypsin-CAT fusion plasmid (pBR.Cht.CAT) containing the 711bp chymotrypsin 5' flanking region (Appendix I). Figure 11b shows the sequence in the altered region in the deletion and duplication mutants and the CAT activities obtained, expressed as a percentage of that for the wild type 5' flanking region. While the duplication of 4bp produces a slight reduction in CAT activity (to about 60% that of the wild type sequence), the 4bp deletion causes a drastic loss of enhancer activity (to about 2% wild type). These results may suggest that the 4bp which are deleted are critical for enhancer activity, either directly as part of a protein recognition sequence or indirectly as essential spacing between functional elements. Since duplication of the same 4bp does not greatly reduce enhancer activity, a presumptive factor binding site may be only slightly disrupted by the addition of 4bp. Alternatively, an additional 4bp in a spacer sequence may have a less drastic effect than a deletion of 4bp.

The synthetic oligonucleotide comprising the minimal amylase enhancer was constructed by C. Erwin using 8 single stranded complementary oligonucleotides (joined together with overlapping ends) to allow facile construction of mutated derivatives. Only two of the

short oligonucleotides must be synthesized for each new mutation. C. Erwin has produced and tested several mutants, and the results are summarized in figure 11a. Alteration of the sequence TTCC at position -155 to -152 to CAGT results in a cell type specific enhancer which is active in the normal orientation but completely inactive in the inverted orientation. In contrast, replacement of the sequence GAGA at position -149 to -146 with TCAG or replacement of TGTG at -125 to -122 with CAGT abolishes enhancer activity when the fragment is in either orientation. Thus, C. Erwin has begun to identify short regions within the minimal amylase enhancer which are critical for activity and for bidirectional function.

## Discussion

The rat amylase and chymotrypsin genes contain enhancer elements which function in exocrine cells of the pancreas and are completely inactive in fibroblasts. These enhancers act on the heterologous tk promoter when located at various distances from the transcription initiation site, either 5' or 3' of the CAT gene and in either the normal or inverted orientation. The distinctive feature shared by the amylase and chymotrypsin enhancers is their specificity for cells which express the endogenous amylase and chymotrypsin genes. The information available to date suggests that cell type specific enhancers, including those described for pancreatic exocrine genes, function in the same manner as viral enhancer elements. The functional characteristics of cell type specific and viral enhancers are very similar: for example, the activity of the MSV enhancer at various locations in pTE1 parallels the activity of the exocrine or endocrine cell enhancers at the same

positions (Edlund et al., submitted).

Sequences contributing to enhancer activity are located between -154 and -41 in the amylase 5' flanking region and between -225 and -109 in the chymotrypsin upstream region. Deletion analysis of these elements indicates that each is composed of sequences critical for enhancer function and discrete sequences which are not essential but augment enhancer activity.

Many experiments in which enhancer regions have been analyzed by extensive deletion mapping and dissection into subfragments have suggested that these elements are composed of multiple domains. The activity of a single copy of the SV40 72bp repeat is decreased at least 15 fold by deletion of 17bp from the 5' end or 23bp from the 3' end (Moreau et al., 1981; Hen et al., 1982; Sassone-Corsi, 1985). A deletion of 21 or 22bp from the center of the 72bp repeat reduces enhancer activity but to a lesser extent (Moreau et al., 1981; Hen et al., 1982). Herr and Gluzman (1985) have proposed that the SV40 enhancer is composed of at least three elements: the enhancer "core" region (Weiher et al., 1983) and two stretches of alternating purine and pyrimidine residues which are thought to contribute to the formation of Z DNA. A single copy of the 72bp repeat carrying mutations in each of the regions of alternating purines and pyrimidines is only weakly active as an enhancer. However, enhancer activity can be recovered by tandem duplication of 45 to 135bp in the enhancer region. A 15bp region of the enhancer, including the enhancer "core" sequence, is duplicated in all revertants.

Analysis of several single and multiple point mutations in the 40bp enhancer region of bovine papilloma virus (BPV) identified two critical

regions (Weiher & Botchan, 1984). Laimins et al. (1984) have shown that the Rous sarcoma virus (RSV) enhancer contains three distinct regions which are not active enhancers on their own but are active when adjacent domains are combined in pairs. In addition, duplication or triplication of the middle domain produces an enhancer with 30% or 100%, respectively, the activity of a pair of adjacent domains.

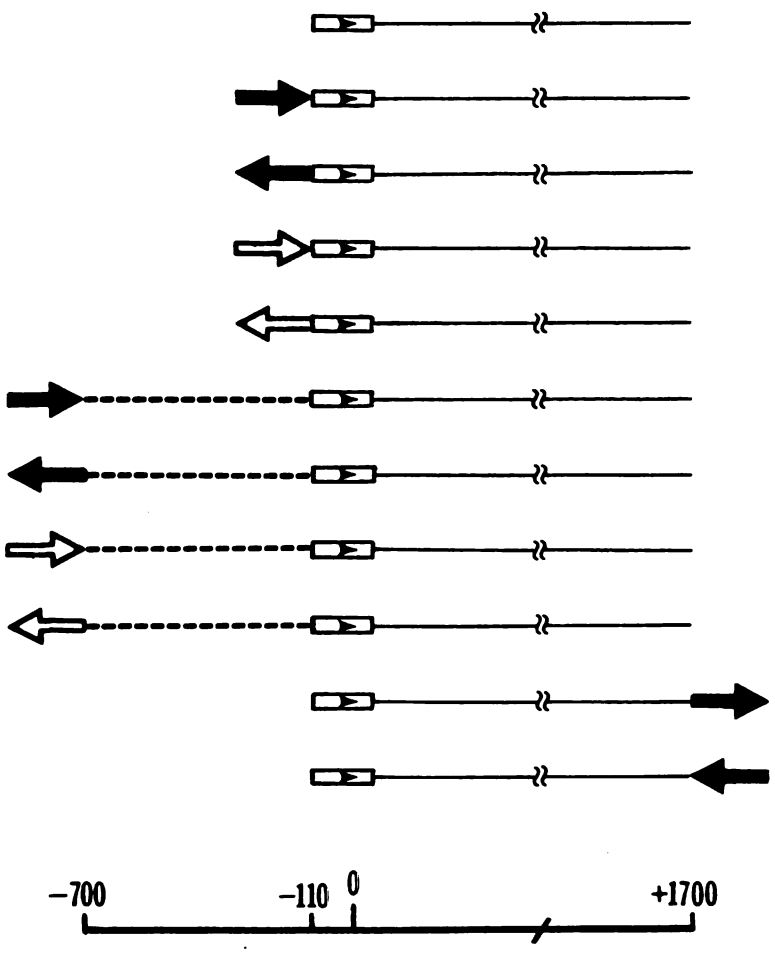
While the RSV enhancer domains function only when paired with the adjacent domain or when duplicated or triplicated (in the case of domain B), the polyoma virus enhancer region can be separated into at least two distinct active enhancer fragments (Herbomel et al., 1984). Element A is 3 fold more active as an enhancer (acting on the alpha 2 collagen promoter) than element B in 3T6 mouse fibroblast cells. Whereas element A is less active in mouse embryonal carcinoma cells than in fibroblasts by a factor of 3.5, element B shows approximately the same enhancer activity in the two cell types. A single point mutation within element B which increases enhancer activity about 2.5 fold (in both cell types) renders element B more active than element A in embryonal carcinoma cells. Weber et al. (1984) determined that a fragment with a duplication of the polyoma sequences containing element A was able to enhance expression from the beta globin promoter to nearly the same degree as the complete polyoma enhancer region (containing elements A and B).

The importance of cellular factors in enhancer function was first suggested by the host cell specificity of viral enhancers. The interaction of the glucocorticoid receptor with the hormone-responsive enhancer element in mouse mammary tumor virus (Chandler et al., 1983; Payvar et al., 1983) and the results of *in vivo* competition experiments (Scholer & Gruss, 1984) support this view. With this concept in mind, a

number of possible roles can be proposed for multiple enhancer domains. Binding of a single factor to two or more sites may be required for optimal enhancer activity. Multiple domains could also imply that two or more factors, each binding to a separate domain, are necessary for enhancer function. Alternatively, certain enhancer domains may not act as factor recognition sites but may, for instance, help to establish some feature of enhancer region DNA or chromatin structure which contributes to factor binding or to production of the enhancer effect. If enhancers do contain recognition sites for different factors or regions with different functions, it would appear that the loss of one component can be compensated by duplication of another enhancer domain. (Of course, what has been identified as a single domain may perform multiple critical functions.)

The downstream domains of the amylase and chymotrypsin enhancers, which contribute to but are not absolutely required for enhancer activity, are reminiscent of the domains identified in the RSV and SV40 enhancers. We have not yet tested duplications of these regions for enhancer activity, nor have we carried out experiments to determine whether they will contribute to the activity of heterologous enhancers. The isolated upstream domains, which are critical for enhancer function, show approximately 30% the activity of the -235 to -41 amylase fragment and the -275 to -93 chymotrypsin fragment, respectively. Again, we have not assayed the activity produced after duplication of these regions.

FIGURE 4: CAT activities directed by the tk promoter in the presence of the chymotrypsin SstI fragment (-275 to -93) or the amylase RsaI fragment (-235 to -41) at various plasmid locations. The scale at the bottom shows distance and position relative to the tk cap site. Arrows indicate the orientation of the inserted fragments; solid arrows and open arrows represent the chymotrypsin and the amylase fragment, respectively. The open box with a small arrow represents the tk promoter, the solid line represents the CAT coding region and SV40 sequences, and the dashed line represents pBR322 sequences. Activities in AR4-2J and XC cells are expressed relative to that obtained with the tk promoter alone (pTE1), set at a value of 1. Activities in AR4-2J cells are averages obtained from several transfection experiments. XC activities are derived from a single representative transient assay.




| Relative CAT Activity | |
|---|---|
| AR4-2J | XC |
| 1 | 1 |
| 41 | 1.1 |
| 48 | 1.4 |
| 44 | 1.2 |
| 33 | 1.3 |
| 3 | 0.6 |
| 2 | 0.6 |
| 11 | 0.6 |
| 10 | 0.9 |
| 4 | 1.1 |
| 5 | 1.4 |

FIGURE 4

FIGURE 5: Promoter dependence of amylase and chymotrypsin upstream enhancers. AR4-2J cells were transfected with recombinant plasmids in which the amylase RsaI (-235 to -41) or the chymotrypsin SstI (-275 to -93) fragment is fused directly to the CAT coding region by deletion of the tk promoter sequences. CAT activity is expressed as a percentage of that obtained in parallel transfections with the corresponding plasmids containing the tk promoter. Arrows signify fragment orientation; solid arrows represent the -275 to -93 chymotrypsin fragment and open arrows represent the amylase -235 to -41 fragment.

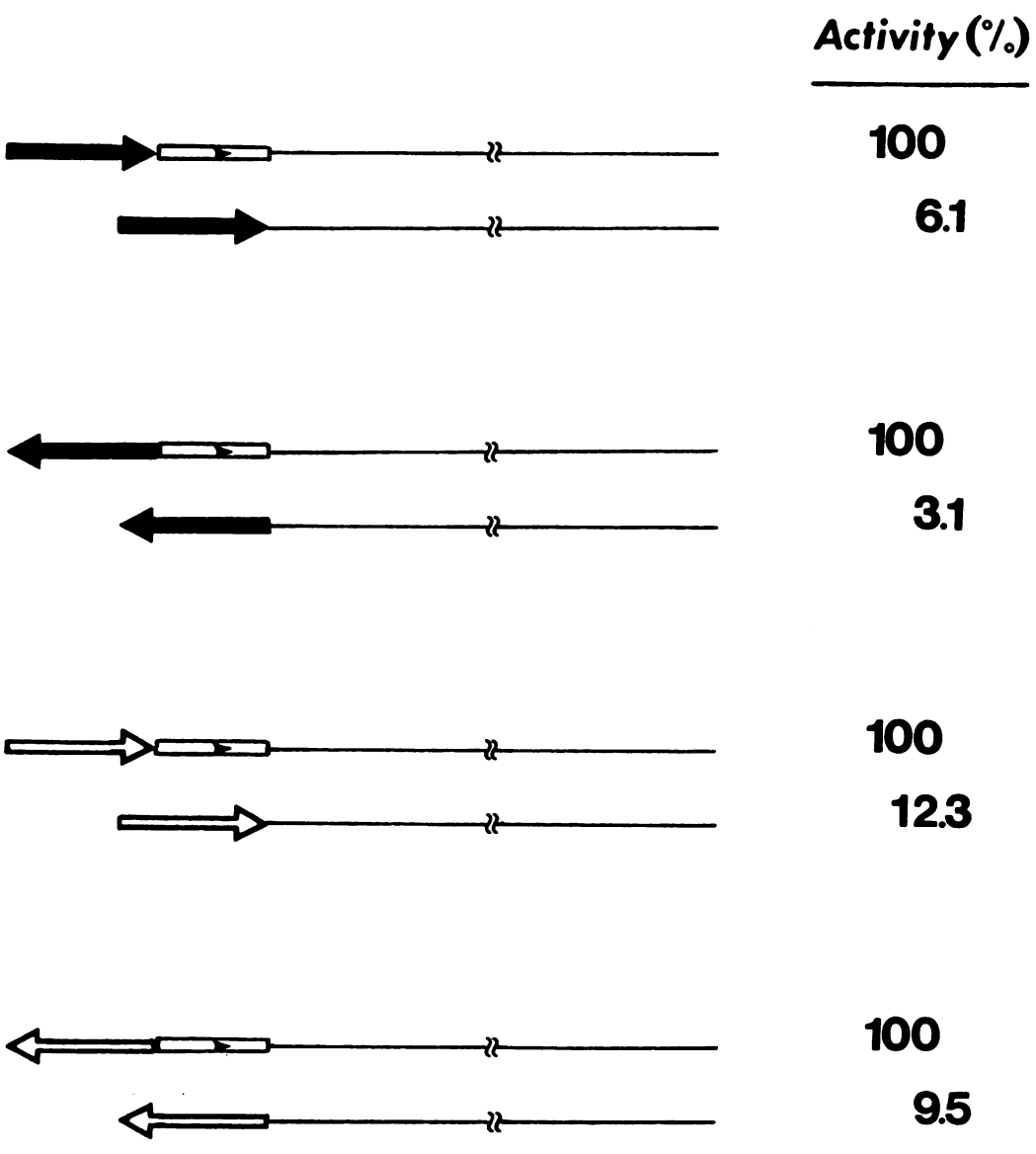
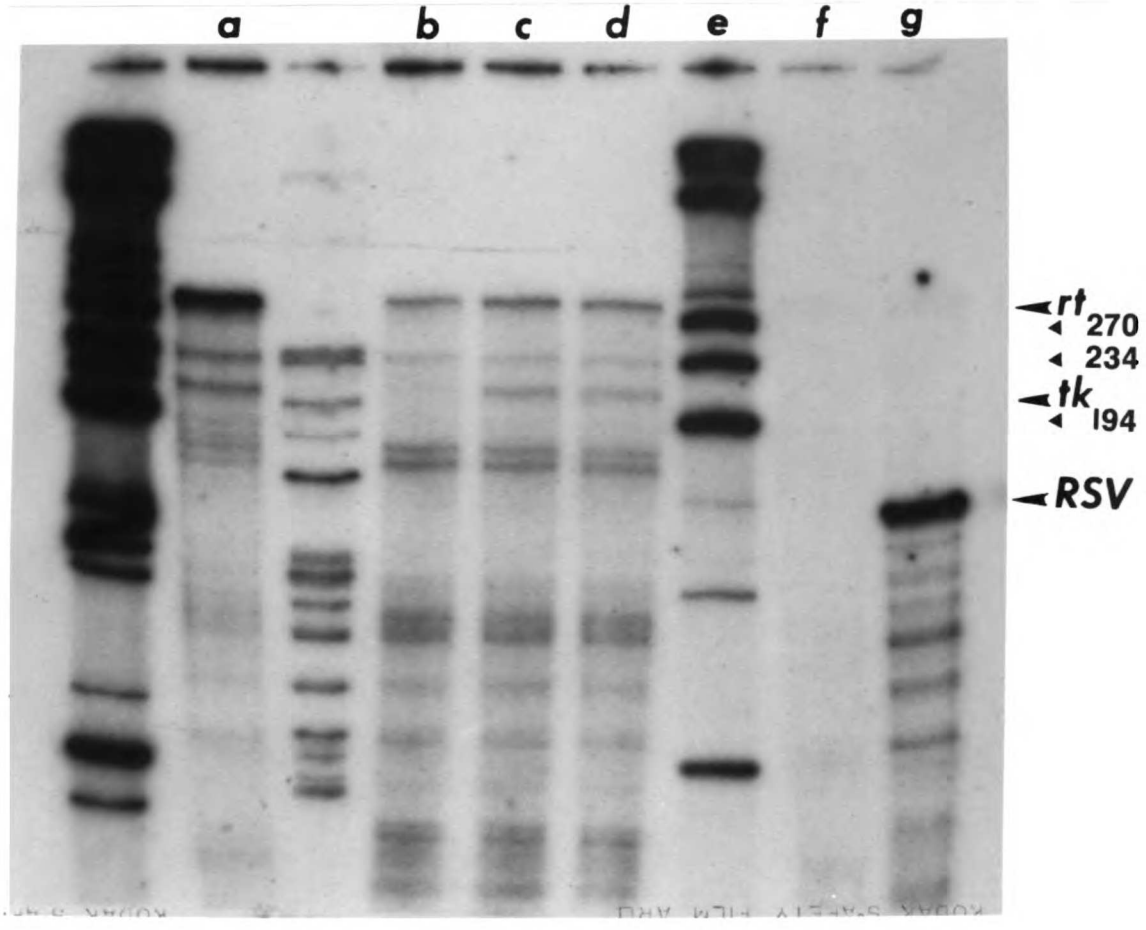



FIGURE 5

FIGURE 6: Analysis of CAT RNA transcripts in transfected cells using Sp6 RNA probes. Twenty micrograms of total RNA from transfected XC or AR4-2J cells was hybridized to a $^{32}$P-labeled Sp6 RNA probe complementary to tk-CAT sequences and digested with RNases A and T1. Protected RNA fragments were analyzed on a 5% polyacrylamide-urea gel. RNA samples were isolated from a) XC cells transfected with a pTE1 derivative in which CAT transcription is driven by the tk promoter, b) AR4-2J cells transfected with a plasmid containing the tk promoter alone, c) AR4-2J cells transfected with a plasmid containing the amylase RsaI fragment (-235 to -41) fused directly to the tk promoter, d) AR4-2J cells transfected with a plasmid containing the chymotrypsin SstI fragment (-275 to -93) fused directly to the tk promoter, e) size markers, f) untransfected AR4-2J cells, and g) AR4-2J cells transfected with RSV.CAT. The diagram below the figure shows the structure of the Sp6 template, the RNA probe transcribed after digestion of the template with EcoRI, RNA species expected in transfected cells, and the fragments generated after hybridization and RNase treatment. Regions of RNA not homologous to the probe are indicated by dotted lines. The band in the autoradiogram designated "RT" corresponds to the fragment protected by a readthrough RNA initiated upstream from the tk promoter. The band designated "RSV" is generated from RNAs initiating at, and upstream from, the RSV promoter in RSV.CAT. The band labeled "tk" is generated from transcription initiating at the tk cap site.


a
b
c
d
e
f
g
rt 270
234
tk 194
RSV

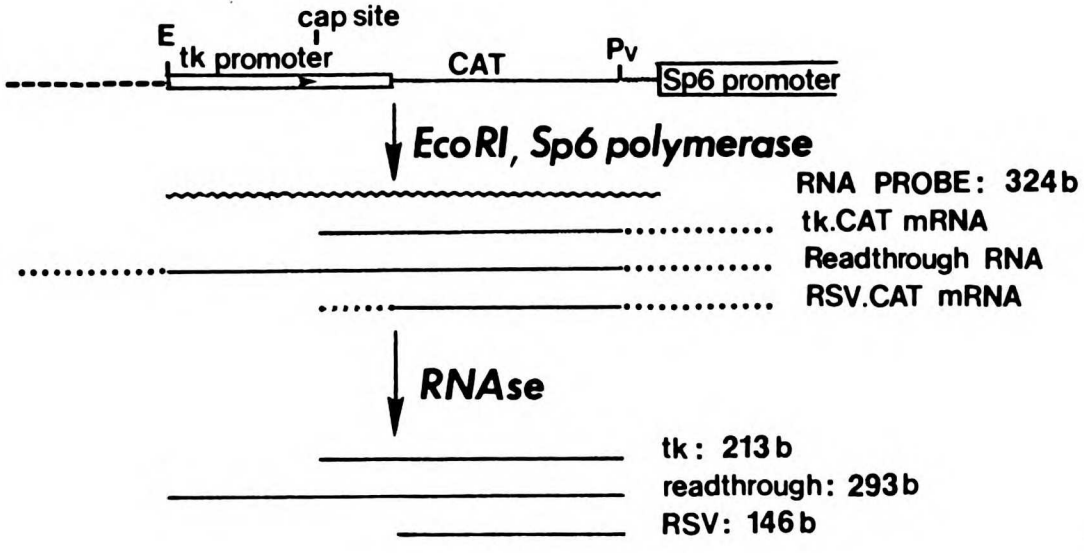

E
cap site
tk promoter
CAT
Pv
Sp6 promoter
EcoRI, Sp6 polymerase
RNA PROBE: 324b
tk.CAT mRNA
Readthrough RNA
RSV.CAT mRNA
RNAse
tk: 213b
readthrough: 293b
RSV: 146b

FIGURE 6

FIGURE 7: Effect of combinations of the amylase and chymotrypsin enhancers on CAT activity directed by the tk promoter. Activities are expressed relative to that obtained with the tk promoter alone set at a value of 1. Average CAT activities for the two groups of constructions were derived from separate series of experiments. Arrows indicate orientation of the inserted fragments; solid and open arrows represent chymotrypsin and amylase enhancers, respectively.

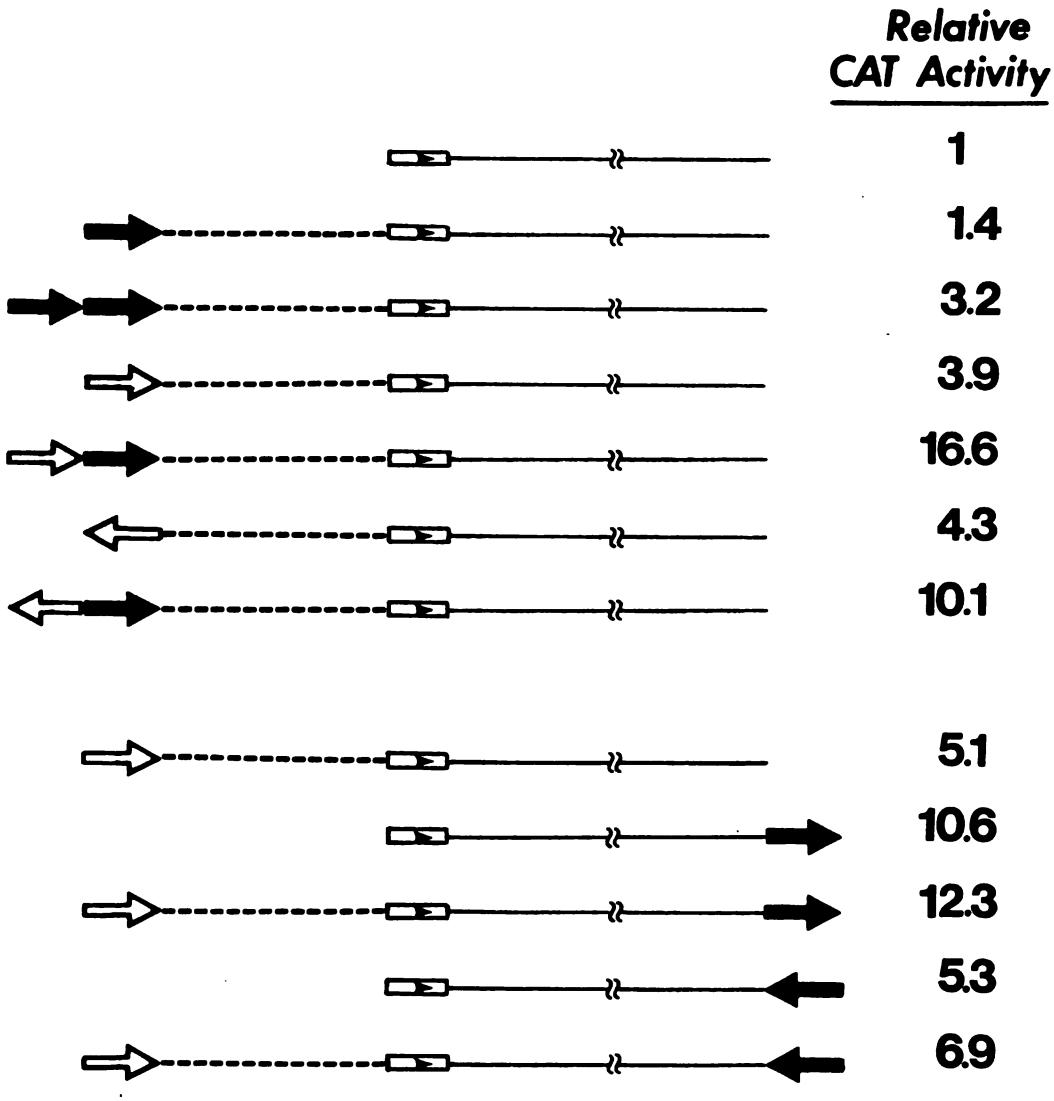



FIGURE 7

FIGURE 8: CAT activities directed by the tk promoter in the presence of chymotrypsin fragments containing 3' deletions or both 5' and 3' deletions. Bars represent retained regions of the chymotrypsin SstI fragment (-275 to -93), and numbers above the bars indicate deletion end points with respect to the chymotrypsin cap site. Activities are expressed as a percentage of that obtained when the intact -275 to -93 chymotrypsin fragment is located directly upstream from the tk promoter.

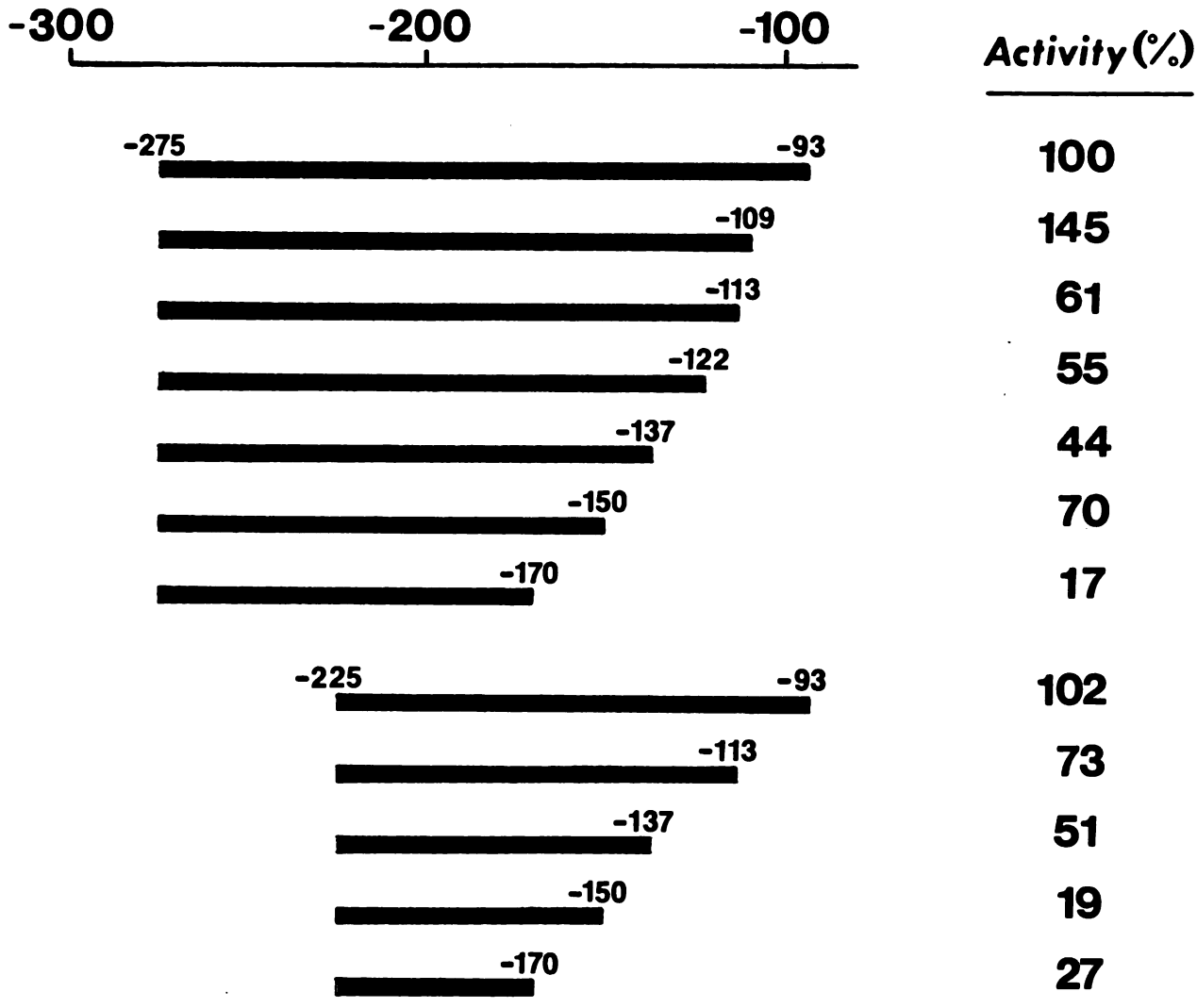

-300
-200
-100
Activity (%)
-275
-93
100
-109
145
-113
61
-122
55
-137
44
-150
70
-170
17
-225
-93
102
-113
73
-137
51
-150
19
-170
27

FIGURE 8

FIGURE 9: CAT activities directed by the tk promoter in the presence of amylase fragments containing 5' or 3' deletions. Bars represent retained amylase sequences and numbers above the bars indicate deletion end points in terms of nucleotide position with respect to the amylase cap site. Activities are expressed relative to that obtained with the intact -235 to -41 amylase RsaI fragment inserted directly upstream from the tk promoter. The bottom line shows the CAT activity produced when a 57bp oligonucleotide containing the indicated amylase upstream sequence is fused directly to the tk promoter in pTE1.

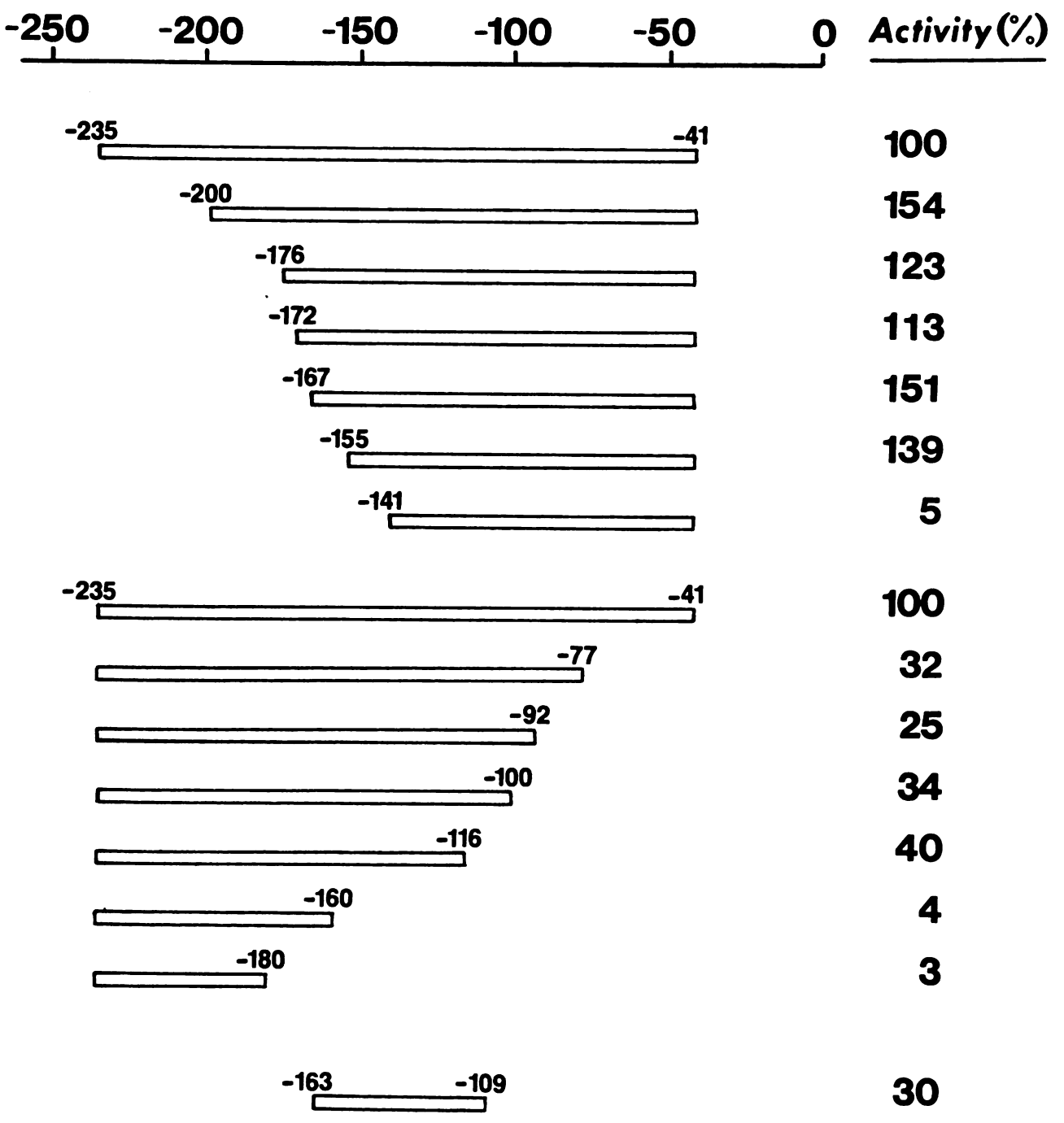

-250
-200
-150
-100
-50
0
Activity (%)
-235
-41
100
-200
154
-176
123
-172
113
-167
151
-155
139
-141
5
-235
-41
100
-77
32
-92
25
-100
34
-116
40
-160
4
-180
3
-163
-109
30

FIGURE 9

FIGURE 10: Effect of orientation on enhancer activity of amylase fragments containing 5' or 3' deletions. Arrows denote the orientation of amylase fragments and numbers indicate fragment end points in terms of nucleotide position with respect to the amylase cap site. Activities are expressed as a percentage of that obtained with the intact -235 to -41 amylase fragment in the orientation in which the deletion mutants were constructed.

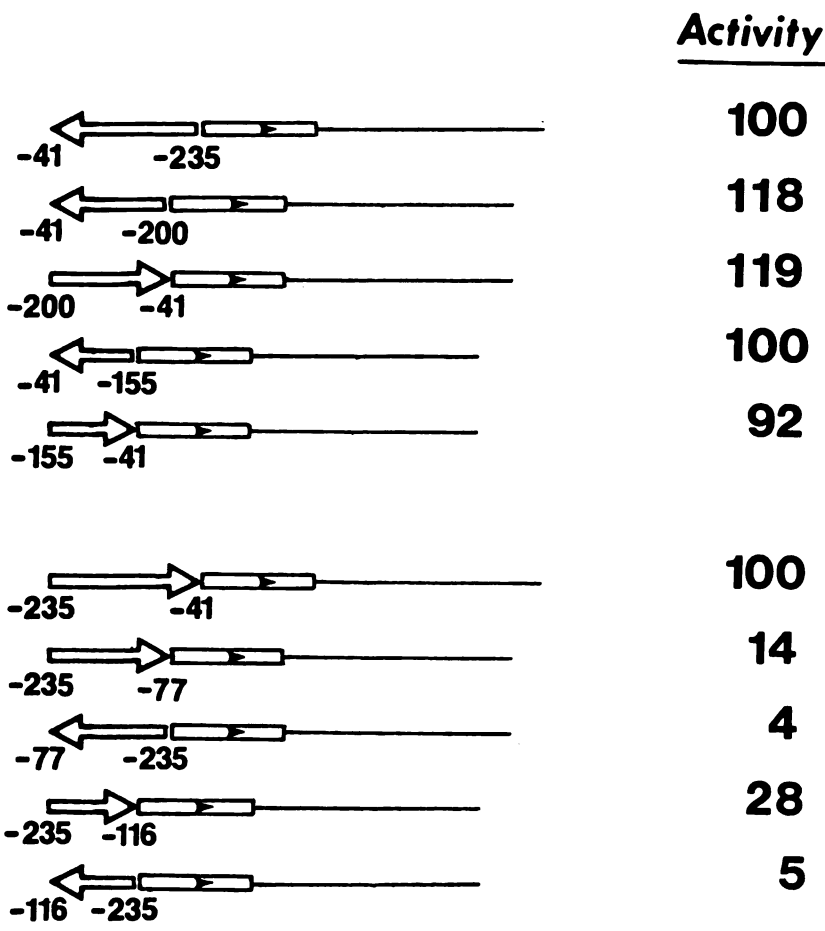

Activity (%)
-41
-235
100
-41
-200
118
-200
-41
119
-41
-155
100
-155
-41
92
-235
-41
100
-235
-77
14
-77
-235
4
-235
-116
28
-116
-235
5

FIGURE 10

FIGURE 11: Effect of internal mutations on enhancer activity of the chymotrypsin 5' flanking region (-711 to -3) and the amylase 57bp oligonucleotide.

A. Amylase mutants: Altered nucleotides are underlined and their positions with respect to the amylase cap site are shown. Activities are expressed as a percentage of that obtained with the wild type amylase enhancer oligonucleotide in each orientation.

B. Chymotrypsin mutants: The region from -208 to -182 in the chymotrypsin 5' flanking region is shown. The triangle marks the site from which four nucleotides were deleted by S1 nuclease treatment as described in Appendix I. The four nucleotides duplicated by T4 polymerase treatment are underlined. Activities are expressed as a percentage of that obtained with the wild type chymotrypsin 5' flanking region (-711 to -3).

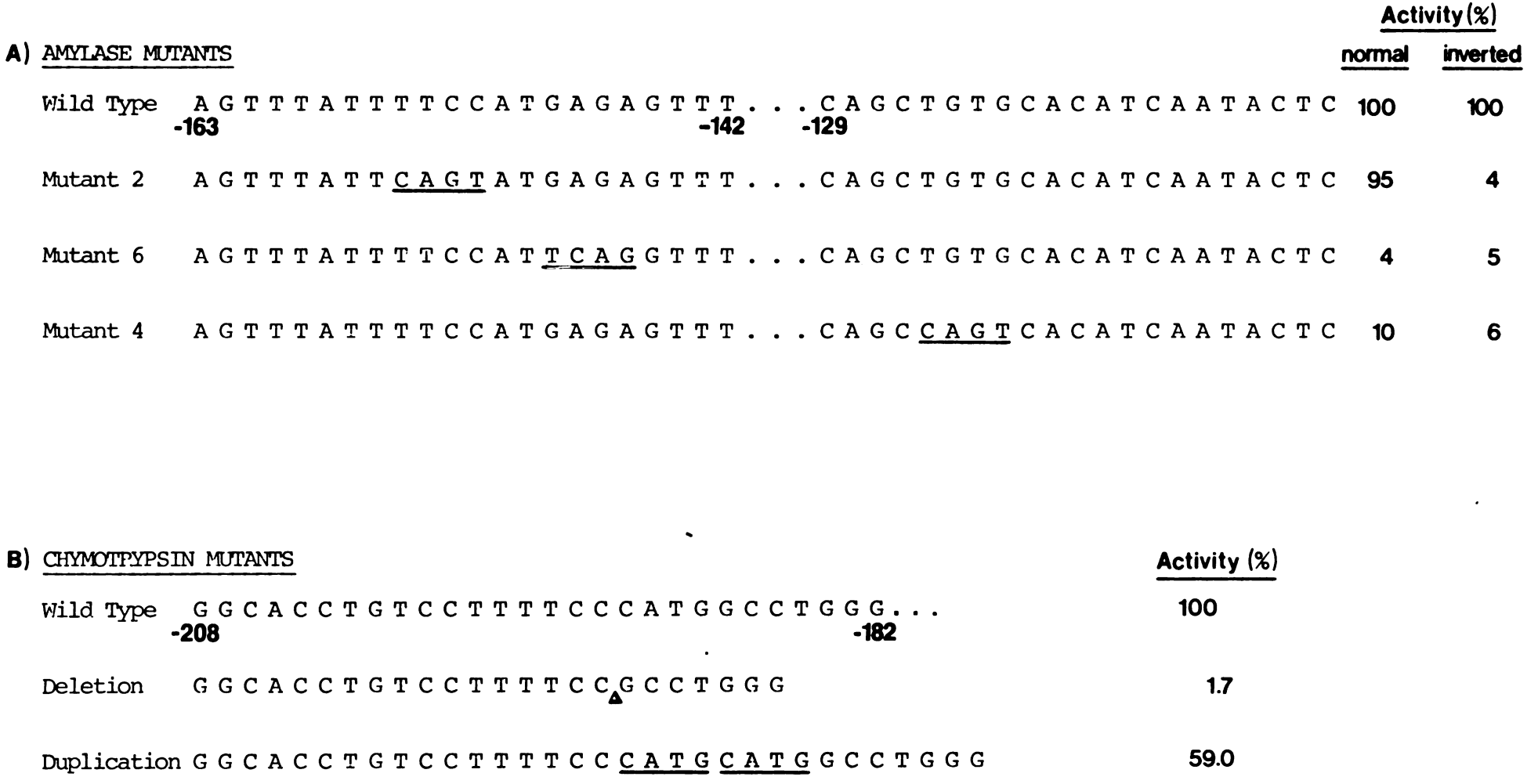

Activity (%)
A) AMYLASE MUTANTS
normal
inverted
Wild Type A G T T T A T T T T C C A T G A G A G T T T . . . C A G C T G T G C A C A T C A A T A C T C
-163
-142
-129
100
100
Mutant 2 A G T T T A T T C A G T A T G A G A G T T T . . . C A G C T G T G C A C A T C A A T A C T C
95
4
Mutant 6 A G T T T A T T T T C C A T T C A G G T T T . . . C A G C T G T G C A C A T C A A T A C T C
4
5
Mutant 4 A G T T T A T T T T C C A T G A G A G T T T . . . C A G C C A G T C A C A T C A A T A C T C
10
6
B) CHYMOTPYPSIN MUTANTS
Activity (%)
Wild Type G G C A C C T G T C C T T T T C C C A T G G C C T G G G . . .
-208
-182
100
Deletion G G C A C C T G T C C T T T T C C G C C T G G G
1.7
Duplication G G C A C C T G T C C T T T T C C C A T G C A T G G C C T G G G
59.0

FIGURE 11

# CHAPTER 3

# COORDINATE CONTROL OF PANCREATIC EXOCRINE GENE EXPRESSION

Introduction

The exocrine pancreas system is well-suited to the study of coordinate control of gene expression in a differentiated cell type. Several genes are expressed at high levels in the pancreatic exocrine or acinar cells, with single mRNA species comprising between 1 and 10% on the average, or as much as 25% for amylase mRNA, of the polyA+ RNA. The results of experiments described above show that enhancer elements in the 5' flanking regions of the amylase and chymotrypsin genes direct high level cell type specific expression in an exocrine pancreatic cell line. We have proposed that trans-acting positive regulatory factors interact with these control elements to activate gene expression in pancreatic endocrine or exocrine cells (Walker et al., 1983). In vivo dimethyl sulfate protection experiments (Ephrussi et al., 1985; Church et al., 1985) and in vivo competition experiments (Mercola et al., 1985) provide evidence that immunoglobulin heavy chain enhancer sequences bind a factor or factors present only in cells of the B lineage. Pancreatic exocrine cells may contain analogous factors which specifically recognize the amylase and chymotrypsin enhancer sequences.

Models which can be proposed for the control of coordinate cell type specific expression in the exocrine pancreas generally fall between two extremes. The expression of all exocrine genes could be controlled by the same factor or set of factors or each gene could be controlled by a unique factor or set of factors. In both cases, a second level of cell type specific regulation must be responsible for activation of the expression of genes encoding regulatory factors. In a situation in which each gene is controlled by a different factor, a mechanism for achieving

coordinate expression of all factor genes must also be proposed.

Several lines of evidence argue against a model in which all genes are controlled by the same factor or factors. The AR4-2J cell line produces high levels of many exocrine gene mRNAs, but the relative levels of expression of these mRNAs compared to adult rat pancreas vary from 42% for amylase down to 0.9% for trypsin mRNA. This may indicate that AR4-2J cells contain factors sufficient for nearly optimal amylase expression, but lack factors required for similar relative levels of trypsin gene expression. A pancreatic acinar cell tumor analyzed in our laboratory manifests yet another pattern of exocrine mRNA expression (Han et al., in preparation). It is not known to what degree the transformed nature and the probable aneuploidy of these cells contributes to the differences in relative mRNA levels.

The patterns of activation of the different exocrine genes during embryonic development are not identical. For instance, the levels of chymotrypsin, amylase, and carboxypeptidase A1 mRNAs begin to increase 2-3 days earlier than levels of trypsin and elastase mRNAs (Han & Rutter, in preparation), and the levels of trypsin and elastase I mRNA show a greater increase after birth than other mRNAs.

Some of the above observations could be explained by postulating that the different genes have different affinities for a single cell type specific factor. More plausible models for coordinate control in the exocrine pancreatic system would propose that a number of cell type specific regulatory factors or a single cell type specific factor and other factors which are, for example, developmentally regulated are involved. The expression of a particular gene may be controlled by a single factor or several factors acting together, and the sets of

factors interacting with different genes may be partly overlapping. For example, high level amylase gene expression may require only a single positive factor while optimal expression of the trypsin genes may require this factor plus additional factors. Further investigation of the sequences required for cell type specific expression, studies of factors which recognize and bind to these sequences, and identification of factors essential for expression in an in vitro system should give a better idea of the number of factors involved and the degree to which different genes share the same regulatory molecules.

I have initiated studies representing two distinct but complementary approaches to the question of coordinate control. The first approach involves direct examination of the sequences in the 5' flanking regions of the exocrine genes to identify primary DNA sequence homologies or structural similarities which may point to common regulatory elements. The role of homologous sequences or structural features can then be assessed by site-directed mutagenesis. Searches for homology have been focused to a large extent by the identification of sequences in the amylase and chymotrypsin gene 5' flanking regions involved in cell type specific expression. In the second approach, in vivo competition experiments are used to address questions concerning the abundance of positive regulatory factors and the ability of the regulatory regions of different genes to compete for factors which may be present in limited quantities in AR4-2J cells. Competition experiments carried out to date have focused mainly on optimization of the experimental system and will not be discussed here. In this chapter, I describe the results of an analysis of upstream sequences of nine pancreatic exocrine genes and discuss possible implications of the

regions of homology which have been identified.

## A Consensus Sequence in the 5' Flanking Region of Pancreatic Exocrine Genes

Regulatory regions often map to the 5' flanking regions of protein coding genes (Chandler et al., 1983; Hearing & Shenk, 1983; Walker et al., 1983; Karin et al., 1984; Renkawitz et al., 1984; Ornitz et al., 1985). The sequences contained within the 55bp and 57bp minimal enhancer fragments from the chymotrypsin and amylase 5' flanking regions, respectively, are known to be critical for cell type specific expression in our assay system. Thus, a search for a common factor recognition sequence which plays a role in the control of cell type specific expression can be narrowed down to these relatively short stretches of DNA sequence. Regions in nine pancreatic exocrine genes containing related sequences are aligned in figure 12. Some of these sequences show a striking degree of homology while others are less closely related. The nucleotides which are found at a particular position in more than half of the sequences (5 out of 10), comprise a consensus of 20bp in an uninterrupted sequence. Three additional conserved nucleotides are located a few bases upstream and downstream from the 20bp sequence. Two short regions of homology, identified in rat pancreatic serine protease genes by Swift et al. (1984) and Craik et al. (1984), overlap with the consensus region. The location of the 3' end of each sequence relative to the cap site of the respective gene is indicated at the right. Also shown is the number of nucleotides in each sequence which match the consensus sequence. Amylase sequence 1 matches the 5' half of the consensus sequence at 8 out of 10 positions, but matches the 3' half in

only 4 out of 10 positions. Amylase sequence 2 contains an 8 out of 10bp match to the 3' half with only 1 match in the 5' half. The consensus sequence of the chymotrypsin gene lies entirely within the 55bp minimal enhancer fragment. The 20bp consensus region of amylase sequence 1 and the 3' half of the amylase sequence 2 consensus are contained within the 57bp amylase minimal enhancer region. The entire elastase I consensus-related sequence is included in the 213bp of upstream sequences which are capable of directing pancreas specific expression in transgenic mice (Ornitz et al., 1985).

In general, the effects of enhancer mutations described in Chapter 2 support the proposed functional significance of the homologous sequences in the 5' flanking regions of pancreatic exocrine genes. The 4bp deletion in the chymotrypsin enhancer region which abolishes activity destroys a single match at the 3' end of the consensus sequence. One of the two 4bp block changes which completely abolish activity of the 57bp amylase enhancer, that at -125 to -122, affects nucleotides within the consensus region of amylase sequence 2. The wild type amylase sequence matches the consensus at these four positions, and the mutation changes each position to a nonconsensus nucleotide. The second inactive mutant has an alteration in a single consensus position of amylase sequence 1. A 4bp block mutation which affects the bidirectionality of amylase enhancer function (mutant 2) alters 4 consensus nucleotides in amylase sequence 2.

## Discussion

We have identified homologous sequences within the 5' flanking regions of nine genes expressed at high levels in the exocrine pancreas.

A consensus sequence of 20bp can be derived from comparison of the ten homologous regions. The locations of the amylase and chymotrypsin consensus sequences coincide with the regions required for cell type specific enhancer function (Chapter 2). Apart from the sequences shown in figure 12, there is no extensive homology between the amylase and chymotrypsin enhancer sequences.

The enhancer elements of several viruses and cellular genes do not manifest any overall sequence homology, but there has been much speculation on the possible significance of a number of short sequences found in several enhancer regions. Hypothetical roles for these enhancer "core" sequences include providing binding sites for specific regulatory proteins and contributing to the formation of secondary or tertiary structures important for enhancer activity (Weiher et al., 1983). Sequences related to the SV40 enhancer "core", 5'GTGG$^{AAA}_{TTT}$G3', are found in most of the viral enhancers and in immunoglobulin and class II $E_B$ gene enhancer regions (Weiher et al., 1983). However, many enhancers lack the "core" sequence (for example, E1A and rat insulin), and a deletion analysis of the kappa immunoglobulin gene enhancer region suggests that its "core"-related sequence does not play a major role in enhancer function (Queen & Stafford, 1984). The pancreatic exocrine gene consensus sequence does not contain this enhancer "core", but a related sequence, 5'GT$^{G}_{C}$CTTTTC3', can be identified. Other short conserved sequences have been found by Lusky et al. (1983) and Hearing and Shenk (1983). Nordheim and Rich (1983) have proposed that segments of alternating purine and pyrimidine residues, by contributing to the formation of Z DNA, may be important for enhancer activity.

Short repeated sequences have been implicated in the mechanisms for control of several sets of coordinately regulated genes (reviewed in Davidson et al., 1983). A 9bp consensus sequence is found in multiple copies upstream from four yeast genes encoding enzymes involved in amino acid biosynthesis which are induced during amino acid starvation. The functional significance of these sequences has been demonstrated for both the his3 (Struhl, 1982b) and his4 (Donahue et al., 1983) genes. Pelham (1982) has shown that a 15bp consensus sequence, located upstream from Drosophila heat shock genes, is involved in heat induction of these genes. The degree of homology to the consensus sequence in the two systems ranges from 54 to 92% for the best consensus match found upstream of a particular gene.

Single or multiple copies of short homologous sequences also appear to be involved in transcriptional activation by glucocorticoid hormones (Payvar et al., 1983; Karin et al., 1984) and heavy metals (Karin et al., 1984; Stuart et al., 1984; Carter et al., 1984). The importance of consensus sequences found within glucocorticoid response elements is supported by in vitro glucocorticoid receptor binding studies (Payvar et al., 1983) and in vivo assays of wild type and mutant response elements (Chandler et al., 1983; DeFranco et al., 1984). Glucocorticoid response elements act as hormone-dependent transcriptional enhancers (Chandler et al., 1983; Majors & Varmus, 1983; Yamamoto, 1983). It has been suggested that specific hormone receptor:DNA interactions activate the GRE enhancer (Payvar et al., 1983). If a similar mechanism is utilized in other systems, one might expect to find homologous sequences in the vicinity of several genes which are regulated by a common DNA binding factor.

Church et al. (1985) have derived an immunoglobulin enhancer consensus sequence based on comparisons of sequences showing B cell specific DMS protection patterns and sequences of other established or putative immunoglobulin enhancer regions. Since single copies of a sequence similar to the 10bp consensus sequence are found in pBR322 and SV40, the authors consider it unlikely that this element alone is sufficient for enhancer activity.

The presence of short homologous sequences in the nine pancreatic exocrine genes examined, the location of these sequences, and the demonstration that specific nucleotides within the consensus region are critical for cell type specific enhancer function suggest that these sequences play some role in the control of cell type specific gene expression. However, these short sequences are almost certainly not the sole component involved in regulation of exocrine pancreatic genes. It is clear that the degree of relatedness to the consensus sequence does not correlate well with the levels of expression of the nine genes in the adult rat pancreas (see Chapter 1). Data described in Chapter 2 indicated that DNA sequences outside of the minimal enhancer regions, and therefore outside of the consensus sequence regions, of the amylase and chymotrypsin genes contribute to full cell type specific enhancer function. Thus, sequences in addition to those within the consensus region help to establish the strength and properties of the cell type specific enhancers.

FIGURE 12: Comparison of homologous sequences in the 5' flanking regions of nine pancreatic exocrine genes. The positions of the 3' ends of each sequence relative to the cap site are shown at the right. Nucleotides present at a particular position in 5 or more sequences are tabulated, with the number of occurrences designated below. Also included are pairs of nucleotides which each occur in 4 of the sequences. The consensus sequence derived from consecutive conserved nucleotides is shown at the bottom. The last column lists the number of matches to the 20 nucleotide consensus sequence in the homologous region of each gene.

| GENE | | DISTANCE FROM CAP SITE (bp) | MATCHES TO 20bp CONSENSUS |
|---|---|---|---|
| AMYLASE(1) | A A C C T T C A G C T G T G C A C A T C A A T A C T C C T | -107 | 12 |
| AMYLASE(2) | A A A G A A T C A A G T T T A T T T T C C A T G A G A G T | -144 | 10 |
| CARBOXY A1 | T C T C T C C A C C T G C C T T G T T C C C C G A T A C T | -142 | 14 |
| CARBOXY A2 | A T C A T A A A C G T C T G A T T T G C C C T T G G T A T | -213 | 13 |
| CHT B | T C A G G G C A C C T G T C C T T T T T C C C A T G G C C T | -185 | 17 |
| ELASTASE I | T C A T G T C A C C T G T G C T T T T T C C C T G C C T T C | -90 | 20 |
| ELASTASE II | T T T G G A C T C C A G C C C T T T A T T C C A C T G G G | -89 | 12 |
| RNase | T T G G G T T T C A A T G G C T T T T T C C T T G G C A G C | -145 | 13 |
| TRYPSIN I | C C T T G T C A C C T G T A G G T C T C C A A G T G A C A | -159 | 14 |
| TRYPSIN II | T T T C A A G A A T C A C C T A C T T T C C T T G T C C T | -164 | 8 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | | G | T | C | A | C | C | T | G | T | G | C | T | T | T | T | C | C | C | T | G | | | C | T |
| | 6 | | 5 | 4 | 6 | 7 | 7 | 6 | 6 | 6 | 6 | 4 | 5 | 7 | 7 | 8 | 8 | 8 | 8 | 6 | 6 | 5 | | | 5 | 6 |
| | | | | A | | | | | | | | C | | | | | | | | | | | | | | |
| | | | | 4 | | | | | | | | 4 | | | | | | | | | | | | | | |
| CONSENSUS | | | G | T/A | C | A | C | C | T | G | T | G/C | C | T | T | T | T | C | C | C | T | G | | | | |

FIGURE 12

# CONCLUSIONS AND PERSPECTIVES

I have identified sequences in the 5' flanking regions of the rat amylase and chymotrypsin genes which elicit preferential expression of the CAT reporter function in cells expressing the endogenous amylase and chymotrypsin genes at high levels. These sequences stimulate transcription from the homologous promoters and from the HSV thymidine kinase promoter. The amylase and chymotrypsin elements are cell type specific enhancers: they function in a relatively position and orientation independent manner, but their activity is restricted to cells of the exocrine pancreas. Sequences critical for enhancer function are located approximately 100 to 200bp upstream of the transcription start sites in the two genes.

In a number of systems, short stretches of DNA sequence homology are located in upstream regions of genes sharing similar regulatory responses (reviewed in Davidson et al., 1983). A 5' flanking region consensus sequence for pancreatic exocrine genes can be derived based on comparison of related sequences in nine genes. The location of the consensus-related sequences with respect to the transcription start sites and the correlation between the consensus regions and sequences known to be involved in cell type specific regulation suggest that these sequences are functionally significant. Several lines of evidence indicate that the consensus-related sequences are not sufficient for the establishment and maintenance of cell specific expression.

M. Walker and T. Edlund have recently demonstrated that sequences between the insulin enhancer and the cap site are able to promote endocrine cell specific expression under the appropriate conditions (Edlund et al., submitted). Thus, it appears that an initiation element within the promoter region may be recognized in a cell specific manner by the transcription machinery. However, the possibility that this element is a component of the

insulin enhancer which is not active on its own cannot be ruled out completely. Preliminary experiments using the chymotrypsin and the amylase promoters indicate that similar cell specific elements are not present in these genes. In the case of the amylase gene, sequences contributing to enhancer activity extend into a region which in other genes contains essential promoter elements (McKnight & Kingsbury, 1982; Everett et al., 1983; Dierks et al., 1983). This feature of the amylase 5' flanking region makes it difficult to distinguish between distal promoter elements and enhancer sequences. Since the -77 to -41 region contributes to enhancer activity when the 193bp RsaI fragment is inserted in either orientation upstream from the tk promoter, it appears to constitute a part of the enhancer rather than the amylase promoter. It is possible that the amylase "promoter" does not contain distal elements: the amylase enhancer may require only proximal promoter elements, such as the TATA sequence, for function.

Gene Regulation: Common Themes

Many of the basic themes of gene regulation have been conserved in evolution. In particular, examples of the utilization of both positive and negative regulatory mechanisms and of sequence specific DNA binding proteins can be found in bacteria, yeast, and complex multicellular organisms. The cell type specific enhancers described here and elsewhere (Gillies et al., 1983; Banerji et al., 1983; Queen & Baltimore, 1983; Neuberger, 1983; Picard & Schaffner, 1984; Gillies et al., 1984; Edlund et al., submitted) appear to be elements of positive control systems: deletion of enhancer sequences leads to a loss of cell type specific expression rather than to an increase in non-specific expression. However, the

existence of regulatory sites involved in the repression of transcription in non-expressing cells cannot be ruled out. The inhibition of TAT expression in cell hybrids (Killary & Fournier, 1984) suggests that repressors of cell type specific genes may be present in fibroblasts.

The involvement of a number of sequence-specific DNA binding proteins in positive and negative regulation of gene expression in eukaryotes has been established (Tjian, 1981; Payvar et al., 1983; Parker & Topol, 1984; Giniger et al., 1985; Johnson & Herskowitz, 1985). Similarly, the lambda repressor and cro recognize and bind to specific DNA sequences to manifest their positive and negative effects on $P_R$ and $P_{RM}$ (Ptashne et al., 1980). The correspondence between the MTV hormone-responsive enhancer region and sites of sequence-specific binding by the glucocorticoid receptor indicates that enhancer activity may be mediated by DNA binding proteins (Payvar et al., 1983; Chandler et al., 1983). Sequences recognized by the glucocorticoid receptor do not show an extensive amount of homology, but common sequence elements can be identified (Payvar et al., 1983). The set of homologous sequences found upstream of several Drosophila heat shock genes are thought to be involved in the coordinate regulation of these genes and are presumably recognized by a common DNA binding protein (Pelham, 1982; Parker & Topol, 1984). B cells appear to contain a factor which affects the dimethyl sulfate reactivity of specific regions of the immunoglobulin enhancer (Ephrussi at al., 1985; Church et al., 1985). Sequences homologous to the protected regions can be found in other established or putative Ig enhancers. Although we have only very preliminary evidence that pancreatic exocrine cells contain factors which recognize specific sequences in the exocrine enhancer regions (E. Fodor, unpublished), the 5' flanking region consensus sequences show a degree of

homology comparable to that of proposed recognition sites in other systems (Pelham, 1982; Davidson et al., 1983; Payvar et al., 1983). Obviously, this evidence is circumstantial and awaits further study of the exocrine enhancers and factors which may recognize sequences within them.

## Gene Regulation: Promoters and Enhancers

The actual mechanisms by which regulatory molecules affect transcription seem to be quite different in prokaryotes and eukaryotes. Dissimilar mechanisms reflect the differences in the nature and complexity of the transcription apparatus. In prokaryotes, positive and negative regulatory factors acting at the level of transcription initiation exert their effects at or near the promoter, a fairly small region close to the transcription start site (reviewed in Raibaud & Schwartz, 1984; Ptashne et al., 1980). Yeast genes contain upstream promoter elements, situated as far as hundreds of base pairs upstream from the start site (reviewed in Guarente, 1984). Although the sequences separating the upstream elements from the TATA region may not play a direct role in transcription initiation, they may have a passive function such as allowing the movement of a transcription factor from an upstream binding site to the start site. Perhaps it is not surprising that the regulatory sites in yeast may also be located at relatively large distances from the start site, overlapping with the upstream promoter elements or falling between them and the TATA region (Guarente, 1984; Johnson & Herskowitz, 1985; Wilson & Herskowitz, submitted; Struhl, 1982b). Higher eukaryotes possess an even more complex array of sequence elements which determine the efficiency of transcription initiation. Enhancers may provide a convenient means for adding the levels of control required in complex multicellular systems. The relative

independence from constraints on precise location with respect to transcription initiation sites allows a great deal of flexibility in regulatory strategies, even more than for the yeast upstream elements, which apparently cannot function from positions downstream of the start site (Guarente & Hoar, 1984; Struhl, 1984). The discovery of cell type specific enhancers in two pancreatic exocrine genes provides additional evidence that enhancers play an important role as regulatory sequences in higher eukaryotes.

The features mentioned above and the unique properties of UAS regions and enhancers predict that the mechanisms proposed for the action of positive and negative regulators in prokaryotes are not directly applicable to eukaryotes. Since relatively little is known about factors required for transcription in eukaryotes, the formulation of models for enhancer function is rather difficult. One model proposes that enhancers act as bidirectional entry sites for RNA polymerase or a transcription factor which then "tracks" the DNA in either direction until it comes in contact with an initiation site or other transcription factors (Moreau et al., 1981). An alternative model postulates an involvement of the enhancer in the formation of an "open" chromatin configuration or an "active" DNA template. These two types of models are not mutually exclusive: enhancer function may involve effects both on DNA structure and on binding of essential components of the transcription machinery, or an effect on chromatin which leads to an increase in factor binding.

The discovery that enhancers from several sources are composed of multiple domains or functional motifs suggests that a number of different factors may act in concert, but to some extent independently, to activate transcription (Laimins et al., 1984; Herbomel et al., 1984; Herr & Gluzman,

1985). The heterogeneity in enhancer sequences and in the arrangement of functional domains may indicate that there are several different mechanisms by which enhancers can produce a similar effect on transcription (Weber et al., 1984). At least some of these apparent differences could be a consequence of differences in the properties of a diverse array of factors which interact with enhancer sequences. The experiments described in this thesis do not directly address the question of enhancer mechanisms. However, the results provide further support for the notion that enhancers are composed of more than a single functional domain. Progressive deletions from the 3' ends of the chymotrypsin and amylase enhancer regions resulted in a gradual loss of activity: the 3' boundaries of these enhancers are not sharply defined. This indicates that optimal activity of these enhancers is dependent upon sequences spread over a considerable distance, some of which can act as independent functional domains. The pancreatic exocrine enhancers also provide additional evidence that factors present only in specific cell types interact with distinct classes of enhancer sequences.

Cell Type Specific Expression & Coordinate Regulation

The amount of data available on the pancreatic exocrine system merits only a very tentative model for the control of cell type specific expression and coordinate regulation of several genes. However, models provide specific predictions which can be tested and a framework into which future results can be incorporated.

Any model proposed must explain several aspects of pancreatic exocrine gene expression. The various genes show different patterns of activation during development, pancreatic tumor cells maintain expression of cell type specific genes to varying extents, and each of the exocrine genes responds

in a different manner to a number of regulatory factors such as hormones. One type of model would propose that each gene is regulated by a separate cell type specific factor. In turn, schemes must be proposed for activation of this set of regulatory factors. If every tissue utilized a similar mechanism, an inordinately large number of cell type specific regulatory molecules would be required. On the other hand, a simple model postulating a single pancreas specific regulatory molecule cannot explain all features of exocrine gene expression.

I propose that a pancreatic exocrine cell specific factor is required for high level expression of all genes coding for pancreatic digestive enzymes. The factor would be a DNA binding protein which specifically recognizes the related short sequences in the 5' flanking regions of the pancreatic exocrine genes. In the absence of this factor, the pancreatic exocrine genes would be expressed only at a very low level. The homologous sequences recognized by the exocrine factor would be associated with enhancers which are inactive unless the factor is present. The strengths of different enhancers may be dependent upon the affinity of the cell type specific factor for a particular sequence, the sequences flanking the recognition site, and the proximity of the enhancer to the transcription initiation site (including the effect of sequences lying between the enhancer and the start site). To explain the relative mRNA levels found in pancreatic tumor cells and the asynchronous regulation of exocrine genes during development, I propose that additional factors are required to achieve appropriate expression of some or all of the genes. These factors may or may not be present only in pancreatic exocrine cells. They may be DNA binding proteins or factors which interact directly with components of the transcription machinery, and could represent transcription factors

which are limiting for a particular promoter or factors which augment the activity of a particular enhancer. In this category of accessory factors, I include regulatory molecules mediating the response to extracellular signals such as hormones which may act at functionally distinct enhancer elements.

One prediction of this model is that the consensus-related sequences in each gene are associated with cell type specific enhancers. This can readily be tested using the CAT system in experiments analogous to those described in Chapter 2.

Current & Future Directions

Experiments currently in progress in Dr. Rutter's laboratory address several aspects of gene regulation in the exocrine pancreas. Chris Erwin is continuing an analysis of the DNA sequences within the amylase enhancer region required for activity. Uri Nir has developed a method to test for the presence of factors which repress transcription of cell type specific genes in fibroblasts. Jang Han and Christian Stratowa are using a bacterial expression system to isolate genes coding for factors which bind to cell type specific enhancer sequences. Finally, Eric Fodor has determined that *in vitro* DNase I protection patterns for the amylase enhancer region after incubation of the DNA fragments with an AR4-2J extract are significantly different from those produced with HIT (pancreatic endocrine) cell extracts.

Future experiments will combine transfection techniques and *in vitro* analysis. For example, DNase I protection patterns predict that certain nucleotides in the amylase enhancer region are involved in the binding of a factor or factors. Specific nucleotides can be mutated *in vitro*, and mutant

enhancers can be tested for their activity on the tk promoter using the CAT system. An in vitro transcription system in which the pancreatic enhancers are active would prove very useful for further investigations. Until such a syatem is available, attempts to sort out the different factor requirements of the various exocrine genes will rely on fine mapping of enhancer elements, with a focus on both identification of critical sequences and characterization of functional domains, and on in vivo competition experiments.

# REFERENCES

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a B-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33, 729-740.

Baty, D., Barrera-Saldana, H.A., Everett, R.D., Vigneron, M., and Chambon, P. (1984). Mutational dissection of the 21bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoters. Nucl. Acids Res. 12, 915-932.

Beckwith, J.R. (1978). in "The Operon" (J.H.Miller and W.S.Reznikoff, eds.) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp11-30.

Beier, D. and Young, E.T. (1982). Characterization of a regulatory region upstream of the ADR2 locus of S. cerevisiae. Nature 300, 724-727.

Bell, G.I., Quinto, C., Quiroga, M., Valenzuela, P., Craik, C.S., and Rutter, W.J. (1984). Isolation and sequence of a rat chymotrypsin B gene. J. Biol. Chem. 259, 14265-14270.

Bergman, Y., Rice, D., Grosschedl, R., and Baltimore, D. (1984). Two regulatory elements for immunoglobulin K light chain gene expression. Proc. Nat. Acad. Sci. USA 81, 7041-7045.

Berk, A.J., Lee, F., Harrison, T., Williams, J. and Sharp, P.A. (1979). Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNAs. Cell 17, 935-944.

Borrelli, E., Hen, R. and Chambon, P. (1984). Adenovirus 2 E1A products

repress enhancer-induced stimulation of transcription. Nature 312, 608-612.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

Carter, A.D., Felber, B.K., Walling, M., Jubier, M-F., Schmidt, C.J. and Hamer, D.H. (1984). Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Nat. Acad. Sci. USA 81, 7392-7396.

Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983). DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. Cell 33, 489-499.

Chao, M.V., Mellon, P., Charnay, P., Maniatis, T., and Axel, R. (1983). The regulated expression of B-globin genes introduced into MEL cells. Cell 32, 483-493.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R., and Maniatis, T. (1984). Differences in human alpha and beta-globin gene expression in MEL cells: The role of intragenic sequences. Cell 38, 251-263.

Church, G.M., Ephrussi, A., Gilbert, W., and Tonegawa, S. (1985). Cell-type-specific contacts to immunoglobulin enhancers in nuclei. Nature 313, 798-801.

Cowing, D., Bardwell, J., Craig, E. and Gross, C., unpublished.

Craik, C.S., Choo, Q-L., Swift, G.H., Quinto, C., MacDonald, R.J., and Rutter, W.J. (1984). Structure of two related rat pancreatic trypsin genes. J. Biol. Chem. 259, 14255-14264.

Darnell, J.E. (1982). Variety in the level of gene control in eukaryotic

cells. Nature 297, 365-371.

Davidson, E.H., Jacobs, H.T., and Britten, R.J. (1983). Very short repeats and coordinate induction of genes. Nature 301, 468-470.

deCrombrugghe, B., Busby, S., and Buc, H. (1984). Activation of transcription by the cyclic AMP receptor protein. Science 224, 831-838.

DeFranco, D., Wrange, O., Merriweather, J., and Yamamoto, K.R. in "Gene Regulation" UCLA Symposium on Molecular and Cellular Biology, New Series. (I.Herskowitz and M.Simon, eds.) Academic Press, New York, 1984.

deVilliers, J. and Schaffner, W. (1981). A small segment of polyoma-virus DNA enhances the expression of a cloned rabbit B-globin gene over a distance of at least 1400 base pairs. Nucl. Acids Res. 9, 6251-6264.

deVilliers, J., Olson, L., Tyndall, C., and Schaffner, W. (1982). Transcriptional "enhancers" from SV40 and polyoma virus show a cell type preference. Nucl. Acids Res. 10, 7965-7976.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit B-globin gene in mouse 3T6 cells. Cell 32, 695-706.

Donahue, T.F., Daves, R.S., Lucchini, G. and Fink, G.R. (1983). A short nucleotide sequence required for the regulation of HIS4 by the general control system of yeast. Cell 32, 89-98.

Dunn, T.M., Hahn, S., Ogden, S. and Schleif, R.F. (1984). An operator at -280 base pairs that is required for repression of $ara_{BAD}$ operon promoter: Addition of DNA helical turns between the operator and

promoter cyclically hinders repression. Proc. Nat. Acad. Sci. USA 81, 5017-5020.

Dynan, W.S. and Tjian, R. (1983a). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell 32, 669-680.

Dynan, W.S. and Tjian, R. (1983b). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35, 79-87.

Edlund, T., Walker, M.D., Barr, P.J. and Rutter, W.J. Cell specific expression of the rat insulin gene: Evidence for role of two distinct 5' flanking elements, submitted.

Emerson, B.M. and Felsenfeld, G. (1984). Specific factor conferring nuclease hypersensitivity at the 5' end of the chicken adult B-globin gene. Proc. Nat. Acad. Sci. USA 81, 95-99.

Ephrussi, A., Church, G.M., Tonegawa, S., and Gilbert, W. (1985). B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science 227, 134-140.

Episkopou, V., Murphy, A.J.M., and Efstratiadis, A. (1984). Cell-specified expression of a selectable hybrid gene. Proc. Nat. Acad. Sci. USA 81, 4657-4661.

Everett, R.D., Baty, D., and Chambon, P. (1983). The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucl. Acids Res. 11, 2447-2464.

Gaynor, R.B. and Berk, A.J. (1983). Cis-acting induction of adenovirus transcription. Cell 33, 683-693.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the

major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Gillies, S.D., Folsom, V., and Tonegawa, S. (1984). Cell type specific enhancer element associated with a mouse MHC gene, $E_B$. Nature 310, 594-597.

Giniger, E., Varnum, S.M. and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell 40, 767-774.

Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982a). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2, 1044-1051.

Gorman, C., Merlino, G.T., Willingham, M.C., Pastan, I., and Howard, B. (1982b). The Rous sarcoma virus LTR is a strong promoter when introduced into a variety of eukaryotic cells by DNA mediated transfection. Proc. Nat. Acad. Sci. USA 79, 6777-6781.

Gorman, C., Padmanabhan, R., and Howard, B.H. (1983). High efficiency DNA-mediated transformation of primate cells. Science 221, 551-553.

Graham, F.L. and van der Eb, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52, 456-467.

Graves, B.J., Eisenman, R.N. and McKnight, S.L. Three distinct cis-acting transcriptional control signals occur in the Moloney murine sarcoma virus LTR, submitted.

Grossman, A.D., Erickson, J.W. and Gross, C.A. (1984). The htpR gene product of E.coli is a sigma factor for heat-shock promoters. Cell 38, 383-390.

Guarente, L. (1984). Yeast promoters: Positive and negative elements. Cell 36, 799-800.

Guarente, L. and Hoar, E. (1984). Upstream activation sites of the CYC1

gene of S.cerevisiae are active when inverted but not when placed downstream of the "TATA" box. Proc. Nat. Acad. Sci. USA 81, 7860-7864.

Guarente, L. and Mason, T. (1983). Heme regulates transcription of the CYC1 gene of S.cerevisiae via an upstream activation site. Cell 32, 1279-1286.

Guarente, L., Yocum, R.R. and Gifford, P. (1982). A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc. Nat. Acad. Sci. USA 79, 7410-7414.

Guertin, M., Baril, P., Bartowiak, J., Anderson, A. and Belanger, L. (1983). Rapid suppression of alpha$_1$-fetoprotein gene transcription by dexamethasone in developing rat liver. Biochemistry 22, 4296-4302.

Haltiner, M., Kempe, T., and Tjian, R. (1985). A novel strategy for constructing clustered point mutations. Nucl. Acids Res. 13, 1015-1025.

Harding, J.D., Przybyla, A.E., MacDonald, R.J., Pictet, R.L., and Rutter, W.J. (1978). Effects of dexamethasone and 5-bromo-deoxyuridine on the synthesis of amylase mRNA during pancreatic development in vitro. J. Biol. Chem. 253, 7531-7537.

Hawley, D.A. and McClure, W.R. (1982). Mechanism of activation of transcription initiation from the lambda $P_{RM}$ promoter. J. Mol. Biol. 157, 493-525.

Hawley, D.A. and McClure, W.R. (1983). Compilation and analysis of E.coli promoter DNA sequences. Nucl. Acids Res. 11, 2237-2255.

Hay, R.J., Rookstool, P., Arner, B., and Lavappa, K.S. (1977). Clonal isolation of an epithelial cell line from adult rat pancreas. In

Vitro 13, 167.

Hearing, P. and Shenk, T. (1983). The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell 33, 695-703.

Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M.P., and Chambon, P. (1982). Sequences upstream from the TATA box are required in vivo and in vitro for efficient transcription from the adenovirus serotype 2 major late promoter. Proc. Nat. Acad. Sci. USA 79, 7132-7136.

Herbomel, P., Bourachot, B., and Yaniv, M. (1984). Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. Cell 39, 653-662.

Herr, W. and Gluzman, Y. (1985). Duplications of a mutated SV40 enhancer restore its activity. Nature 313, 711-714.

Hochschild, A., Irwin, N. and Ptashne, M. (1983). Repressor structure and the mechanism of positive control. Cell 32, 319-325.

Irani, M.H., Orosz, L. and Adhya, S. (1983). A control element within a structural gene: The gal operon of E.coli. Cell 32, 783-788.

Jessop, N.W. and Hay, R.J. (1980). Characteristics of two rat pancreatic exocrine cell lines derived from transplantable tumors. In Vitro 16, 212.

Johnson, A.D. and Herskowitz, I. (1985). A repressor (MAT alpha2 product) and its operator control expression of a set of cell-type specific genes in yeast. Cell, in press.

Johnston, M. and Davis, R. (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in S.cerevisiae. Mol. Cell. Biol. 4, 1440-1448.

Johnston, S.A. and Hopper, J.E. (1982). Isolation of the yeast

regulatory gene GAL4 and analysis of its dosage effects on the galactose/melibiose regulon. Proc. Nat. Acad. Sci. USA 79, 6971-6975.

Jones, N. and Shenk, T. (1979). An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc. Nat. Acad. Sci. USA 76, 3665-3669.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984). Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-IIa gene. Nature 308, 513-519.

Keller, J.M. and Alwine, J.C. (1984). Activation of the SV40 late promoter: Direct effects of T antigen in the absence of viral DNA replication. Cell 36, 381-389.

Killary, A.M. and Fournier, R.E.K. (1984). A genetic analysis of extinction: Transdominant loci regulate expression of liver-specific traits in hepatoma hybrid cells. Cell 38, 523-534.

Kondoh, H., Yasuda, K., and Okada, T.S. (1983). Tissue-specific expression of a cloned chick delta-crystallin gene in mouse cells. Nature 301, 440-442.

Korc, M., Owerbach, D., Quinto, C., and Rutter, W.J. (1981). A pancreatic islet-acinar cell interaction: Amylase mRNA levels are determined by insulin. Science 213, 351-353.

Laimins, L.A., Khoury, G., Gorman, C., Howard, B., and Gruss, P. (1982). Host-specific activation of transcription by tandem repeats from SV40 and Moloney murine sarcoma virus. Proc. Nat. Acad. Sci. USA 79, 6453-6457.

Laimins, L.A., Tsichlis, P., and Khoury, G. (1984). Multiple enhancer

domains in the 3' terminus of the Prague strain of Rous sarcoma virus. Nucl. Acids Res. 12, 6427-6442.

Laughon, A. and Gesteland, R.F. (1982). Isolation and preliminary characterization of the GAL4 gene, a positive regulator of transcription in yeast. Proc. Nat. Acad. Sci. USA 79, 6827-6831.

Lee, N.L., Gielow, W.O. and Wallace, R.G. (1981). Mechanism of araC autoregulation and the domains of two overlapping promoters, $P_C$ and $P_{BAD}$, in the L-arabinose region of E.coli. Proc. Nat. Acad. Sci. USA 78, 752-756.

Levinson, B., Khoury, G., Van de Woude, G., and Gruss, P. (1982). Activation of SV40 genome by 72 base pair tandem repeats of Moloney sarcoma virus. Nature 295, 568-572.

Logsdon, C.D., Moessner, J., Williams, J.A., and Goldfine, I.D. Glucocorticoids regulate the differentiation of pancreatic acinar AR42J cells: Effects on morphology, secretion and amylase mRNA levels, in press.

Loyter, A., Scangos, G.A., and Ruddle, F.H. (1982). Mechanism of DNA uptake by mammalian cells: Fate of exogenously added DNA monitored by the use of fluorescent dyes. Proc. Nat. Acad. Sci. USA 79, 422-426.

Luciw, P.A., Bishop, J.M., Varmus, H.E., and Capecchi, M.R. (1983). Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell 33, 705-716.

Lusky, M., Berg, L., Weiher, H., and Botchan, M. (1983). Bovine papilloma virus contains an activator of gene expression at the distal end of the early transcription unit. Mol. Cell. Biol. 3,

1108-1122.

Luthman, H. and Magnusson, G. (1983). High efficiency polyoma DNA transfection of chloroquine treated cells. Nucl. Acids Res. 11, 1295-1308.

MacKay, V. and Manney, T.R. (1974). Mutations affecting sexual conjugation and related processes in S.cerevisiae. II. Genetic analysis of nonmating mutants. Genetics 76, 272-288.

Majors, J. (1975). Initiation of in vitro mRNA synthesis from the wild type lac promoter. Proc. Nat. Acad. Sci. USA 72, 4394-4398.

Majors, J. and Varmus, H.E. (1983). A small region of the mouse mammary tumor virus LTR confers glucocorticoid hormone regulation on a linked heterologous gene. Proc. Nat. Acad. Sci. USA 80, 5866-5870.

Maniatis, T., Fritsch, E.F., and Sambrook, J. Molecular Cloning - A Laboratory Manual (Cold Spring Harbor Laboratory, New York, 1982).

McKnight, S.L. (1982). Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell 31, 355-365.

McKnight, S.L. and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein coding gene. Science 217, 316-325.

McKnight, S.L., Gavis, E.R., Kingsbury, R., and Axel, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: Identification of an upstream control region. Cell 25, 385-398.

McKnight, S.L., Kingsbury, R.C., Spence, A., and Smith, M. (1984). The distal transcription signals of the herpes virus tk gene share a common hexanucleotide control sequence. Cell 37, 253-262.

Mercola, M., Goverman, J., Mirell, C., and Calame, K. (1985).

Immunoglobulin heavy chain enhancer requires one or more tissue-specific factors. Science 227, 266-270.

Miller, D.M., Turner, P., Nienhuis, A.W., Axelrod, D.E. and Gopalakrishnan, T.V. (1978). Active conformation of the globin genes in uninduced and induced MEL cells. Cell 14, 511-521.

Moessner, J., Logsdon, C.D., Williams, J.A., and Goldfine, I.D. in press.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl. Acids Res. 9, 6047-6068.

Musso, R.E., diLauro, R., Adhya, S. and deCrombrugghe, B. (1977). Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. Cell 12, 847-854.

Myers, R.M., Rio, D.C., Robbins, A.K. and Tjian, R. (1981). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell 25, 373-384.

Neidhardt, F.C. and VanBogelen, R.A. (1981). Positive regulatory gene for temperature-controlled proteins in E.coli. Biochem. Biophys. Res. Commun. 100, 894-900.

Neuberger, M.S. (1983). Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells. EMBO J. 2, 1373-1378.

Nevins, J.R. (1981). Mechanism of activation of early viral transcription by the adenovirus E1A gene product. Cell 26, 213-220.

Nordheim, A. and Rich, A. (1983). Negative supercoiled SV40 DNA contains

Z-DNA segments within transcriptional enhancer sequences. Nature 303, 674-679.

Oi, V.T., Morrison, S.L., Herzenberg, L.A., and Berg, P. (1983). Immunoglobulin gene expression in transformed lymphoid cells. Proc. Nat. Acad. Sci. USA 80, 825-829.

Ornitz, D.M., Palmiter, R.D., Hammer, R.E., Brinster, R.L., Swift, G.H., and MacDonald, R.J. (1985). Specific expression of an elastase-human growth hormone fusion gene in pancreatic acinar cells of transgenic mice. Nature 313, 600-602.

Palmiter, R., Mulvihill, E., McKnight, S. and Senear, A. (1978). Cold Spring Harb. Symp. quant. Biol. 42, 639-647.

Parker, C.S. and Topol, J. (1984). A Drosophila RNA polymerase II transcription factor specific for the heat-shock gene binds to the regulatory site of an hsp70 gene. Cell 37, 273-283.

Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J-A., and Yamamoto, K.R. (1983). Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell 35, 381-392.

Pelham, H.R.B. (1982). A regulatory upstream promoter element in the Drosophila hsp70 heat shock gene. Cell 30, 517-528.

Picard, D. and Schaffner, W. (1984). A lymphocyte-specific enhancer in the mouse immunoglobulin kappa gene. Nature 307, 80-82.

Platt, T. (1981). Termination of transcription and its regulation in the tryptophan operon of E.coli. Cell 24, 10-23.

Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M. and Sauer, R.T. (1980). How the lambda repressor and cro work. Cell 19, 1-11.

Queen, C. and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell 33, 741-748.

Queen, C. and Stafford, J. (1984). Fine mapping of an immunoglobulin gene activator. Mol. Cell. Biol. 4, 1042-1049.

Raibaud, O. and Schwartz, M. (1984). Positive control of transcription initiation in bacteria. Ann. Rev. Genet. 18, 173-206.

Rall, L., Pictet, R., Githens, S., and Rutter, W.J. (1977). Glucocorticoids modulate the in vitro development of the embryonic rat pancreas. J. Cell. Biol. 75, 398-409.

Rao, K.N., Takahashi, S., and Shinozuka, H. (1980). Acinar cell carcinoma of the rat pancreas grown in cell culture and in nude mice. Cancer Res. 40, 592-597.

Renkawitz, R., Beug, H., Graf, T., Matthias, P., Grez, M., and Schutz, G. (1982). Expression of a chicken lysozyme recombinant gene is regulated by progesterone and dexamethasone after microinjection into oviduct cells. Cell 31, 167-176.

Renkawitz, R., Schutz, G., von der Ahe, D. and Beato, M. (1984). Sequences in the promoter region of the chicken lysozyme gene required for steroid regulation and receptor binding. Cell 37, 503-510.

Rice, D. and Baltimore, D. (1982). Regulated expression of an immunoglobulin K gene introduced into a mouse lymphoid cell line. Proc. Nat. Acad. Sci. USA 79, 7862-7865.

Ringold, G.M., Yamamoto, K.R., Bishop, J.M. and Varmus, H.E. (1977). Glucocorticoid-stimulated accumulation of mouse mammary tumor virus RNA: Increased rate of synthesis of viral RNA. Proc. Nat. Acad. Sci. USA 74, 2879-2883.

Rio, D.C. and Tjian, R. (1983). SV40 T antigen binding site mutants that affect autoregulation. Cell 32, 1227-1240.

Rutter, W.J., Przybyla, A.E., MacDonald, R.J., Harding, J.D., Chirgwin, J.M. and Pictet, R.L. (1978). in "Cell Differentiation and Neoplasia" (G.F.Saunders, ed.) Raven Press, New York.

Sakonju, S., Bogenhagen, D.F., and Brown, D.D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of transcription: The 5' border region. Cell 19, 13-25.

Sassone-Corsi, P., Wildeman, A., and Chambon, P. (1985). A trans-acting factor is responsible for the SV40 enhancer activity in vitro. Nature 313, 458-463.

Scholer, H.R. and Gruss, P. (1984). Specific interaction between enhancer-containing molecules and cellular components. Cell 36, 403-411.

Sprague, G.F., Jensen, R. and Herskowitz, I. (1983). Control of yeast cell type by the mating type locus: Positive regulation of the alpha-specific STE3 gene by the MAT alpha1 product. Cell 32, 409-415.

Strathern, J., Hicks, J. and Herskowitz, I. (1981). Control of cell type in yeast by the mating type locus: The alpha1-alpha2 hypothesis. J. Mol. Biol. 147, 357-372.

Struhl, K. (1982a). The yeast HIS3 promoter contains at least two distinct elements. Proc. Nat. Acad. Sci. USA 79, 7385-7389.

Struhl, K. (1982b). Regulatory sites for HIS3 gene expression in yeast. Nature 300, 284-287.

Struhl, K. (1984). Genetic properties and chromatin structure of the yeast GAL regulatory element: An enhancer-like sequence. Proc. Nat.

Acad. Sci. USA 81, 7865-7869.

Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L., and Palmiter, R.D. (1984). A 12 base pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Nat. Acad. Sci. USA 81, 7318-7322.

Swift, G.H., Craik, C.S., Stary, J., Quinto, C., Lahaie, R.G., Rutter, W.J., and MacDonald, R.J. (1984). Structure of the two related elastase genes expressed in the rat pancreas. J. Biol. Chem. 259, 14271-14278.

Tjian, R. (1981). T antigen and the control of SV40 gene expression. Cell 26, 1-2.

Ucker, D.S., Firestone, G.L. and Yamamoto, K.R. (1983). Glucocorticoids and chromosomal position both modulate MTV transcription by affecting the efficiency of promoter utilization. Mol. Cell. Biol. 3, 551-561.

Velcich, A. and Ziff, E. (1985). Adenovirus E1A proteins repress transcription from the SV40 early promoter. Cell 40, 705-716.

Walker, M.D., Edlund, T., Boulet, A.M., and Rutter, W.J. (1983). Cell specific expression controlled by the 5' flanking region of insulin and chymotrypsin genes. Nature 306, 557-561.

Wasylyk, B., Wasylyk, C., Augereau, P., and Chambon, P. (1983). The SV40 72bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell 32, 503-514.

Weber, F., deVilliers, J., and Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell 36, 983-992.

Weiher, H., Konig, M., and Gruss, P. (1983). Multiple point mutations

affecting the SV40 enhancer. Science 219, 626-631.

Weiher, H. and Botchan, M.R. (1984). An enhancer sequence from bovine papilloma virus DNA consists of two essential regions. Nucl. Acids Res. 12, 2901-2916.

Weintraub, H. and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science 193, 848-856.

Weisbrod, S. (1982). Active chromatin. Nature 297, 289-295.

West, R., Yocum, R. and Ptashne, M. (1984). S.cerevisiae GAL1-GAL10 divergent promoter region: Location and function of the upstream activating sequence $UAS_G$. Mol. Cell. Biol. 4, 2467-2478.

Wilson, K.L. and Herskowitz, I. (1984). Negative regulation of STE6 gene expression by the alpha2 product of yeast. Mol. Cell. Biol. 4, 2420-2427.

Wilson, K.L. and Herskowitz, I. A site upstream of STE6 is required for its regulation by the mating type locus in S.cerevisiae, submitted.

Wright, S., deBoer, E., Grosveld, F.G. and Flavell, R.A. (1983). Regulated expression of the human B-globin gene family in murine erythroleukemia cell hybrids. Nature 305, 333-336.

Wright, S., Rosenthal, A., Flavell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of B-globin genes in MEL cells. Cell 38, 265-273.

Wu, C. (1980). The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNaseI. Nature 286, 854-860.

Wu, C. and Gilbert, W. (1981). Tissue-specific exposure of chromatin structure at the 5' terminus of the rat preproinsulin II gene. Proc. Nat. Acad. Sci. USA 78, 1577-1580.

Yamamoto, K.R. (1983). in "Steroid Hormone Receptors: Structure and

Function" (H.Erikkson, J.-A.Gustafsson and B.Hogberg, eds.) Nobel Symposium 57, Elsevier/ North Holland Biomed Press, Amsterdam, p 285.

Yocum, R., Hanley, S., West, R. and Ptashne, M. (1984). Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in S.cerevisiae. Mol. Cell. Biol. 4, 1985-1998.

Zaret, K.S. and Yamamoto, K.R. (1984). Reversible and persistent changes in chromatin structure accompany activation of a glucocorticoid-dependent enhancer element. Cell 38, 29-38.

Zinn, K., DiMaio, D., and Maniatis, T. (1983). Identification of two distinct regulatory regions adjacent to the human B-interferon gene. Cell 34, 865-879.

# APPENDIX I: PLASMID CONSTRUCTIONS

Plasmids Used For Constructions and Reference

pBR.CAT(constructed by T. Edlund): This plasmid was used to test the 5' flanking region fragments of various genes for the ability to direct cell type specific expression. DNA fragments with HindIII ends can be inserted at a unique HindIII site directly upstream from the coding sequences for chloramphenicol acetyltransferase (CAT). Recombinants in which transcription from the promoter of interest reads toward the CAT gene are used for subsequent assays. (Plasmids containing the inserted fragment in the opposite orientation serve as controls.) A 1600bp HindIII to BamHI fragment, including the CAT coding region and SV40 splice sites and polyadenylation signals, was used to replace the HindIII to BamHI region of pBR322 (Figure 13; Walker et al., 1983).

pTE0 (constructed by T. Edlund): A number of modifications were made in the basic pBR.CAT plasmid to produce pTE0. The CAT and SV40 sequences are identical to those in pBR.CAT, but the BamHI to NruI region of pBR322 (map positions 375 to 971) at the 3' end of the CAT gene in pBR.CAT was deleted. The BamHI site was retained while the NruI site was destroyed. The same BamHI to NruI region of pBR322 was inserted upstream from the CAT gene, with both sites retained. In addition, a polylinker was inserted at the 5' end of this relocated pBR322 fragment between the HindIII and BamHI sites, and an SstI/SacI site was created directly upstream from the CAT sequences (figure 13).

pTE1 (constructed by T. Edlund): This plasmid was derived by the insertion of the HSV tk promoter at the SstI/SacI site at the 5' end of the CAT coding region in pTE0. The tk promoter fragment was obtained from a mutant produced by McKnight et al. (1981) which encompasses the

region between -109 and +55bp relative to the cap site. This fragment contains all elements required to direct efficient transcription initiation. The BamHI to BglII tk fragment was blunt-ended and SacI linkers were added for introduction into pTE0 (figure 13; Edlund et al., submitted).

RSV.CAT (Gorman et al., 1982b): In this construction, 524bp containing the promoter and upstream regulatory elements from the Rous sarcoma virus 3' LTR was inserted into the HindIII site of pSV0.CAT.

Fusion of Exocrine Gene 5' Flanking Regions to the CAT Gene

Chymotrypsin, pBR.Cht.CAT: A 709bp EcoRI/HindIII fragment, containing the 5' flanking region sequences of the rat chymotrypsin b gene from -711 to -3 relative to the cap site (Bell et al., 1984), was isolated and blunt ended with T4 polymerase. This repaired fragment was then inserted at the HindIII site of pBR.CAT using HindIII linkers (figure 14).

Amylase, Amy-17.CAT: An Fnu4HI fragment containing approximately 1350bp of the rat pancreatic amylase gene 5' flanking region and 32bp of transcribed sequences, including the template for the first AUG codon (W. Swain, unpublished), was isolated and cloned into the HindIII site of pBR.CAT using HindIII linkers (figure 14). Upon transfection into AR4-2J cells and rat fibroblasts, little or no cell type specificity was observed. The CAT activity directed by this plasmid, termed pBR.Amy.CAT, was very low in both cell lines. Since the template sequence for the first amylase AUG codon is not in frame with the CAT gene AUG in this construction, translation initiation at the amylase AUG would not result in synthesis of the CAT enzyme. The amylase AUG template sequence was

deleted to determine whether cell type specific expression was masked by a preference for the first AUG codon in the mRNA transcript. The amylase 5' flanking region fragment was cloned into the HindIII site of pTE0. This intermediate plasmid was linearized at the XbaI site at the 3' end of the amylase insert (figure 15). Linearized plasmid molecules were treated with exonuclease III and S1 nuclease to remove 30-50bp from each susceptible end. The amylase sequences were then fused to the CAT coding region by digestion with NruI and blunt end ligation. A few of the resulting plasmids were analyzed by dideoxy DNA sequencing, and a deletion which retains amylase sequences upstream from -17bp relative to the cap site was chosen for further analysis and designated Amy-17.CAT.

Trypsin I, pBR.TrpI.CAT (constructed by C. Erwin): A 1162bp HgiAI fragment containing sequences between -1172 and -10 relative to the transcription start site of the rat trypsin I gene (Craik et al., 1984) was inserted into the HindIII site of pBR.CAT using HindIII linkers (figure 14).

pTE1 Derivatives: Enhancer Assay Plasmids

A 182bp SstI fragment from the chymotrypsin 5' flanking region (-275 to -93 relative to the cap site) was cloned, in both orientations, into the unique BglII site of pTE1 using BamHI linkers (figure 16). A 193bp RsaI fragment (-235 to -41) and an approximately 450bp AhaIII fragment (about -530 to -77) from the amylase 5' flanking region were inserted, in both orientations, into the HindIII site of pTE1 using HindIII linkers (figure 16). In these pTE1 constructions, the amylase and chymotrypsin fragments are located approximately 700bp upstream from the transcription start site of the tk promoter. To move the inserts to

a position 110bp from the start site (directly upstream from the tk promoter sequences), each plasmid was digested with SalI and NruI, blunt ended with T4 polymerase, and recircularized with DNA ligase. This procedure deletes the pBR322 sequences between the polylinker and the tk promoter in pTE1 (figure 16). The chymotrypsin SstI fragment was also inserted, in both orientations, into the BamHI site at the 3' end of the CAT gene/SV40 sequences using BamHI linkers (figure 16).

Deletions and Mutations

5' Deletions from pBR.Cht.CAT: Deletions from the 5' end of the chymotrypsin flanking region were generated by partial digestion of pBR.Cht.CAT with SstI, treatment with Bal31, digestion with ClaI, treatment with T4 polymerase to produce blunt ends, and ligation (figure 17). Using this protocol each deletion end point is joined to the same sequences at the ClaI site of pBR322.

Deletions from Chymotrypsin and Amylase Enhancer Fragments: Deletions from the 3' end of the chymotrypsin fragment and from the 5' and 3' ends of the amylase fragment were generated from the pTE1 derivatives in which the SstI insert (normal orientation) and the RsaI insert (normal and inverted orientations) are separated from the tk promoter by 600bp of pBR322 sequences (figure 18). These plasmids were digested with SalI (chymotrypsin pTE1 derivative) or XbaI (amylase pTE1 derivatives) to cut the respective plasmids adjacent to the insert. The linearized plasmids were then treated with exonuclease III for various periods of time. After treatment with S1 nuclease, deleted molecules were digested with NruI and recircularized with ligase to join each deletion end point directly to the tk promoter sequences. The extent of

each deletion was determined by dideoxy DNA sequencing.

Combination of 5' and 3' Deletions of the Chymotrypsin Enhancer Region: NcoI fragments containing the 3' half of the chymotrypsin insert, the tk promoter, and the 5' portion of the CAT coding region were isolated from four pTE1 derivatives with 3' deletions (figure 19). These fragments were used to replace the NcoI fragment, consisting of the chymotrypsin 5' flanking region sequences from -192 to -3 and the 5' portion of the CAT gene, in the pBR.Cht.CAT derivative with a 5' deletion to -225. In the resulting constructions, chymotrypsin fragments with coordinates -225 to -113, -137, -150, or -170 are located directly upstream from the tk promoter fused to the CAT gene.

Deletion of the HSV tk Promoter Region: pTE1 derivatives containing the amylase and chymotrypsin fragments directly adjacent to the tk promoter region were digested with SstI, separated from the small tk promoter fragment and recircularized. In the resulting plasmids, the amylase and chymotrypsin fragments are juxtaposed to the CAT coding sequences (figure 20).

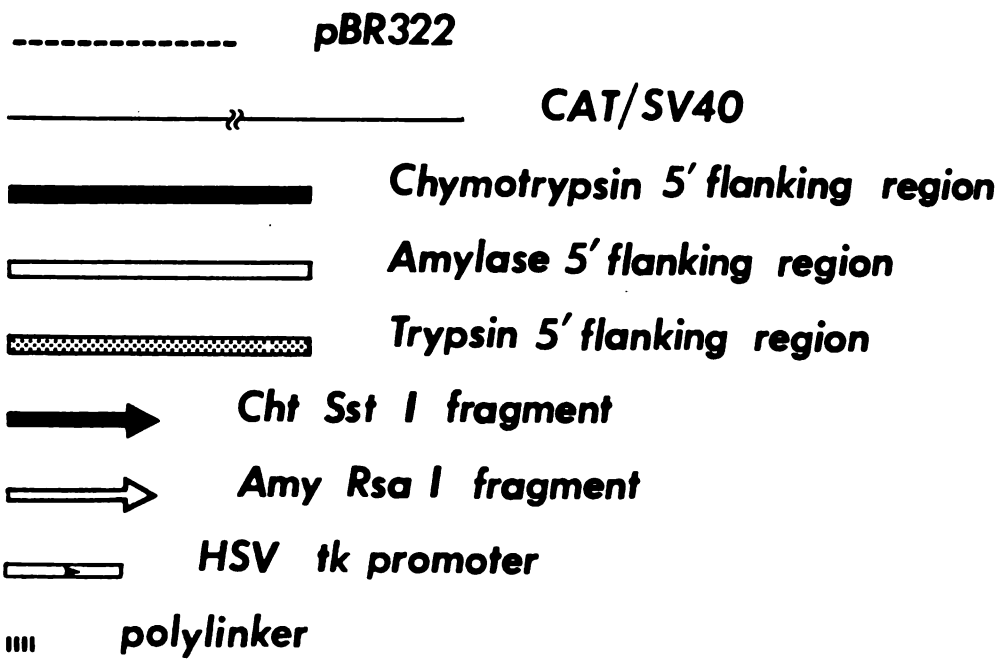
Mutations Within the Chymotrypsin Enhancer Region: pBR.Cht.CAT was partially digested with NcoI and molecules with a single cut were gel purified. One portion of this DNA was treated with T4 polymerase to fill in the single stranded 5' overhanging ends. A second portion was treated with S1 nuclease to digest the single stranded regions. Both samples were then recircularized with T4 ligase, and the plasmids in which only the NcoI site at -192 in the chymotrypsin 5' flanking region had been destroyed were isolated. S1 treatment results in a 4bp deletion at the NcoI site while T4 polymerase repair produces a 4bp duplication (figure 21).

Amylase Enhancer Region 5' and 3' Deletions, Reversed Orientations: Two amylase enhancer region 5' and 3' deletions (the shortest deletion and the most extensive deletion retaining detectable enhancer activity) were digested with HindIII, treated with T4 polymerase to create blunt ends, and digested with BamHI. The HindIII/blunt-BamHI fragments were used to replace the BglII-NruI region in pTE1 (figure 22). In the resulting constructions, the deleted amylase fragments are joined directly to the tk promoter sequences in the orientation opposite that in which they were originally produced and tested (5' deletions are now in the normal orientation and 3' deletions are in the reversed orientation).

Upstream Amylase Enhancer/Downstream Chymotrypsin Enhancer Combinations: The ClaI/HpaI fragments of pTE1 derivatives containing the amylase enhancer located 600bp upstream from the tk promoter fragment were isolated and used to replace the analogous ClaI/HpaI fragments of pTE1 derivatives containing the chymotrypsin enhancer at the 3' end of the CAT/SV40 sequences. In the resulting constructions, the amylase enhancer is located 710bp upstream from the cap site and the chymotrypsin enhancer is 1700bp downstream. Two combinations were tested; those with the amylase enhancer in the normal orientation and the chymotrypsin enhancer in the normal and inverted orientations.

Tandem Upstream Enhancers: The amylase and chymotrypsin enhancer fragments were inserted, using HindIII linkers, into the HindIII site of the pTE1 derivative with the chymotrypsin fragment in the normal orientation at a position 600bp upstream from the tk promoter region.

FIGURE 13: Structure of plasmids used for CAT constructions. Relevant portions of circular plasmids are shown. Symbols are defined in the key below, except that the tk promoter is represented by an arrow in pTE1. The polylinker, containing multiple restriction sites, is indicated by the wavy line drawn above the pTE1 and pTE0 diagrams. Restriction sites are labeled as follows: B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; N, NruI; P, PstI; S, SalI; Ss, SstI/SacI; X, XbaI.



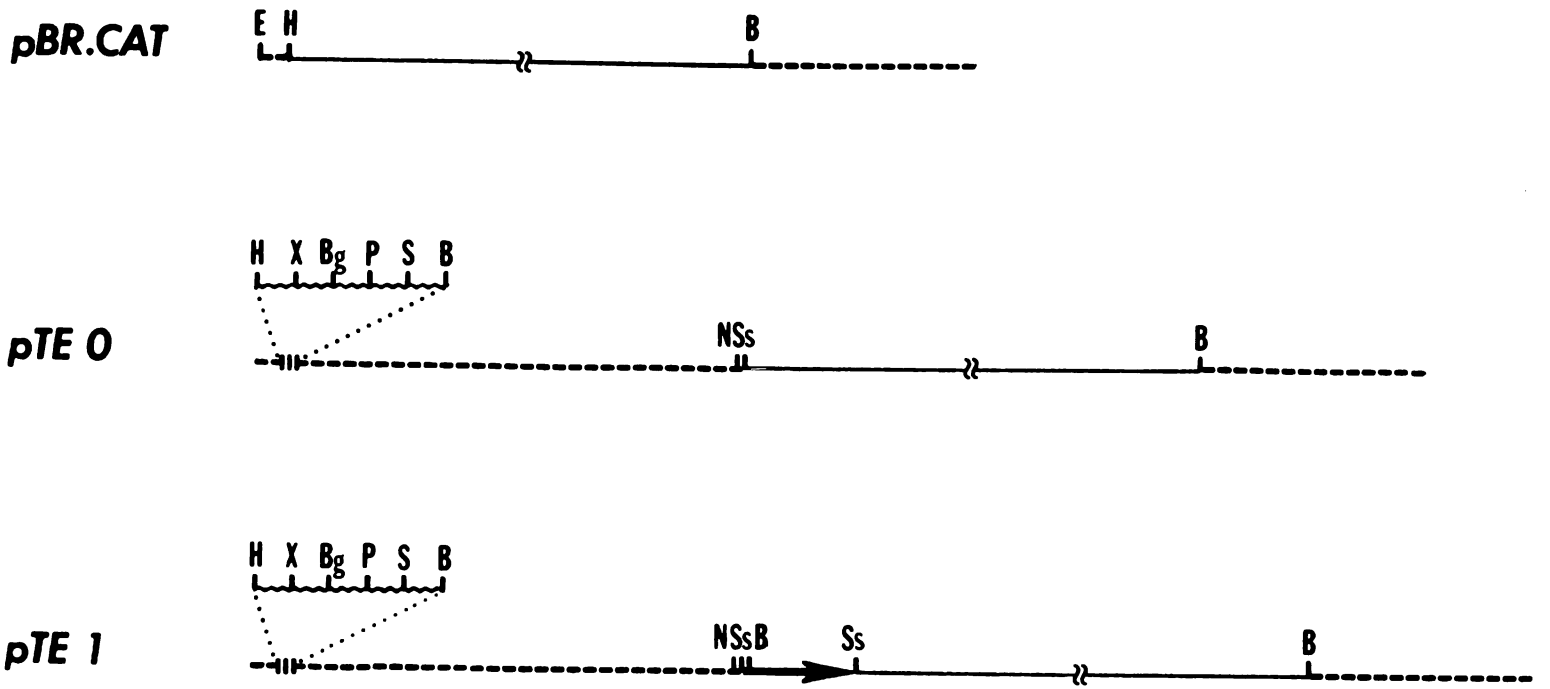



FIGURE 13

FIGURE 14: Structure of plasmids containing the 5' flanking regions of the rat chymotrypsin, amylase, and trypsin I genes fused to CAT coding sequences. Symbols and restriction sites are as described in the legend to figure 13. Numbers indicate nucleotide positions with respect to the cap site of each gene.

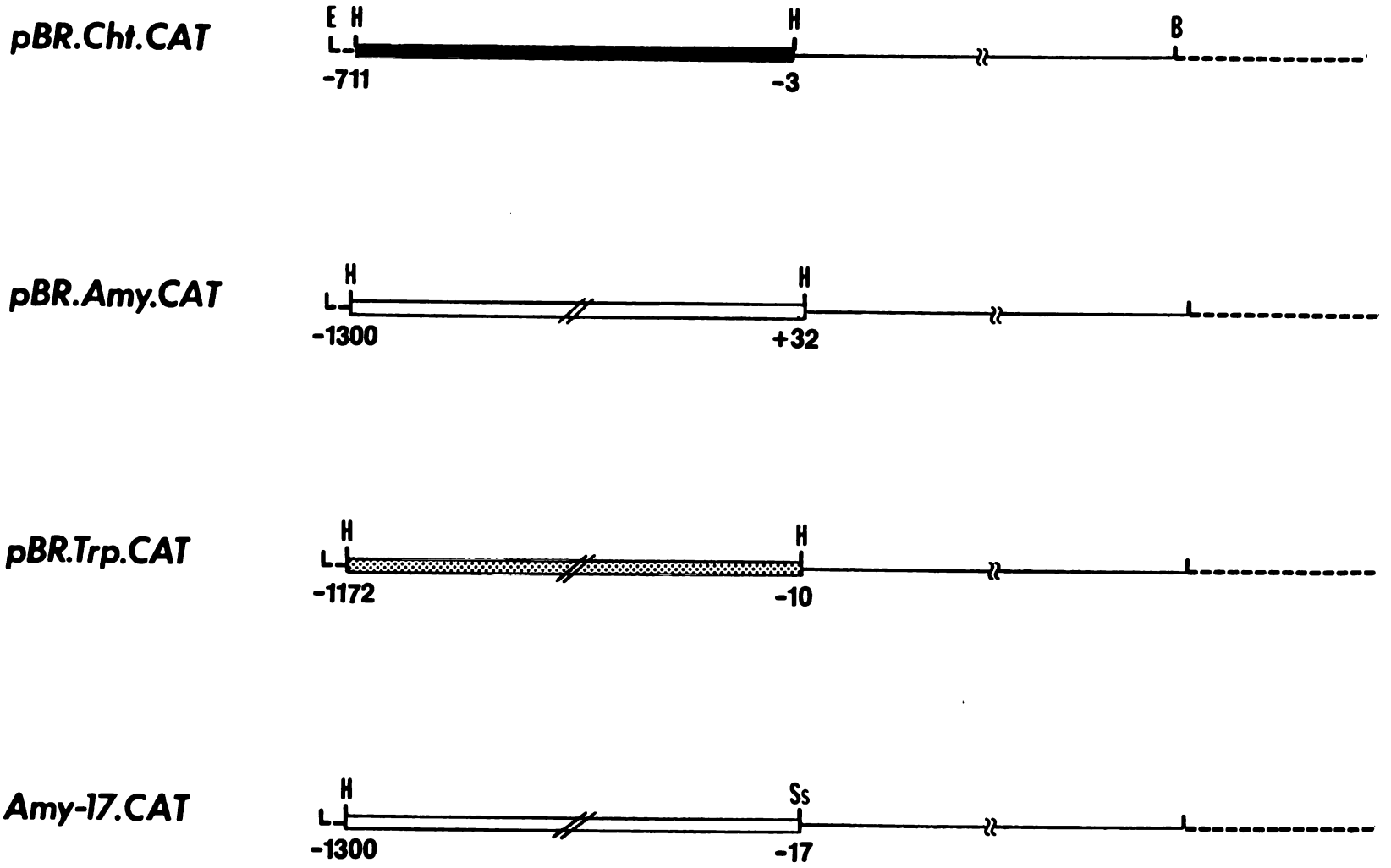

pBR.Cht.CAT
E H
H
B
-711
-3
pBR.Amy.CAT
H
H
-1300
+32
pBR.Trp.CAT
H
H
-1172
-10
Amy-17.CAT
H
Ss
-1300
-17

FIGURE 14

FIGURE 15: Construction of Amy-17.CAT. Details of the procedure are given in the text and Appendix II. Symbols and restriction sites are as described in the legend to figure 13. Numbers beneath the amylase 5' flanking region indicate nucleotide position with respect to the amylase cap site.

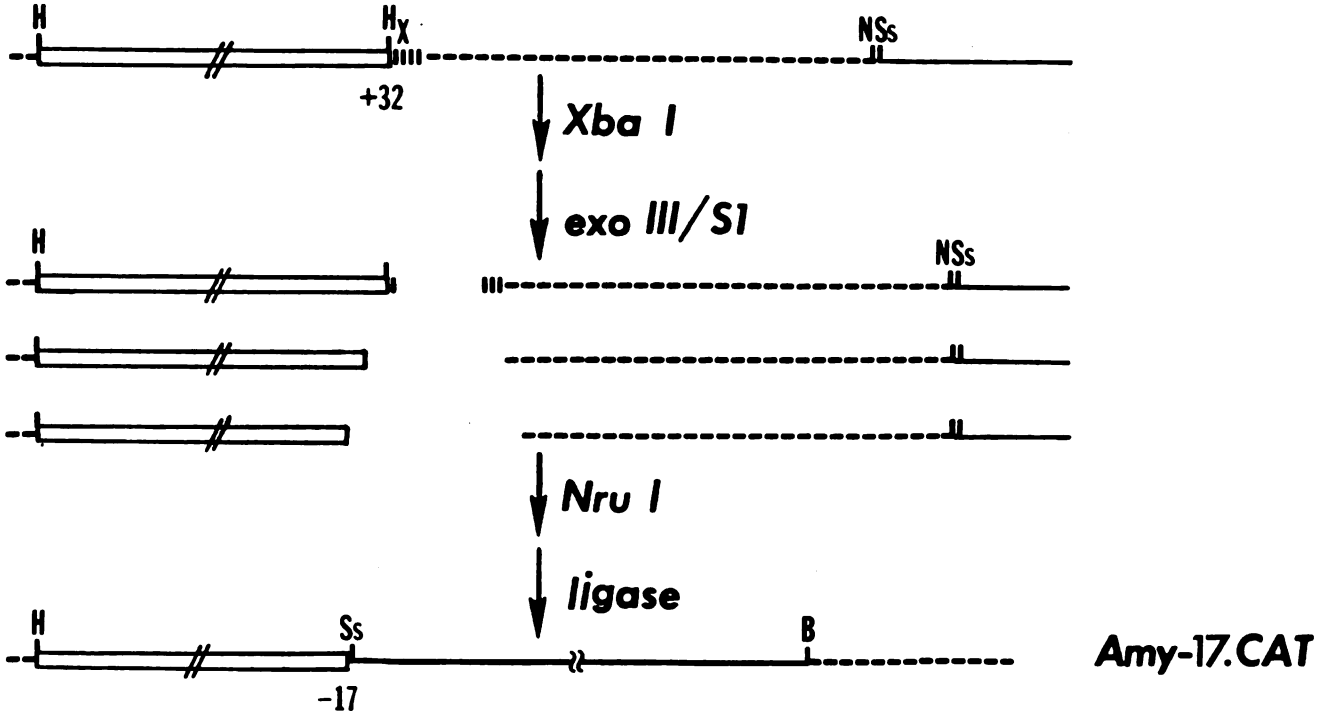



FIGURE 15

FIGURE 16: Construction of pTE1 derivatives containing the amylase RsaI or AhaIII fragment or the chymotrypsin SstI fragment. Details of the methods used are given in the text. Scales at the top and bottom and numbers beneath the isolated fragments indicate nucleotide position relative to the cap site of the respective gene. Symbols and restriction site abbreviations are as described in the legend to figure 13 with the following additions: A, AhaIII; Nc, NcoI; Pv, PvuII; R, RsaI. The amylase AhaIII fragment, the amylase RsaI fragment, and the chymotrypsin SstI fragment, respectively, are labeled 1, 2, and 3. Numbered arrows above pTE1 and below pTE1 with a deletion of the sequences between the SalI and NruI sites indicate restriction sites at which the corresponding fragments were inserted.

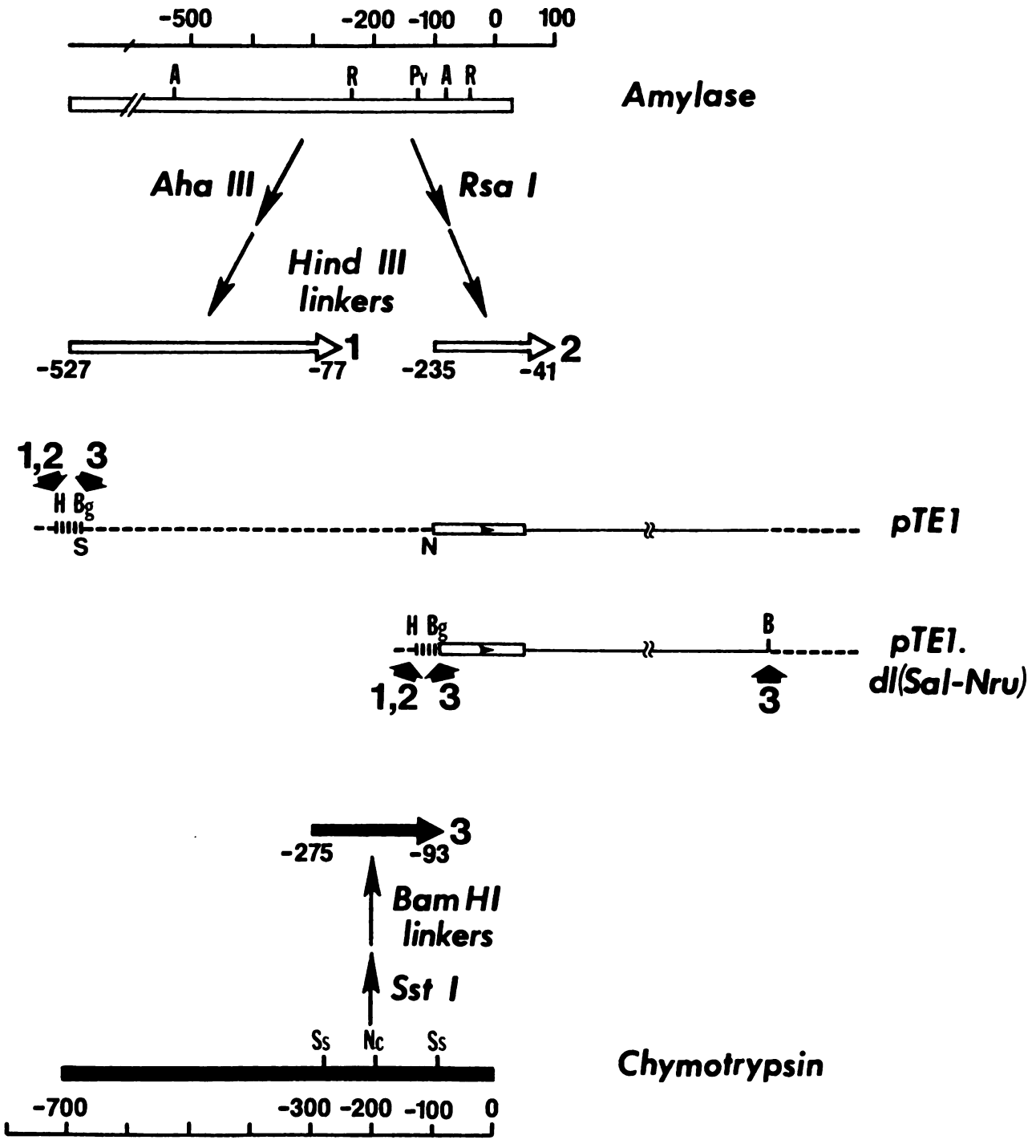

-500
-200
-100
0
100
A
R
Pv
A
R
Amylase
Aha III
Rsa I
Hind III
linkers
1
2
-527
-77
-235
-41
1,2
3
H
Bg
S
N
pTE1
H
Bg
B
pTE1.
dl(Sal-Nru)
1,2
3
3
-275
-93
3
Bam HI
linkers
Sst I
Ss
Nc
Ss
Chymotrypsin
-700
-300
-200
-100
0

FIGURE 16

FIGURE 17: Construction of 5' deletions from the SstI site at position -275 in the chymotrypsin 5' flanking region fragment. Details of the methods are given in the text. Symbols and restriction sites are as described in the legend to figure 13 with the following addition: C, ClaI. Numbers below the diagrams indicate nucleotide position with respect to the chymotrypsin cap site.

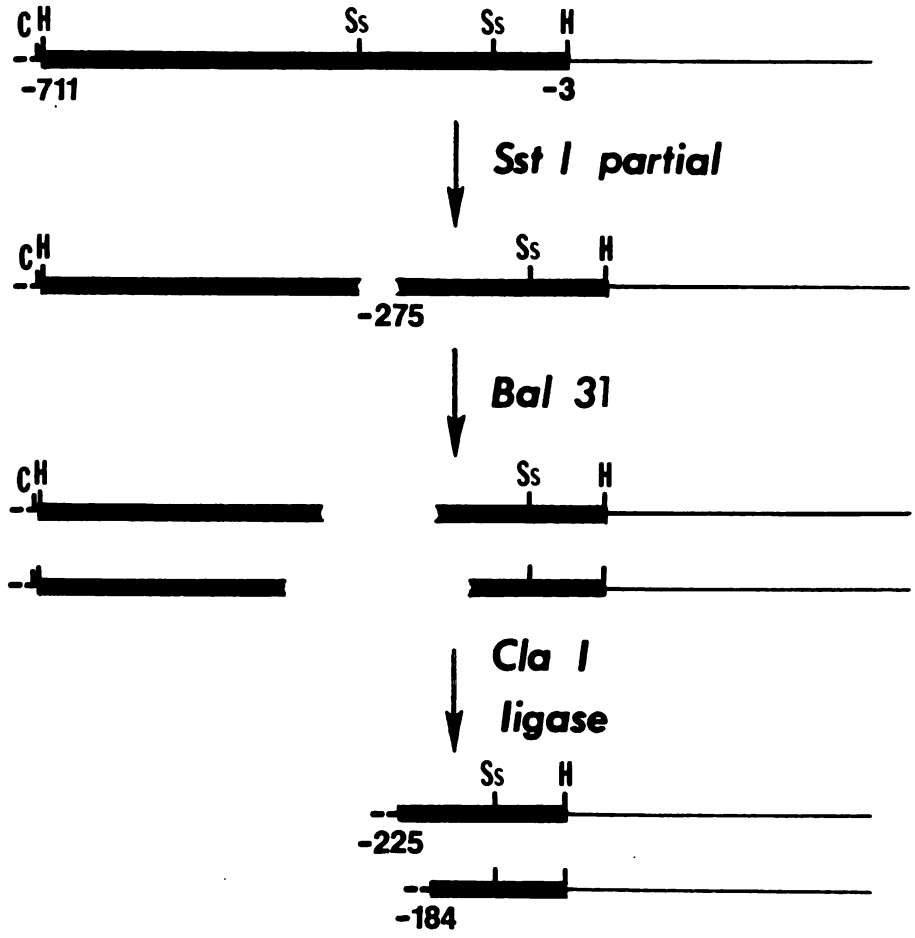

C H
Ss
Ss
H
-711
-3
Sst I partial
C H
Ss
H
-275
Bal 31
C H
Ss
H
Cla I
ligase
Ss
H
-225
-184

FIGURE 17

FIGURE 18: Construction of 3' deletions from the -275 to -93 chymotrypsin SstI fragment and 5' and 3' deletions from the -235 to -41 amylase RsaI fragment. Details of the methods are given in the text. Symbols and restriction sites are as described in the legend to figure 13.

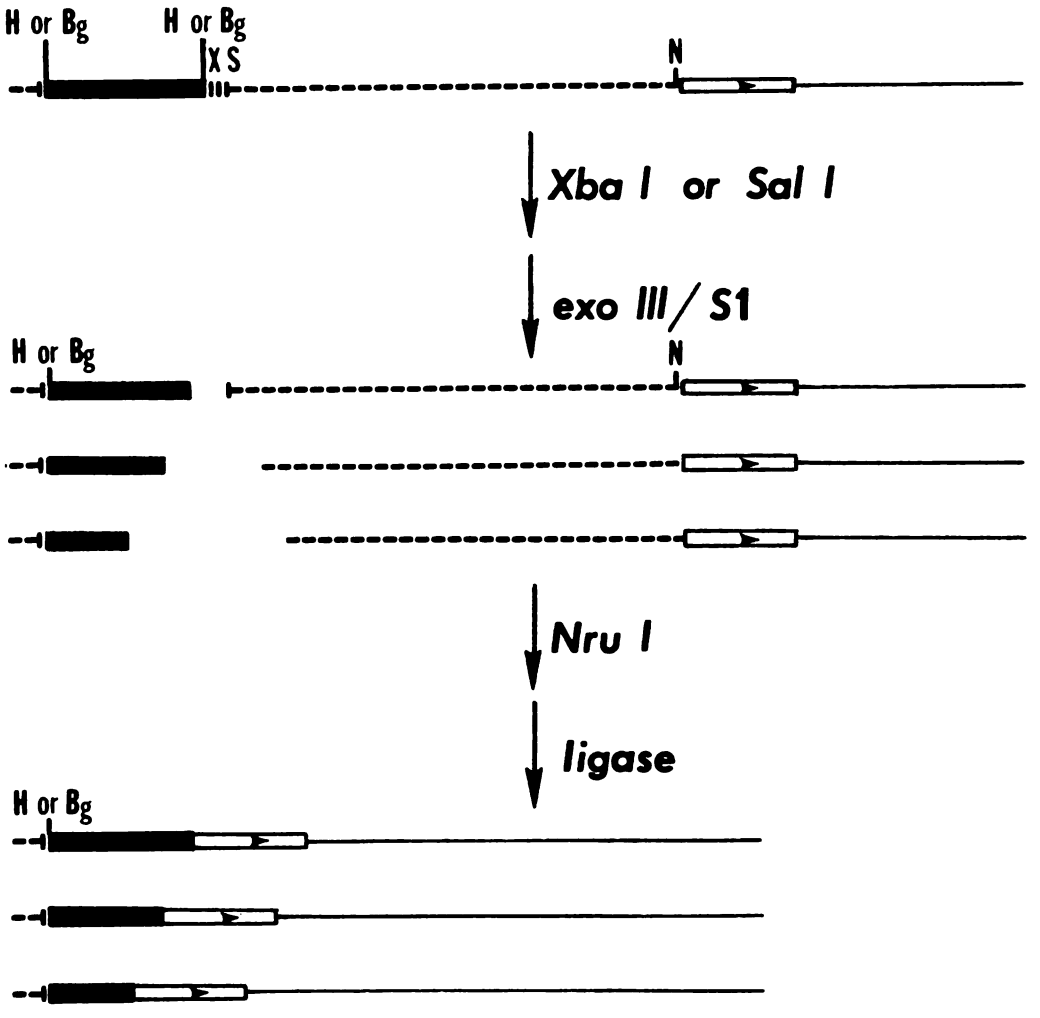

H or Bg
H or Bg
X S
N
Xba I or Sal I
exo III / S1
H or Bg
N
Nru I
ligase
H or Bg

FIGURE 18

FIGURE 19: Combination of 5' and 3' deletions from the chymotrypsin enhancer region. Details of the methods are given in the text. Symbols and restriction site abbreviations are as described in the legend to figure 13. The line at top left represents the pBR.Cht.CAT derivative containing a 5' deletion to -225. The lines at upper right represent pTE1 derivatives in which chymotrypsin fragments with 3' deletions are fused to the tk promoter. Numbers below the lines indicate nucleotide positions with respect to the chymotrypsin cap site.

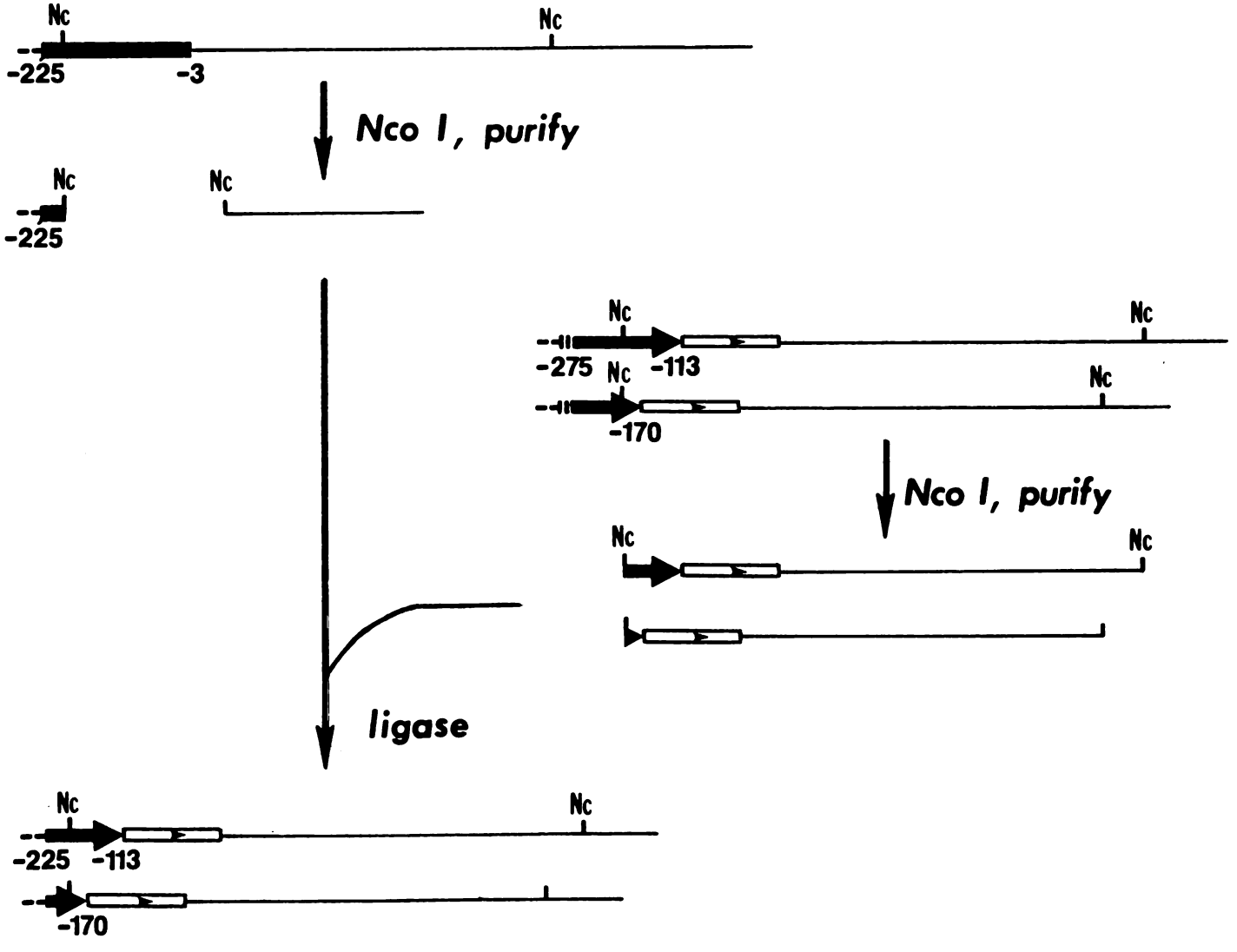

Nc
Nc
-225
-3
Nco I, purify
Nc
Nc
-225
Nc
Nc
-275
Nc
-113
Nc
-170
Nco I, purify
Nc
Nc
ligase
Nc
Nc
-225
-113
-170

FIGURE 19

FIGURE 20: Deletion of the tk promoter fragment from pTE1 derivatives. Symbols and restriction site abbreviations are as described in the legend to figure 13.

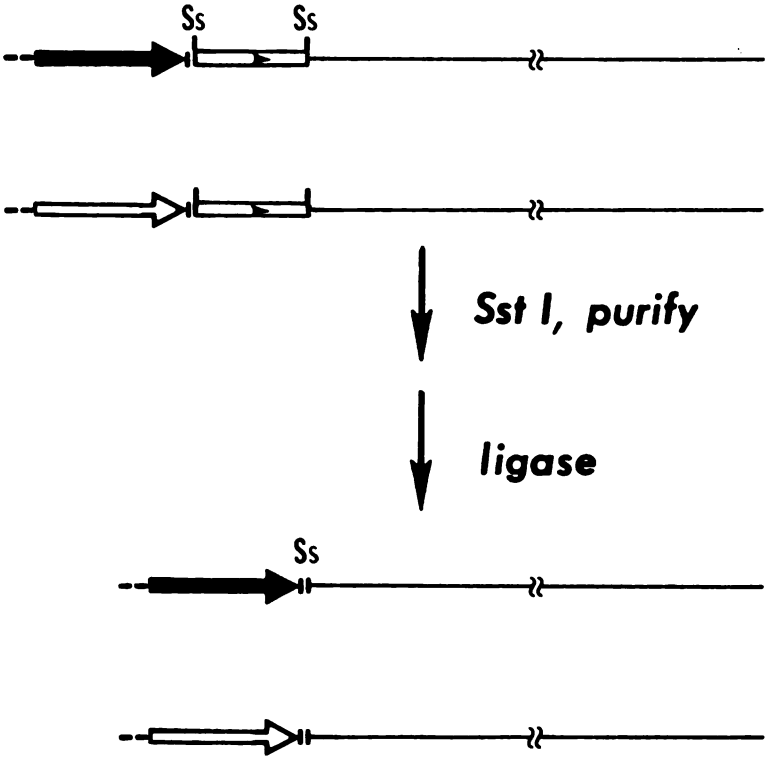



FIGURE 20

FIGURE 21: Construction of plasmids containing an internal deletion or duplication in the chymotrypsin 5' flanking region fragment. The relevant DNA sequence is shown. Numbers indicate nucleotide positions with respect to the chymotrypsin cap site. Details of the methods used are given in the text. The four base pair duplication is underlined, and the site of the four base pair deletion is marked with a triangle.

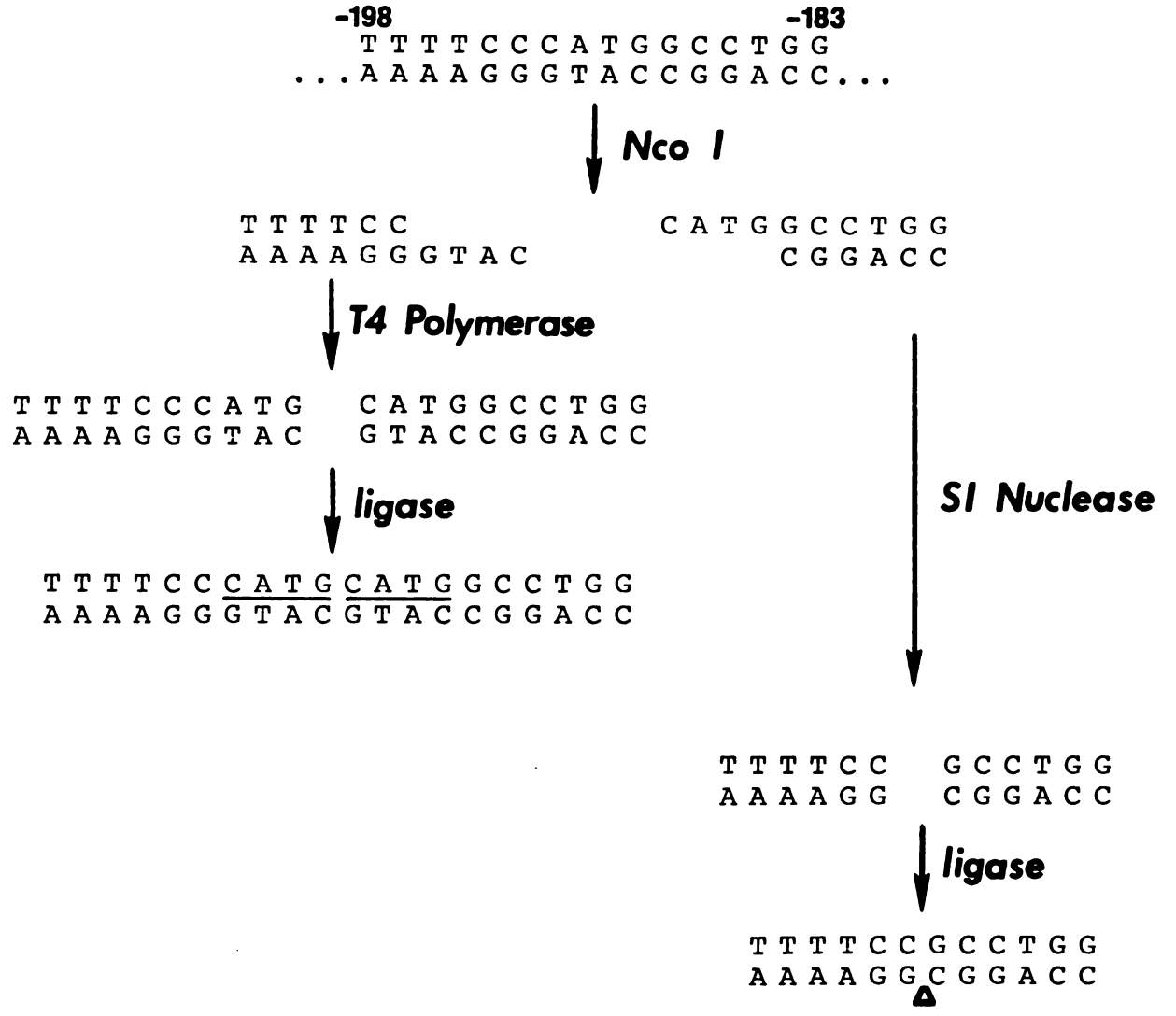

-198
-183
T T T T C C C A T G G C C T G G
. . . A A A A G G G T A C C G G A C C . . .
Nco I
T T T T C C
A A A A G G G T A C
C A T G G C C T G G
C G G A C C
T4 Polymerase
T T T T C C C A T G
A A A A G G G T A C
C A T G G C C T G G
G T A C C G G A C C
ligase
T T T T C C C A T G C A T G G C C T G G
A A A A G G G T A C G T A C C G G A C C
S1 Nuclease
T T T T C C
A A A A G G
G C C T G G
C G G A C C
ligase
T T T T C C G C C T G G
A A A A G G C G G A C C

FIGURE 21

FIGURE 22: Construction of plasmids in which the orientation of amylase fragments with 5' or 3' deletions is reversed. Symbols and restriction sites are as described in the legend to figure 13. The letters "bl" denote a blunt end; N/bl points out that the enzyme NruI produces blunt ends; B/Bg indicates that the cohesive BglII and BamHI ends have been ligated.

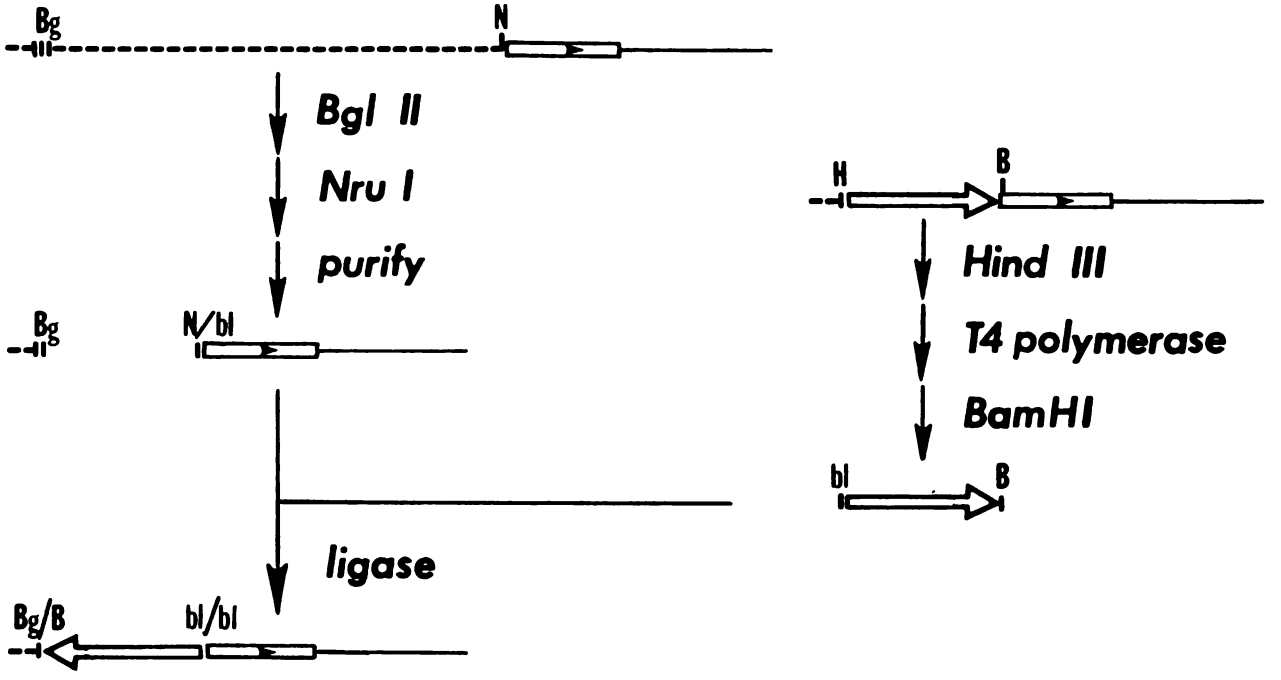

Bg
N
Bgl II
Nru I
purify
Bg
N/bl
ligase
Bg/B
bl/bl
H
B
Hind III
T4 polymerase
BamHI
bl
B

FIGURE 22

# APPENDIX II: METHODS

## Plasmid Construction and Isolation

In general, standard recombinant DNA procedures were used for the construction of plasmids described in Appendix I. Plasmids were isolated using the alkaline/SDS protocol of Maniatis et al. (1982). Plasmids for transfection experiments were purified by two cycles of banding in CsCl gradients.

Bal31 digestions were carried out essentially according to conditions suggested by the supplier. Exonuclease III/S1 nuclease deletions were produced according to the procedure described in Sakonju et al. (1980).

Dideoxy sequencing of double stranded plasmid DNA was carried out essentially as described in Haltiner et al. (1985) except that plasmid DNAs were linearized by restriction enzyme digestion prior to annealing of the primer (O. Wrange, unpublished).

## Cell Culture

Rat exocrine pancreatic and rat fibroblast cell lines were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and containing penicillin and streptomycin.

## Transfection of AR4-2J Cells

Cultured cell lines vary widely in their ability to take up and express exogenous DNAs. Initial experiments indicated that it is particularly difficult to obtain detectable levels of expression of exogenous DNAs transfected into AR4-2J cells. A number of modifications were introduced into the standard calcium phosphate coprecipitation

technique for transfection (Graham and van der Eb, 1973) which greatly improved the expression of plasmid DNAs containing promoter elements active in these cells. The best results were obtained when chloroquine was added to the media immediately following addition of the calcium phosphate-DNA precipitate, presumably reducing the degradation of DNA through inhibition of lysosomal enzyme function (Luthman and Magnusson, 1983). Glycerol shock treatment four hours after the addition of the DNA was also required for efficient transfection. The pH of the media and the concentration of $CO_2$ in the incubator while the cells are exposed to the calcium phosphate-DNA precipitate was found to be critical for optimal transfection efficiency. The pH of the media used for the AR4-2J cells is quite alkaline under 5% $CO_2$ incubation (pH about 8). By increasing the $CO_2$ concentration to 10%, a pH for the media of about 7.3 to 7.4 was attained, closer to that generally chosen to achieve optimal transfection efficiency using the calcium phosphate technique (for example, see Gorman et al., 1983). Since buffering the media to approximately the same pH value did not improve the transfection efficiency to the same extent as incubation under 10% $CO_2$ (only about 50% efficiency, not shown), it appears that the $CO_2$ concentration per se may be important for transfection. Figure 23 shows the signals obtained when the same DNA was transfected into cells incubated under different $CO_2$ concentrations. Transfection efficiency at 10% $CO_2$ is significantly greater than at either 5% or 7.5% $CO_2$ (approximately 6 to 7 fold).

Previous studies indicated that the amount of plasmid DNA added to each plate of cells could have a large effect on transfection efficiency. The amount of DNA which gave the highest levels of expression in mouse L cells was found to be about 10-20ug per 60mm dish

plated with $5x10^5$ cells about 24 hours prior to transfection (Loyter et al., 1982). The optimal amount of DNA might be different for AR4-2J cells under the slightly modified conditions used for transfection. $2.5x10^6$ cells, or 5 fold more than for parallel experiments using XC cells, are typically seeded to 60mm plates the day prior to transfection, a modification necessitated by the slow growth rate of AR4-2J cells. This number gives a good transfection efficiency and provides sufficient cells for harvesting 48 hours after transfection. In addition, Loyter et al. (1982) incubated L cells with the DNA calcium phosphate for 6 hours (relative to 4 hours for AR4-2J cells), and did not subject the cells to glycerol shock. The optimal amount of DNA was determined by preparing calcium phosphate precipitates with increasing amounts of test plasmid DNA, from 1 to 50ug. Figures 24 and 25 show the results of this experiment using two different plasmid DNAs, one that gives a very strong signal (RSV.CAT, figure 24) and one that gives a weaker signal (figure 25). The results show that 15ug of DNA per 60mm plate gives the greatest signal. In addition, the experiment shown in figure 24 was carried out in an incubator at 5-6% $CO_2$ while the experiment shown in figure 25 was done at 10% $CO_2$. Thus, it appears that the overall increase in transfection efficiency obtained by altering the incubation conditions does not change the amount of DNA which gives optimal signals.

CAT Assay

Cells were harvested with a pasteur pipet after brief treatment with 150mM NaCl, 10mM Tris-HCl, pH 7.5, 10mM EDTA, washed, and resuspended in 250mM Tris-HCl, pH 7.8. Crude cell extracts were prepared

by sonication and centrifugation to remove cell debris. Concentration of protein in cell extracts was determined by standard Lowry assays. CAT assay reactions contained, in a final volume of 180ul: 140mM Tris-HCl, pH 7.5, 2.2 or 4.4mM acetyl Coenzyme A, 0.2uCi C14-chloramphenicol, and cell extract containing 50ug of protein. Reaction mixtures were incubated at 37°C for 15 minutes to 12 hours depending on the DNAs, cells, and conditions used. Samples were extracted into 1ml ethyl acetate, dried, redissolved in 20ul ethyl acetate, and spotted onto TLC plates. Ascending chromatography in chloroform:methanol (95:5, v/v) was used to separate products from unreacted substrate. After exposure of X-ray film to visualize radioactive spots, activities were quantitated by counting appropriate regions in a liquid scintillation spectrometer. Results are expressed in terms of percent conversion of C14-chloramphenicol to the more abundant monoacetylated derivative.

RNA Preparation

RNA was prepared from transfected cells using the guanidine thiocyanate-lithium chloride procedure. Cells were harvested after trypsinization and washed once with 150mM NaCl, 10mM Tris-HCl, pH 7.5. Cells were taken up in guanidine thiocyanate lysis solution (5M GnSCN, 50mM Tris-HCl, pH7.5, 10mM EDTA), BME was added to 1.6M final concentration, and the solution was vortexed and incubated 15 minutes at room temperature. LiCl was then added to a final concentration of 3.3M, and the solution was incubated overnight at 4°C. The LiCl precipitate was spun down, washed with 3M LiCl, 4M urea and respun. The pellet was dissolved in 10mM Tris-HCl, pH 7.5, 1mM EDTA with 0.1% SDS, phenol extracted twice, chloroform/isoamyl alcohol extracted once, and ethanol

precipitated twice.

Sp6 RNA Mapping

The Sp6 radioactively labeled probe was prepared according to the protocol suggested by the supplier of the Riboprobe System (Promega Biotec) using alpha-$^{32}$P-rGTP. The probe template, pSp6.CATtk (constructed by M. Walker) consists of 147bp of CAT sequences fused to the tk promoter fragment (-109 to +55) inserted downstream from the Sp6 promoter so as to produce an antisense RNA transcript. RNAse mapping experiments followed the method described by the manufacturer with modifications from Zinn, et al (1983). 15ug of total RNA was hybridized with 100,000cpm of RNA probe in 30ul hybridization buffer for approximately 16 hours at 37°C. RNA hybrids were digested by addition of 350ul of RNAse digestion buffer containing 40ug/ml RNAse A and 2ug/ml RNAse T1 and incubation at 37°C for 60 minutes. Resistant RNA duplexes were analyzed on denaturing polyacrylamide-urea gels.

FIGURE 23: Effect of $CO_2$ concentration on the transfection efficiency of AR4-2J cells by the calcium phosphate coprecipitation method. Duplicate 60mm plates were transfected with 10ug of RSV.CAT plasmid DNA. CAT assays contained 30ug of crude extract protein except for 3a which contained only 15ug. Reactions were incubated at 37°C for 6.5 hours. The average CAT activities for each pair of plates were as follows: 5% $CO_2$, 9.6% conversion; 7.5% $CO_2$, 11.7% conversion; 10% $CO_2$, 69.1% conversion.

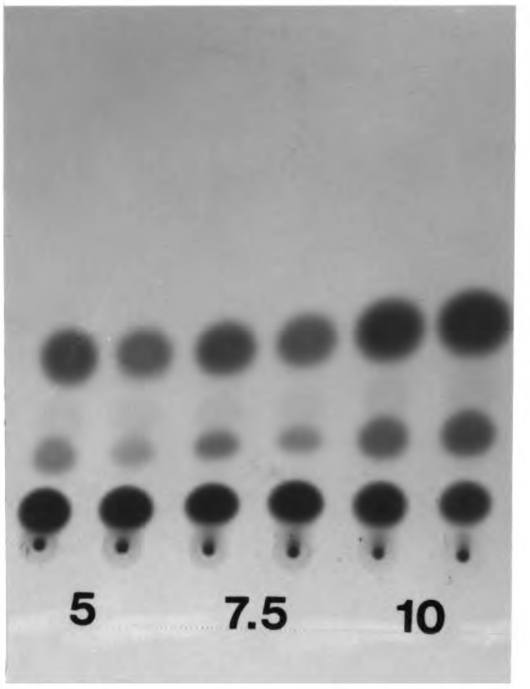
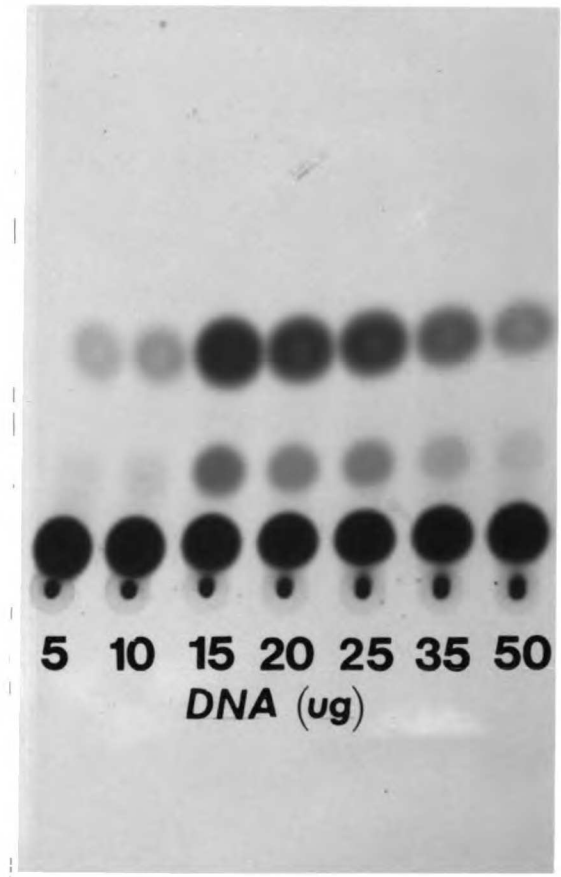



FIGURE 23

FIGURE 24: CAT activities obtained when increasing amounts of DNA are added per plate of cells. Individual 60mm plates of AR4-2J cells were transfected with the indicated amounts of RSV.CAT DNA. CAT assays containing 50ug of extract protein were allowed to proceed for 12 hours at 37°C. The graph below plots percent conversion of substrate versus amount of DNA added per plate.

FIGURE 24


5 10 15 20 25 35 50
DNA (ug)

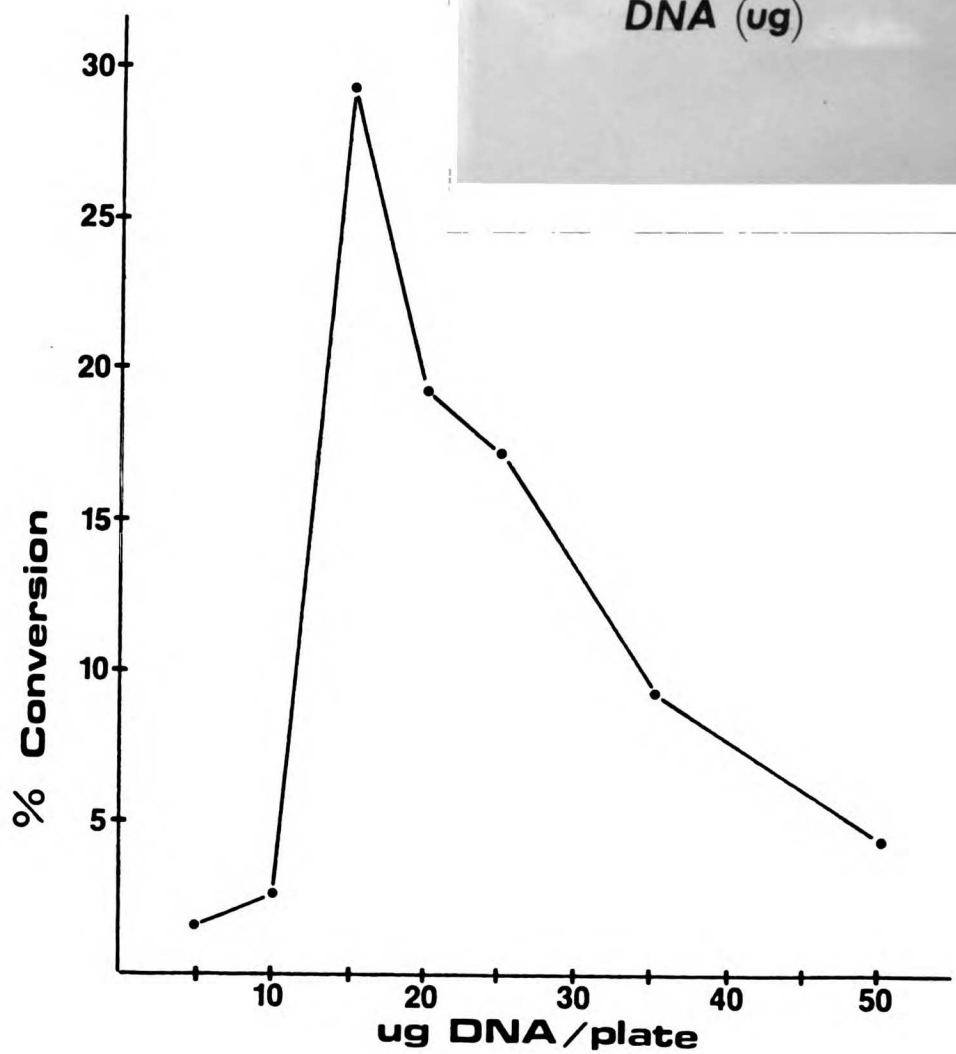
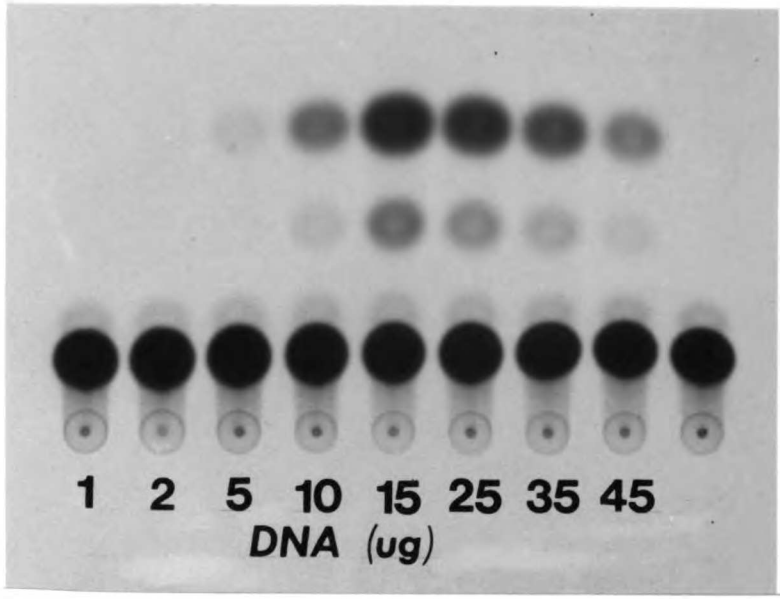
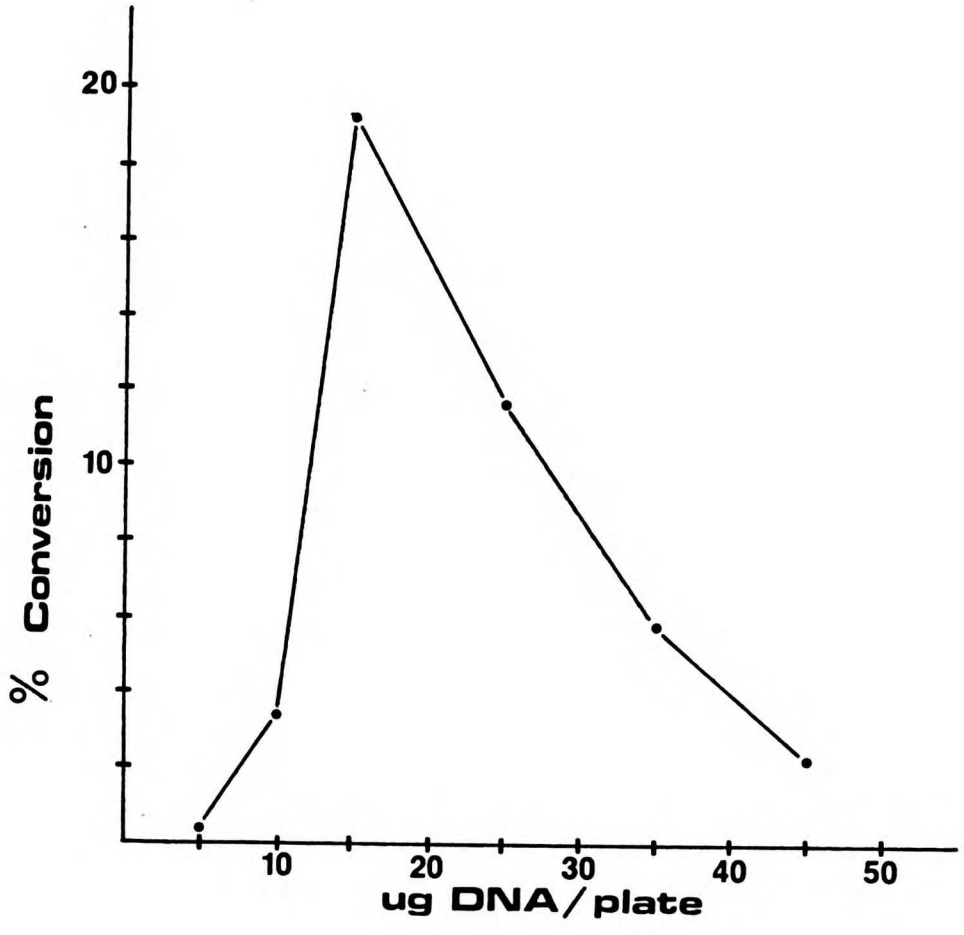

% Conversion
30
25
20
15
10
5
10
20
30
40
50
ug DNA/plate

FIGURE 25: CAT activities obtained when increasing amounts of DNA are added per plate of cells. Individual 60mm plates of AR4-2J cells were transfected with the indicated amounts of a pTE1 derivative in which the amylase enhancer fragment is located directly upstream from the tk promoter. CAT assays containing 50ug of crude extract protein were incubated at 37°C for 2.5 hours. The graph below plots percent conversion of substrate versus amount of plasmid DNA added per plate.

FIGURE 25


1 2 5 10 15 25 35 45
DNA (ug)


20
10
% Conversion
10 20 30 40 50
ug DNA/plate