The Role of Alternative Splicing in the Regulation of Cathepsin B Expression in Human Tumors

Qiuming Gong

Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Molecular Biology Commons

Recommended Citation
Gong, Qiuming, "The Role of Alternative Splicing in the Regulation of Cathepsin B Expression in Human Tumors" (1994). Dissertations. 3392.
https://ecommons.luc.edu/luc_diss/3392

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License. Copyright © 1994 Qiuming Gong
THE ROLE OF ALTERNATIVE SPLICING IN THE REGULATION
OF CATHEPSIN B EXPRESSION IN HUMAN TUMORS

By
Qiuming Gong

A Dissertation Submitted to the Faculty of the Graduate
School of Loyola University of Chicago in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy
January
1994
ACKNOWLEDGEMENTS

I would like to express sincere thanks to my advisor and mentor, Dr. Allen Frankfater. His guidance, patience and encouragement were essential to the completion of this dissertation.

I also wish to thank my committee members, Drs. Richard M. Schultz, Mark R. Kelley, Sally A. Amero, Gary L. Engelmann and Russell Pieper for their critical readings and discussions of this dissertation.

My appreciation is also extended to all members of the Department of Biochemistry for providing the means to complete this study.

Finally, I wish to acknowledge my parents, my husband and my son, whose sacrifice and understanding were of great help to me.
VITA

The author, Qiuming Gong, was born on August 24, 1957, in Beijing, China.

She entered Beijing Medical University in Beijing, China in March, 1978. In December of 1982, she received her Doctor of Medicine degree. In December of 1985, she received her Master of Science degree in Biochemistry. From December of 1985 to August of 1987, she worked as a research associate in the Department of Biochemistry at Beijing Medical University. In August of 1988, she enrolled in the Graduate School at Loyola University of Chicago in the Department of Molecular and Cellular Biochemistry.

Qiuming has completed her dissertation work under the direction of Dr. Allen Frankfater. She was the recipient of Arthur J. Schmitt Dissertation Fellowship in 1992-1993.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF PUBLICATIONS</td>
<td>xi</td>
</tr>
</tbody>
</table>

## CHAPTER

### I. INTRODUCTION

1

### II. REVIEW OF THE RELATED LITERATURE

6  

A. Cellular Proteinases  
6  
1. Serine Proteinases  
6  
2. Cysteine Proteinases  
7  
3. Aspartic Proteinases  
8  
4. Metallo-Proteinases  
9  
B. Cathepsin B  
9  
C. Correlation Between Cathepsin B Expression and Malignant Tumors  
15  
D. Alternative Pre-mRNA Splicing  
22  
1. Constitutive Splicing  
23  
2. Modes of Alternative Splicing  
25  
3. Factors Affecting Splicing-Site Selection  
27  
E. Functional Effects of Alternative Splicing of Pre-mRNA  
33  
1. Protein Localization  
34  
2. Modulation of Protein Function  
36  
3. Effects on RNA Stability and Translation Efficiency  
39

### III. METHODS

42  

A. Materials  
42  
B. Solutions  
44
CHAPTER

C. Cell Culture ........................................... 45
D. Human Tissues ......................................... 46
E. Total RNA Isolation .................................... 46
F. Quantification of Cathepsin B mRNA .................. 47
G. Northern Blot Analysis .................................. 49
H. RNase Protection Assay .................................... 49
I. Polymerase Chain Reaction (PCR) Amplification
   of cDNA ................................................... 53
J. Sequencing of PCR Fragments ............................ 57
K. Construction of Expression Vectors ...................... 58
L. In Vitro Transcription and Translation .............. 63
M. Analysis of the Translation Products .................. 64
N. Folding of an In Vitro Translation Product .......... 65
O. Measurement of Cathepsin B Activity .................. 66

IV. RESULTS .................................................. 69
A. Quantification of Total Cathepsin B mRNA in
   Human Normal Tissues and Tumors by Slot Blot ........ 69
B. Quantification of Variant Cathepsin B mRNAs by
   Northern Blot and RNase Protection Assay ............. 72
C. Amplification of the 5'-Ends of Cathepsin B
   cDNA from Human Normal Tissues and Tumors by
   PCR ...................................................... 77
D. Molecular Cloning of Human Cathepsin B cDNA
   Lacking Exons 2 and 3 .................................. 81
E. Detection of Cathepsin B mRNA that is missing
   exons 2 and 3 by PCR .................................... 83
F. In Vitro Translation of Variant Cathepsin B
   mRNA ...................................................... 88
G. Refolding of a Truncated Cathepsin B Produced
   by the In vitro Translation of mRNA Lacking
   Exons 2 and 3 ........................................... 96

V. DISCUSSION ................................................. 101
A. Multiple Forms of Cathepsin B Transcripts in
   Human tissues ........................................... 101
B. Levels of Expression of Variant Cathepsin B
   mRNAs in Human Normal Tissues and Tumors .......... 107
C. Mechanisms of Alternative Splicing .................... 111
D. Comparison of Translational Activities of
   Cathepsin B mRNAs with Variant 5'-UTRs ............. 115
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. Characteristics of Cathepsin B Transcript</td>
<td></td>
</tr>
<tr>
<td>Missing Exons 2 and 3</td>
<td>120</td>
</tr>
<tr>
<td>VI. SUMMARY</td>
<td>124</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>127</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relative levels of cathepsin B mRNA in human tissues and tumor cells</td>
</tr>
<tr>
<td>2. Relative proportion of 2.3 Kb and 4.0 Kb cathepsin B mRNAs</td>
</tr>
<tr>
<td>3. Relative proportion of cathepsin B mRNA with and without exon 2</td>
</tr>
<tr>
<td>4. Exon-intron organization of the human cathepsin B gene</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Organization of the cathepsin B gene and its mRNA transcript</td>
<td>14</td>
</tr>
<tr>
<td>2.</td>
<td>Human cathepsin B cDNAs in plasmids phCB79-2 and phCB96-2</td>
<td>43</td>
</tr>
<tr>
<td>3.</td>
<td>Scheme for RNase protection assay</td>
<td>50</td>
</tr>
<tr>
<td>4.</td>
<td>The riboprobe for RNase protection assay</td>
<td>51</td>
</tr>
<tr>
<td>5.</td>
<td>Scheme for PCR amplification of 5'-end of cathepsin B cDNA</td>
<td>55</td>
</tr>
<tr>
<td>6.</td>
<td>Scheme for the PCR amplification of cathepsin B cDNA missing exons 2 and 3</td>
<td>56</td>
</tr>
<tr>
<td>7.</td>
<td>Scheme for the construction of plasmid phCB79-2-2</td>
<td>60</td>
</tr>
<tr>
<td>8.</td>
<td>Scheme for the construction of plasmid phCB79-2-2.3</td>
<td>61</td>
</tr>
<tr>
<td>10.</td>
<td>Slot blot hybridization analysis of human normal tissues and cultured melanoma cells for human cathepsin B mRNA</td>
<td>70</td>
</tr>
<tr>
<td>11.</td>
<td>Northern blot of cathepsin B mRNA from human normal tissues and cultured melanoma cells</td>
<td>75</td>
</tr>
<tr>
<td>12.</td>
<td>RNase protection of variant 5'-ends of human cathepsin B mRNAs</td>
<td>78</td>
</tr>
<tr>
<td>13.</td>
<td>Evaluation of the riboprobe used in the RNase protection assay</td>
<td>80</td>
</tr>
<tr>
<td>14.</td>
<td>PCR amplification of the 5'-ends of cathepsin B cDNA from human normal tissues and tumors</td>
<td>82</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>PCR amplification of 5'-end of cathepsin B cDNA missing exons 2 and 3</td>
<td>84</td>
</tr>
<tr>
<td>16.</td>
<td>Sequence of the 306 bp PCR product amplified from cathepsin B cDNA in A375M cells</td>
<td>85</td>
</tr>
<tr>
<td>17.</td>
<td>Detection cathepsin B mRNA that lacks exons 2 and 3 by PCR</td>
<td>87</td>
</tr>
<tr>
<td>18.</td>
<td>Analysis of SP6 directed variant cathepsin B transcripts</td>
<td>90</td>
</tr>
<tr>
<td>19.</td>
<td>Rates of in vitro translation of synthetic cathepsin B mRNAs containing variant 5'-UTRs</td>
<td>92</td>
</tr>
<tr>
<td>20.</td>
<td>SDS-PAGE electrophoresis analysis of in vitro translation products from cathepsin B mRNAs with and without exon 2</td>
<td>94</td>
</tr>
<tr>
<td>21.</td>
<td>In vitro translation of synthetic cathepsin B mRNAs with variant 5'-ends</td>
<td>97</td>
</tr>
<tr>
<td>22.</td>
<td>Refolding of a truncated cathepsin B produced by the in vitro translation of mRNA missing exon 2 and 3</td>
<td>99</td>
</tr>
<tr>
<td>23.</td>
<td>Variant cathepsin B transcripts in human tissues</td>
<td>102</td>
</tr>
<tr>
<td>24.</td>
<td>Organization of human cathepsin B gene</td>
<td>104</td>
</tr>
</tbody>
</table>
LIST OF PUBLICATIONS


CHAPTER I

INTRODUCTION

Metastasis is believed to involve a series of sequential events which include intravasation, extravasation and invasion, during which tumor cells must penetrate various host tissue barriers. This has led to a hypothesis that metastasis may require the enzymatic degradation of structural components of the extracellular matrix and basement membrane. Consistent with this hypothesis, many studies have shown increased levels of proteolytic enzymes produced by tumor cells. These include metalloproteinases such as collagenase (Fessler et al. 1984), serine proteinases such as tissue type and urokinase type plasminogen activators (Moscatelli et al. 1980; Schmitt et al. 1992), and several lysosomal proteinases including cathepsin B (Sloane et al. 1981; Recklies et al. 1982; Dufek et al. 1984; Qian et al. 1989), cathepsin L (Gal and Gottesman, 1986; Troen et al. 1988) and cathepsin D (Capony et al. 1987; Maudelonde et al. 1988).

Among these proteinases which may play a role in the metastatic process, cathepsin B has been extensively investigated. Cathepsin B is capable of degrading many of the protein constituents in the extracellular matrix, such as fibronectin, laminin, and type IV collagen and others (Sloane,
1990; Delaisse et al. 1991; Buck et al. 1992; Cardozo et al. 1992). Cathepsin B can also activate a number of latent extracellular proteinases such as procollagenase (Eeckhout and Vaes, 1977) and proplasminogen activator (Kobayashi et al. 1991; Schmitt et al. 1992). Consequently, were cathepsin B to be secreted by tumor cells, it could cause the degradation of the structural barriers through which tumor cells must migrate during invasion and metastasis (Docherty et al. 1983).

Elevated levels of cathepsin B are often associated with cancer (Sloane, 1990). In addition, cathepsin B is not exclusively targeted to lysosomes in tumor cells. It has been reported to be in the nucleus (Pietras and Roberts, 1981) where it could act to modify proteins which regulate DNA replication and gene expression. Cell surface and secreted forms of cathepsin B are also well documented in malignant tumors. In rodents, levels of intracellular and secreted forms of cathepsin B have been correlated with metastatic potential in murine B16 melanoma variants (Sloane et al. 1981; Sloane et al. 1982; Qian et al. 1989), in BDX rat sarcoma variants (Koppel et al. 1984) and Lewis lung tumors (Takenaga, 1984), and in humans, in liver (Ohsawa et al. 1989), colorectal (Keppler et al. 1988a; Maciewicz et al. 1989; Sheahan et al. 1989), gastric (Watanabe et al. 1989) and lung tumors (Kasafirek et al. 1990). Higher molecular weight latent forms of cathepsin B have been detected in body fluids of patients with primary liver cancer, ovarian cancer (Mort et al. 1983;
Dufek et al. 1984) and breast cancer (Petrova-Skalkova et al. 1987). Forms of cathepsin B associated with the cell surface have been reported for human, rat, and murine tumors (Recklies et al. 1982; Mort et al. 1983; Koppel et al. 1984; Dufek et al. 1984; Petrova-Skalkova et al. 1987; Keren and LeGrue, 1988), and have been correlated with tumor invasiveness and/or metastasis in humans (Weiss et al. 1990) and rodents (Sloane et al. 1986). These and similar findings suggest a direct or indirect role for cathepsin B in facilitating tumor metastasis through its ability to degrade the extracellular matrix and to release tumor cells from solid tumors. However the molecular basis for the increased production of cathepsin B and its abnormal localization to cell membranes, nuclei and secretory vesicles of tumor cells is unknown.

Eukaryotic pre-mRNAs undergo a series of nuclear processing events. Splicing of all intron sequences must occur with a high degree of accuracy in order for functional proteins to be generated. However, flexibility in the selection of a splice site, which forms the basis for the alternative splicing, is not unexpected. Alternative splicing can produce different but related messages. The variant proteins derived from these messages could then be targeted to different cellular and subcellular locations (Kwiatkowski et al. 1986; Johnson et al. 1990; Alt et al. 1980). Alternative splicing can also produce messages which vary in their 5' or 3'-untranslated regions (UTRs) and consequently affect mRNA
stability and translational efficiency (Kozak, 1991a,b; Brawerman, 1987). Thus, alternative processing of the cathepsin B pre-mRNA may be one of many possible factors which can be advanced for differences in cathepsin B protein levels between normal tissues and tumors, and for the presence of unusual tumor forms of cathepsin B protein which might be directed to the nucleus, the cell surface or the cell exterior.

Cathepsin B mRNA cloned from human kidney is heterogeneous due to alternative processing of the 5'- and 3'-ends of the primary transcript, resulting in mRNAs which contain alternative 5'- and 3'-UTRs (Gong et al. 1993). The variable removal of exon 2 produces transcripts which differ by 88 bases in the lengths of their 5'-UTR. The sequence of exon 2 has features which could decrease the translational efficiency of the transcript. Consequently, were exon 2 more frequently absent from the 5'-UTR of tumor cell cathepsin B mRNAs, it could explain higher levels of cathepsin B in tumors. In this present work we have found that some human tumors contain a third class of transcripts which lack exon 3. As exon 3 contains the coding sequence for the signal peptide required to initially direct cathepsin B to lysosomes, and a portion of the inhibitory propeptide, mRNA lacking exon 3 could be responsible for the presence of active cathepsin B forms at the cell surface and/or within the nuclei of human tumors.
It is my hypothesis that the higher levels of cathepsin B found in tumor cells is due to differences in cathepsin B pre-mRNA processing between normal tissues and tumor cells. Therefore, the specific aims of the present research are: 1) to determine relative abundances of different mRNA forms between normal human tissues and human tumors; 2) to determine translational activities of the various cathepsin B mRNAs which differ in the length of their 5'-UTRs in an in vitro translation assay; 3) to clone the transcript lacking exon 3 and to study its translational potential in an in vitro assay system; and 4) to determine whether the putative protein products of messages which lack exon 3 possess peptidolytic activity.
CHAPTER II

REVIEW OF THE RELATED LITERATURE

A. Cellular Proteinases

Proteases include two groups: exopeptidase (peptidases) which cleave either amino- or carboxyl-terminal amino acids or dipeptides from proteins, and endopeptidases (proteinases) which cleave peptide bonds in the interior of protein molecules. However, some proteases can have both exo- and endo-peptidase activities.

Proteases can also be subdivided into five classes based on the mechanisms of peptide bond cleavage catalyzed by their active center (Barrett, 1980a): serine proteinases, cysteine proteinases, aspartic proteinases, metalloproteinases, and a fifth category which includes enzymes with different or undetermined catalytic mechanisms. Each of the first four classes will be briefly discussed, and then cathepsin B will be introduced.

1. Serine Proteinases

Serine proteinases include trypsin, chymotrypsin, elastase, coagulation factors, plasmin, cathepsin G, chymases, tryptases, plasminogen activators and others. Sequence
homologies among the serine proteinases suggest that they may have a common evolutionary origin (Hartley, 1970). Serine proteinases all have a serine in their active sites and are active at neutral pH without the need of cofactors. An adjacent histidine residue which functions as a general acid-general base catalyst is essential for the activity of these enzymes (Walsh, 1979). They are synthesized as inactive processors that require limited proteolysis for activation. Serine proteinases, which participate in coagulation, are controlled by a cascade of activating enzymes and a system of protein proteinase inhibitors. Serine proteinases that have been of particular interest in cancer metastasis are plasmin and the pro-urokinase and tissue type plasminogen activators.

2. Cysteine Proteinases

Cysteine proteinases contain a cysteine residue in their active sites. An adjacent histidine residue which functions as a general acid-general base catalyst is also essential for the activity of this class of enzymes (Topham et al. 1990). This subclass of proteinases includes lysosomal enzymes (cathepsin B, H and L) (Barrett, 1980b), the plant proteinase papain, and the cytosolic calcium-activated neutral proteinases, calpains (Barth and Elce, 1981; Dayton and Schollmeyer, 1981). The amino acid sequence of papain and of cathepsin B and H have been found to be significantly homologous (Takio et al. 1983). However, cathepsin H and papain appear more closely related than are papain and
cathepsin B or cathepsin H and cathepsin B. The active sites of cathepsin B and papain are functionally homologous as well as structurally homologous (Bajkowski and Frankfater, 1983a; Bajkowski and Frankfater, 1983b). The lysosomal cysteine proteinases are synthesized as larger inactive precursors which undergo proteolytic maturation. Their mature forms are normally confined to the lysosome and thought to play an important role in intracellular protein degradation. The lysosomal cysteine proteinases exhibit pH optima of about 6.0, and are unstable above pH 7 (Barrett and Kirschke, 1981; Barrett, 1984). However, some lysosomal cysteine proteinases, such as cathepsin B, are secreted into extracellular environment by tumor cells, and appear to retain activity at and above pH 7.0 (Mullins and Rohrlich, 1983).

3. Aspartic Proteinases

Members of the family of aspartic proteinases include pepsin, renin and cathepsin D. All of these enzymes contain aspartate in their active sites, and are synthesized as inactive precursors which are processed either autocatalytically (e.g. pepsin) or by other enzymes (e.g. renin) to remove an N-terminal propeptide. Pepsin, which functions in the acidic milieu of the stomach, is responsible for the initial steps in protein digestion. Cathepsin D is a lysosomal enzyme and plays an important role in intracellular proteolysis in the acidic environment of the lysosome. Renin is thought to have originated from the hypothetical cathepsin
D-like archetype. However, unlike cathepsin D it is stable and active at neutral pH, and is highly specific for the Leu-Leu bond that generates angiotensin I from angiotensinogen (Reid et al. 1978).

4. Metalloproteinases

Metalloproteinases are characterized by the presence of Zn$^{2+}$ in their active sites. Examples of this group include pancreatic carboxypeptidase, thermolysin and collagenases. The metalloproteinases have a neutral to slightly alkaline pH optimum, and their activities can be inhibited by EDTA and thiols. They seem well suited to function extracellularly.

Type IV collagenase, a metalloproteinase, has been found to be secreted by a variety of malignant tumors (Liotta et al. 1982; Fessler et al. 1984). The type IV collagenase degrades the basement membrane type IV collagen molecule at a single site in the helical domain at about one fourth the distance from the N-terminal end (Fessler et al. 1984). Accordingly, the enzyme may be able to decompose the collagen network of basement membrane and thus facilitates penetration by tumor cells.

B. Cathepsin B

Cathepsin B is a cysteine proteinase located in the lysosomes of mammalian cells. Cathepsin B has been isolated from tissues of various mammalian species, including rat liver (Otto, 1971), bovine spleen and liver (Otto, 1971), and human
spleen and liver (Barrett, 1973). Using immunochemical methods, cathepsin B has been located within macrophages (Howie et al. 1985), fibroblasts (Mort et al. 1981), neurons (Howie et al. 1985), myocytes (Bird et al. 1985) and epithelial cells (Howie et al. 1985) of various tissues. Cathepsin B is thought to play an important role in a number of physiological and pathological processes including: degradation of endocytosed proteins (Huisman et al. 1974; Bystryn and Perlstein, 1982) and intracellular proteins (Kar and Pearson, 1977; Shaw and Dean, 1980); processing of secreted proteins (Quinn and Judah, 1978), enzymes (Figarella et al. 1988), hormones (Docherty and Steiner, 1982) and neuromodulatory polypeptides (Docherty and Steiner, 1982); tissue resorption and remodeling (Bayliss and Ali, 1978); inflammation (Daries et al. 1978) and mediation of hormonal response (Pietras et al. 1975). At present, the full spectrum of the biological role of cathepsin B is still not very clear but its broad tissue distribution suggests it has many important functions.

Cathepsin B is both an endopeptidase and a peptidyl dipeptidase (Aronson and Barrett, 1978). Below pH 5, cathepsin B is an exopeptidase (Polgar and Csoma, 1987) catalyzing the successive removal of dipeptides from the carboxyl terminus of proteins (Aronson and Barrett, 1978). Between pH 5 and 7, cathepsin B shows endopeptidase activity (Polgar and Csoma, 1987) with a broad specificity (Barrett and
Kirschke, 1981; Mason, 1989). Synthetic substrates containing the \(-\text{Arg-Arg-}\) sequence, such as Benzoyl-L-arginine-L-arginine-2-naphthylamide, are extremely sensitive to cathepsin B (McDonald and Ellis, 1975). Fluorogenic and chromogenic synthetic substrates are the most common substrates used in enzyme assays for determination of cathepsin B activity (Barrett, 1973; Bajkowski and Frankfater, 1975).

Mature cathepsin B has a molecular weight of 27,000 – 29,000. It is derived from a larger precursor, procathepsin B, which in its glycosylated state has a molecular weight of about 43,000 (Barrett, 1987; Nishimura et al. 1988). Proteolytic processing of the precursor yields both an active single chain form and a two chain form (Hanewinkel et al. 1987). Procathepsin B can also be activated \textit{in vitro} by limited proteolytic cleavage with pepsin (Achkar et al. 1990; Rowan et al. 1992). Procathepsin B is in turn derived from preprocathepsin B which is synthesized on the rough endoplasmic reticulum. Preprocathepsin B is converted to procathepsin B by cleavage with a signal peptidase which removes the N-terminal signal peptide in the lumen of the rough endoplasmic reticulum.

Procathepsin B is directed to the lysosome where it is converted to biologically active mature forms (McDonald and Ellis, 1975). In fibroblasts, it is known that mannose-6-phosphate receptors segregate newly synthesized lysosomal enzymes from secretory proteins in the \textit{trans} Golgi network.
The lysosomal enzyme-receptor complex is delivered to an intermediate acidified compartment where dissociation of the complex occurs. Receptors are then returned to the Golgi apparatus and cell surface for reuse while lysosomal enzymes in turn are transferred to the lysosome.

There have been many reports of cathepsin B-like activities associated with polypeptides of Mr greater than 26,000-27,000. Steiner and coworkers have shown that $^{125}$I-tyrosyl-Arg-ArgCH$_2$Cl, a cathepsin B substrate analog, covalently labeled three polypeptides of Mr 39,000, 31,500 and 25,000 in rat islet cells (Docherty et al. 1983) and in a transplantable rat insulinoma (Docherty et al. 1984). The Mr = 31,500 form cross-reacted with anti-rat liver cathepsin B antibodies. Pepsin treatment of the Mr = 39,000 form converted it to a 31,500 protein which was also precipitated with anti-rat liver cathepsin B antibodies, suggesting that the 39,000 and 31,500 polypeptides are related as precursor to product. These workers also found that cathepsin B in rat islets is synthesized initially as a precursor of approximate Mr = 43,000. In the insulinoma the Mr = 39,000 form was found in insulin secretory granules, the Mr = 31,500 form was found in secretory granules and lysosomes, whereas the Mr = 25,000 species was only found in the lysosomal fraction. Consequently, it has been suggested that the larger precursor forms of cathepsin B are sorted into secretory granules, possibly as a consequence of their incomplete targeting to
lysosomes via the mannose-6-phosphate pathway.

Preprocathepsin B cDNAs of human (Fong et al. 1986), rat (Chan et al. 1986) and mouse (Fong et al. 1986) have been cloned and sequenced. Each predicts a primary structure of 339 amino acids. The amino acid sequences of the rat and mouse enzymes are similar, with 90% identity. The human and rat amino sequences are 84% identical in the regions of the mature proteins, and about 50% identical in the preproregions and C-terminal peptide extensions.

The entire human cathepsin B gene has been cloned and its structure elucidated. The 350 nucleotides preceding the transcription start site, large portions of introns 1 and 2 (Gong et al. 1993), and all 12 exons have been sequenced (Fong et al. 1986). There are several Sp1 binding sites in the 5' flanking region and in intron 1 but no TATA box appropriately positioned with respect to the transcription initiation site.

Cathepsin B cDNA cloned from human tissues is heterogeneous due to differences in processing of the 5' and 3'-ends of the pre-mRNA (Figure 1). Alternative splicing, "exon 2 skipping", produces mRNAs which differ by 88 bases in the lengths of their 5'-untranslated regions. The occasional selection of an alternative intron donor site in exon 11 can result in the loss of an adjacent polyadenylation signal in exon 11 and termination after a downstream signal in exon 12. This produces two kinds of transcripts (2.3Kb and 4.0Kb) which vary in the lengths of their 3'-untranslated regions.
Figure 1. Organization of the cathepsin B gene and its mRNA transcript. Intron positions (†) and polyadenylation signals (△) are identified. The coding regions for the prepeptide, activation propeptide and mature enzyme are indicated by boxes. Also shown are the sizes of the first six exons.
The mouse cathepsin B gene and its 5'-upstream region has also been cloned (Qian et al. 1991b). The gene has 10 exons and 9 introns spanning about 20 kb. Exon 10 contains three potential polyadenylation sites which are alternatively used in transcription of cathepsin B mRNA in B16 melanoma. Like that of human cathepsin B, the 5' upstream region of the mouse cathepsin B gene has several Sp1 binding sites and lacks a TATA box in the immediate vicinity of the transcription start site (Qian et al. 1991b).

C. Correlation Between Cathepsin B Expression and Malignant Tumors

Sufficient evidence has accumulated to implicate proteolytic enzymes as playing an important role in tumor metastasis. Proteinases, either on the cell surface or released by tumor cells, are thought to facilitate tumor cell colonization (Yee and Shiu, 1980; Mainferme et al. 1985). Consequently, many studies have focused on enzymes that are known to degrade connective tissue constituents, such as collagenase (Liotta et al. 1979), neutral proteinase (Edmonds-Alt et al. 1980), plasminogen activator (Laug et al. 1983), glycosidase (Kramer et al. 1986) and cathepsin B (Recklies et al. 1982; Qian et al. 1989). There may also exist a proteolytic cascade in which one or more proteinases act in concert and a newly activated protein in turn may activate others (Kramer et al. 1986; Kobayashi et al. 1992). In the case of cathepsin B, increased enzyme activity has
been found in a large number of metastatic tumors. Elevated cathepsin B activity has been reported in humans for liver (Dufek et al. 1985), colorectal (Dufek et al. 1984; Keppler et al. 1988a; Maciewicz et al. 1989; Sheahan et al. 1989), gastric (Watanabe et al. 1989), breast (Poole et al. 1978; Recklies et al. 1980; Dengler et al. 1991; Lah et al. 1992), lung (Kasafirek et al. 1990) and cervical cancers (Pietras and Roberts, 1981). In addition, positive correlations have been observed between intracellular levels of cathepsin B and the metastatic phenotype in the murine B16 melanoma variants (Sloane et al. 1981; Sloane et al. 1982), in cultured three Lewis lung clones (Takenaga, 1984) and in cultured BDX rat sarcoma variants (Koppel et al. 1984). The presence of the elevated cathepsin B activity in tumor cells suggests a direct or indirect role for this enzyme in the degradation of the extracellular matrix and in the release of tumor cells from solid tumors. Cathepsin B can degrade fibronectin, laminin and type IV collagen (Sloane, 1990), and can activate latent collagenase (Eeckhout and Vaes, 1977) and pro-urokinase type plasminogen activator (Kobayashi et al. 1991).

Further evidence in support of a role for cathepsin B in tumor invasion is the observation of increased cathepsin B activity in plasma membrane of tumor cells, at the margins of actively growing tumors in direct contact with the extracellular matrix, and the observation of the release of cathepsin B from tumor cells. Cathepsin B activity has been
found in the plasma membrane fraction from anaplastic sarcoma in BDX rat (Koppel et al. 1984), from murine fibrosarcomas (Keren and LeGrue, 1988), from mouse B16 melanoma variants in proportion to the metastatic potential of the variants (Sloane et al. 1986), and from a human pancreatic adenocarcinoma (Zucker et al. 1985). Cathepsin B activity is localized at the invasion front of a rabbit V2 carcinoma (Graf et al. 1981). The secreted enzymes are very similar to lysosomal cathepsin B in terms of their reactivity to synthetic substrates. It was shown that cultured explants of malignant human breast tumors released eleven times more cathepsin B activity than normal breast tissue or nonmalignant tumors (Poole et al. 1978; Recklies et al. 1980). Other reports indicated elevation of cathepsin B activity in the serum of women with genital tract cancer especially in those patients with invasive carcinomas (Pietras et al. 1979; Pietras and Roberts, 1981), in malignant ascites fluid from patients with primary liver and ovarian cancer (Mort et al. 1983; Dufek et al. 1984), in pleural effusions from patients with breast cancer (Petrova-Skalkova et al. 1987), and in conditioned media from human and murine tumors (Recklies et al. 1982; Mort et al. 1983; Qian et al. 1989).

The secreted cathepsin B-like enzyme activities are different from lysosomal cathepsin B in that they have higher molecular weights. Two extracellular high molecular weight forms of human cathepsin B with Mr 40,000 have been observed,
one is latent, the other active. In ascitic fluid and culture medium of ascites cells from cancer patients, a latent form of cathepsin B was discovered which could be activated by pepsin pretreatment (Mort et al. 1981). In addition, malignant human breast tumors in organ culture release a high molecular weight form of cathepsin B that was detected by its ability to hydrolyze synthetic substrates for which lysosomal cathepsin B has high specificity (Mort et al. 1980). The advantages that tumor cells gain by secreting latent and high molecular weight forms of lysosomal proteinase may be related to the instability of low molecular weight lysosomal form above pH 7, resulting in rapid denaturation and inactivation in serum. In contrast, high molecular weight latent and active forms of cathepsin B are more stable under these same conditions (Mort and Rechlies, 1986). In addition, secretion of a latent enzyme, which is then activated some distance from tumor cells, may protect the enzyme from inactivation by endogenous inhibitor while undergoing transport to its site of action. Thus, the active, high molecular weight secreted form of cathepsin B is not inhibitable by plasma proteinase inhibitor, α2-macroglobulin, whereas the lysosomal form is inhibitable (Mort and Rechlies, 1986). Schultz and his associates found that leupeptin, an inhibitor of cathepsin B did not prevent the invasion of B16-F10 melanoma cells across basement membrane in vitro (Persky et al. 1986) or lung colonization in vivo (Ostrowski et al. 1986). In contrast, Belleli et al
(1990) recently found that papain-immunized mice possessed serum antibodies which cross-react with cathepsin B like endopeptidase isolated from B16 melanoma cells. The growth rate, invasion and metastasis of both B16 melanoma and the Lewis lung carcinoma were inhibited in mice immunized with papain. It may be that the secreted and/or membrane associated forms of cathepsin B, which might be insensitive to inhibition by leupeptin but can react with anti-papain antibodies, play a role in the metastasis of B16 melanoma cells.

The molecular basis for increased cellular and secreted levels and altered cellular distribution of cathepsin B in tumor cells is not well understood. Possible factors may include differences in transcription, post-transcriptional processing, translation, and post-translational processing of cathepsin B. Comparisons of mRNA levels for five lysosomal proteinases, cathepsin B, D, H, L, and S in B16 melanoma cells, in rat W-256 carcinoma cells, and in normal rodent tissues revealed that only the mRNA for cathepsin B is elevated in the malignant tumors (Qian et al. 1989; Qian et al. 1990). A correlation between the metastatic potentials of several B16 melanoma variants and cathepsin B mRNA levels has been observed. No evidence was found for gene amplification or gross chromosome rearrangement, or for the presence of more than one gene for cathepsin B in melanoma cells. These findings together with recent results of nuclear run off
assays (F. Qian, unpublished results) suggest that in these cells cathepsin B expression is regulated, at least in part, at the level of RNA transcription.

In addition to the normal 2.2 Kb transcript, two larger transcripts (4.0 and 5.0 Kb) were found in B16 melanomas. These larger transcripts were not seen in normal mouse tissues (Qian et al. 1989). A comparison of the cDNAs derived from the larger transcripts with the cathepsin B gene and cDNA derived from normal message revealed that all three transcripts from mouse melanoma cells encoded for the same form of cathepsin B and differed only in the lengths of their 3'-untranslated regions. The differences in the 3'-UTRs was due to the utilization of alternative polyadenylation signals during the processing of the 3'-end of cathepsin B pre-mRNA, the larger transcripts resulting from the occasional use of downstream polyadenylation signals in melanoma cells (Qian et al. 1991b; Qian et al. 1991c). Thus, the secretion and altered distribution of cathepsin B in murine melanoma cells is not due to alteration in the gene or in the pre-mRNA processing, but more likely due to aberrant proenzyme processing and sorting.

Mannose-6-phosphate receptors and mannose-6-phosphate-containing sugar side chains are required for correct delivery of enzymes to lysosomes (Kornfeld, 1987). Alterations in N-linked oligosaccharides have been found in proprecathespin D secreted by MCF7 mammary cancer cells (Capony et al. 1989).
The secretion of procathepsin L by Kirsten sarcoma virus-transformed fibroblasts (KNIH cell) is reported to be due to poor binding of cathepsin L to mannose-6-phosphate receptor, together with saturation of the mannose-6-phosphate receptor resulting from cathepsin L overexpression (Dong et al. 1989). Moloney murine sarcoma virus-transformed fibroblasts (MMSV cell) express 3-4 fold higher levels of cathepsins B, D, and L, but not H and S (Qian et al. 1991a). In addition, MMSV cells appear to have a general defect in lysosomal enzyme sorting which causes the increased secretion of latent cathepsin B, cathepsin L and an active β-glucuronidase (Achkar et al. 1990). This enhanced lysosomal enzyme secretion was correlated with an absence of cell surface mannose-6-phosphate receptor activity (Achkar et al. 1990) and an alkalinization of vacuolar compartments, which should prevent lysosomal enzyme-mannose-6-phosphate receptor uncoupling (Pracht et al. 1991). Agents which acidified MMSV cells strongly inhibited cathepsin B secretion and restored cellular levels to amounts seen in normal fibroblasts. These results agree with the finding that estrogen increases the secretion of cathepsins D, L, and B by human mammary carcinoma MCF-7 cells by down-regulating mannose-6-phosphate receptor expression (Mathieu et al. 1991).

In contrast to the above results in mouse, cathepsin B expression in human tissues may be regulated at the level of mRNA processing. Multiple forms of the cathepsin B mRNAs
which differ in their 3' and 5'-untranslated regions were also found in human tissues. These different forms of mRNA arise from alternative splicing of the 5' and 3'-untranslated regions (Figure 1). Examples of messages with alternative 5'-untranslated regions have been described (Young et al. 1981; Lowe et al. 1987; Shen et al. 1988; Nonaka et al. 1989; Horiuchi et al. 1990; Waterhouse et al. 1990). Many reports have indicated that changes in the 5'-noncoding sequence of a message can affect the translational efficiency of the message (Rao et al. 1988; Horiuchi et al. 1990; Waterhouse et al. 1990; Greens and Scheffler, 1990). Alternative utilization of multiple polyadenylation signals in 3'-end processing of pre-mRNA has also been described (Setzer et al. 1980; Tosi et al. 1981; Aho et al. 1983; Sanke et al. 1988). Modifications at the 3'-ends of messages have been shown to affect the translational capacity (Greens and Scheffler, 1990) and rates of mRNA turnover (Brawerman, 1987; Ross, 1988) in some cases. The alternative splicing of pre-mRNA and its functional effects are reviewed in next section.

D. Alternative Pre-mRNA Splicing

Eukaryotic pre-messenger RNAs undergo a series of nuclear processing events. In addition to 5' capping with 7-methyl guanosine (Salditt-Georgieff et al. 1980), and the addition of a poly A tail to the 3' end of the nascent transcripts (Birnstiel et al. 1985), introns must be precisely excised and
exons correctly ligated to avoid disrupting the open reading frame. Not unexpectedly, the pre-mRNAs can sometimes be spliced in more than one way. In recent years, alternative pre-mRNA splicing has emerged as a powerful post-transcriptional mechanism for regulating gene expression. By utilizing different splice sites within the same precursor RNA, multiple transcripts and corresponding protein products can be generated from a single gene. In this section, alternative splicing and its function will be reviewed after briefly introducing constitutive splicing.

1. Constitutive Splicing

The splicing reaction has been extensively studied in cell-free systems that used exogenous pre-mRNA substrates (Grobowski et al. 1984; Krainer et al. 1984; Ruskin et al. 1984 Lambowitz, 1989). Splicing is a multistep process involving two successive transesterification reactions (Sharp, 1987). Initially, cleavage occurs at the 5' donor splice site, and the 5' end of the intron is joined to an adenosine residue located within the intron (branch-point), giving rise to a lariat form of the intron. The second transesterification reaction between the 5' splice donor and 3' splice acceptor sites results in ligation of the exons and release of the lariat intron. Comparison of sequences of the corresponding genomic DNA with that of cDNA revealed that moderately conserved and short consensus sequences exist at intron-exon boundaries (Mount, 1982). The 5' donor splice
site consensus is AG/GUP GU (where the intron-exon boundary is denoted by the backslash), and the 3' acceptor splice site consensus sequence is P,AG/G. Of these bases, the most highly conserved are the first two and last two bases of the intron.

Although important, these consensus sequences are not sufficient to define efficient splice sites because similar sequences are found in positions other than actual splice sites. Some of these cryptic sites can be activated when the normal splice sites are mutated (Ruskin et al. 1985; Aebi et al. 1987). Several observations had led to a suggestion that small ribonucleoprotein particles (snRNPs) somehow assisted in the splicing reaction. First, the short consensus sequence at the 5' end of introns was found to be complementary to a site near the 5' end of the small nuclear RNA (snRNA), called U1 (Lerner et al. 1980). Second, snRNAs were found in nuclear extracts as part of large particles containing pre-mRNA (Steitz et al. 1988). Subsequently, U1, U2, U4, U5 and U6 have been purified, sequenced and shown to be necessary for splicing (Guthrie, 1991; Ruby and Abelson, 1991). All of these snRNAs are associated in the nucleus with six to ten proteins to form snRNPs. These snRNPs assemble with pre-mRNA and additional proteins to form a large particle, called a spliceosome.

The total population of pre-mRNA in the cell nucleus, known as heterogeneous nuclear RNA (hnRNA), is found in association with a discrete set of proteins, forming repeating
structures termed hnRNP particles. These are visualized as a linear array of globular monoparticles along each pre-mRNA molecule (Dreyfuss, 1986). Cleavage of the linker RNA by brief RNase treatment releases single particles that sediment at 30-40S. These particles contain a segment of hnRNA about 700 nucleotides long and core hnRNP proteins, A1, A2, B1, B2, C1, and C2, in a well-defined stoichiometry (Conway et al. 1988). Because of their abundance and association with hnRNA, hnRNP proteins are thought to be involved in the packing and posttranscriptional processing of all pre-mRNA. In vitro studies under splicing conditions showed sequence-specific binding of hnRNP protein A1 to the 3' splicing site (Swanson and Dreyfuss, 1988). In nuclear extracts, UV-induced cross-linking of hnRNP proteins A1 and C to pre-mRNA requires the presence of U1 and U2 snRNPs, suggesting that specific interaction take place between hnRNP and snRNP particles (Mayrand and Pederson, 1990). Prevention of in vitro splicing reaction by the antibodies against hnRNP proteins indicated a requirement of hnRNP proteins during splicing (Choi et al. 1986; Sierakowska et al. 1986).

The interactions of the cis-acting elements in pre-mRNA with the trans-acting protein and snRNA factors are necessary for constitutive splicing. These interactions are thought to be altered and/or regulated to allow for alternative splicing.

2. Modes of Alternative Splicing

Smith et al (1989) have devised a scheme for classifying
alternative splicing based upon a wide variety of possible regulated events which can give rise to different mRNA transcripts from the same initial gene.

**Retained Intron** Introns can either be spliced out or retained in the mature RNA. The retained intron can result in a premature termination, an altered amino acid segment due to frameshift, or the insertion of a peptide if the open reading frame is maintained. Premature termination or frameshift may cause functionally different products to be made or may produce no functional product. Such examples are found in the genes for fibronectin (Schwarzbauer et al. 1987) and the A-chain of platelet derived growth factor (Tong et al. 1987).

**Alternative 5' Donor Sites and 3' Acceptor Sites** Alternative utilization of 5' donor and/or 3' acceptor sites can excise introns of different lengths, resulting in retained exons of different sizes. Again, this can give rise to the insertion or deletion of a small peptide if the reading frame is maintained, or frameshifting and premature termination.

**Alternative Promoters and Polyadenylation Sites** In some cases, a gene may contain more than one promoter or more than one poly A attachment site. The mRNAs which differ in their 5' and 3' ends can result from differential utilization of more than one promoter or more than one polyadenylation signal. This can lead to variations in the amino acid sequence of the amino- or carboxyl-terminals of a protein, or to variations in the 5' and 3' untranslated regions of the
message. Utilization of alternative promoters dictates the splicing pattern of myosin light chain gene (Periasamy et al. 1984), while an example of alternative 3' end exons resulting from the selection of different polyadenylation sites is found in α-tropomyosin transcripts (Helfman et al. 1986).

Mutually Exclusive Exons Internal exons are sometimes used in a mutually exclusive manner. One member of the pair is always spliced into the mRNA, but never both. Thus, the two exons are never spliced together, nor are they both skipped. Examples of such splicing are found in genes for α and β tropomyosins (Helfman et al. 1986; Goodwin et al. 1991). Isoforms TM-5a and TM-5b of α-tropomyosin are due to alternative inclusion of exon 6a and 6b respectively. The differences in the amino-acid sequence encoded by these alternate exons affects the actin-binding pattern of the tropomyosin.

Cassette Exons Some cassette exons can be included or excluded (skipped). This can lead to insertion of small peptide segment, premature termination, or a shift of reading frame. An example of such cassette exons is found in the troponin-T gene (Breitbart and Nadal-Ginard, 1987). The presence of five cassette exons at 5' end of the gene give rise to hypervariability within the N-terminal region of troponin-T.

3. Factors Affecting Splicing-Site Selection

Splicing of all intron-containing pre-mRNAs must occur
precisely in order to generate functional proteins. Authentic splicing sites, which exhibit sequence conservation, are accurately selected and matched by poorly understood mechanisms. On the other hand, flexibility in the selection of splice sites forms the basis for alternative splicing, an important mechanism for posttranscriptional control of gene expression. Thus, a large number of genes can express several related, but structurally and functionally distinct protein isoforms. The corresponding mRNAs arise by selection of alternative splice sites, a process that is often subject to tissue-specific or developmental control. The molecular mechanisms responsible for the specificity of constitutive and alternative splice-site selection are not well understood. Some cis- and trans-acting factors have been reported to be involved in splice-site selection, some essential for splicing, others affecting the efficiency with which a splice site is utilized. RNA secondary structure is also shown to affect splice-site selection.

**Cis-Acting Factors** The 5' splice site acting as a cis-splicing element is highly conserved (Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987). Mutations within the 5' GU dinucleotide at the intron boundary inhibit in vitro splicing (Aebi et al. 1987; Treiman et al. 1983), and often cause activation of nearby cryptic splice sites in vivo. Variations of the 5' splice-site consensus outside of the GU do not absolutely impair splicing, but can affect the efficiency with
which the splice-site is used. Usually, mutation of 5' splice sites towards the consensus improves their competitive efficiency (Aebi et al. 1986; Eperon et al. 1986; Zhuang et al. 1987; Mayeda and Ohshima, 1988).

The 3' splice-site consensus is another cis-acting factor which can affect the splicing. The AG dinucleotide at the intron-exon boundary is absolutely conserved (Mount, 1982; Padgett et al. 1986; Ohshima and Gotoh, 1987). The point mutation or deletion of the AG dinucleotide always inhibits splicing to the mutated splice site (Aebi et al. 1986; Lamond et al. 1987). However, in some cases, this mutation could result in utilization of the next downstream AG dinucleotide as the 3' splice site.

In addition to 5' and 3' splice site consensus, branch-point sequence and location have also been shown to affect the splicing. The branch-point consensus is commonly found 20-40 nt upstream of the 3' splice site (Smith and Nadal-Ginard, 1989; Helfman and Ricci, 1989). Deletion of the sequence between the branch-point and 3' splice site could result in alternative splice-site selection (Helfman et al. 1990; Guo et al. 1991). The most highly conserved position in the consensus is the A at the site of branch formation. Point mutations of the branch-point sequence often have no effect upon splicing due to the activation of a nearby cryptic branch point (Padgett et al. 1985; Reed and Maniatis, 1985; Ruskin et al. 1985). It has been found that some alternatively spliced
introns contained multiple branch points associated with a single acceptor site (Gattoni et al. 1988; Noble et al. 1988; Helfman and Ricci, 1989) demonstrated that the utilization of different branch points plays a crucial role in the alternative splicing of the SV40 pre-mRNA. The large-T and small-t mRNAs result from the use of two alternative 5' splice sites joined to a single shared 3' splice site.

**Trans-Acting Factors** The observation that identical primary transcripts are processed along alternative splicing pathways, according to cell type or stage of development, clearly demonstrates differences in the trans-splicing environments of these cells (Medford et al. 1984; Bretibart et al. 1985; Leff et al. 1987; Breitbart and Nadal-Ginard, 1987; Tsurushita et al. 1988). Such variability in the splicing environments is likely determined by specific splicing factors. Some RNA and protein factors which are required for constitutive splicing are present in variant forms. These variant forms may increase the possibility for regulation of alternative splice-site selection.

snRNPs, U1, U2, U5, and U4/U6, have all been shown to be essential for splicing by a combination of antibody and RNase H mediated degradation experiments (Krainer et al. 1984; Kramer et al. 1984; Black et al. 1985; Chabot et al. 1985; Black and Steitz, 1986; Berget and Robberson, 1986). Some snRNAs showed heterogeneity. For example, about 15% of the total human U1 RNAs may be composed of variant sequences
In addition, in mouse, multiple U1 RNAs have been detected and appear to be developmentally regulated (Kato and Harada, 1985; Lund et al. 1985). Likewise, U4 RNAs are regulated in a developmental and tissue-specific manner in chicken and Xenopus (Lund and Dahlberg, 1987; Krof et al. 1988). Sequence heterogeneity of U2 RNA has not been detected but it is thought that posttranscriptional base modification may be developmentally regulated and lead to heterogeneous U2 snRNPs (Lund and Dahlberg, 1987). Thus, variant forms of snRNPs which are developmentally regulated and/or tissue specific may facilitate differential splice-site selection.

In addition to the above snRNPs, several other protein factors have to be shown to affect alternative splicing. One cellular factor isolated from mammalian cells termed ASF (alternative splicing factor) or SF2 has been found to affect the choice of alternative 5'-splice sites (Ge and Manley, 1990; Krainer et al. 1990a). In vitro splicing studies with the early region of simian virus (SV40) involving the large T- and small t-splice choices, and a model pre-mRNA substrate derived from the human β-globin gene, demonstrated that high concentrations of ASF/SF2 promote the use of proximal 5'-splice sites, whereas low concentrations favor the use of distal 5'-splice sites. Recently, hnRNP A1 has been reported to specifically counteract the effect of ASF/SF2 on the splice-site selection (Mayeda and Krainer, 1992). The hnRNP A1 is not a general inhibitor of ASF/SF2, since both hnRNP A1
and ASF/SF2 are essential splicing factors (Krainer et al. 1990b). However, when two alternative 5' splice sites are available for splicing, the ratio of hnRNP A1 to ASF/SF2 precisely determines which splice site is used in vitro. High concentrations of ASF/SF2 activates proximal 5' splice sites, whereas high concentrations of hnRNP A1 favor distal 5' splice sites with several model and natural alternatively spliced pre-mRNAs. The regulation of these antagonistic activities may play an important role in the tissue-specific and developmental control of gene expression by alternative splicing of pre-mRNA.

RNA Secondary Structure Several authors have suggested that one way in which sequences other than the primary consensus sequences might influence splicing patterns is to alter the secondary structure within the precursor RNA and thereby alter the utilization of a splice site (Solnick, 1985; Eperon et al. 1988; d'Orval et al. 1991; Libri et al. 1991; Estes et al. 1992). The potential of secondary structure to affect splice-site selection was first demonstrated with artificially constructed transcripts. When an exon was put into the loop of a hairpin by inserting long inverted repeats into the adjoining introns, in vitro splicing apparatus generated a novel, "jump"-spliced RNA that had the flanking exons joined directly together, leaving out the exon in the loop. To a lesser extent, the hairpin caused the same alternative splice in vivo, in transfected HeLa cells
(Solnick, 1985). The extent by which the exon within a stem-loop structure was "skipped" was inversely proportional to the stability of the stem (Solnick and Lee, 1987; Eperon et al. 1988). Recently, secondary structure in the chicken β-tropomyosin pre-mRNA has been shown to play a role in the regulation of the alternative splicing of this message (d'Orval et al. 1991; Libri et al. 1991; Guo et al. 1991). The chicken β-tropomyosin pre-mRNA is spliced in a tissue-specific manner to yield mRNAs coding for different isoforms of this protein. Exons 6A and 6B are spliced in a mutually exclusive manner; exon 6B was included in skeletal muscle, whereas exon 6A was preferred in all other tissues. The distal portion of the intron upstream of exon 6B was shown to form stable double-stranded regions with part of the intron downstream of exon 6B. This structure repressed splicing of exon 6B to 7 in a Hela cell extract. Derepression of splicing occurred on disruption of this structure and repression followed when the structure was reformed. Thus, cellular factors which either stabilize or melt secondary structure may regulate alternative splicing.

E. Functional Effects of Alternative Splicing of Pre-mRNA

The functional effects of alternative splicing of pre-mRNA is a subject of considerable interest. Strong correlations have been made between alternative splicing and biological function. These cases highlight the importance of
alternative splicing as a posttranscriptional regulatory event which can quantitatively and qualitatively control gene expression. Alternative splicing in most cases gives rise to protein isoforms sharing some identical regions, and varying in some specific domains, thus allowing for the modulation of protein function, such as regulation of protein localization, modification of protein activity, and production of novel protein activities. Alternative splicing is also used to quantitatively regulate gene expression, by giving rise to prematurely truncated open reading frames, or by regulating mRNA stability or translational efficiency via variability in the untranslated regions. As a mechanism to regulate gene expression and generate protein diversity, alternative splicing has advantages over gene rearrangement and extensive multigene families. It allows for switches in protein isoforms without permanent changes in the cell genetic content, and without changes in transcriptional activity. Functional effects of alternative splicing are reviewed below.

1. Protein Localization

Regulated alternative RNA processing is often used to specify the localization of proteins. There are a number of cases of proteins which are either present within the cell or are secreted. For example, the actin filament-severing protein gelsolin exists as both a plasma protein and a cytoplasmic protein. The two mature proteins are identical except for an additional 25 amino acid residues at the its
amino terminus of the plasma isoform. The two isoforms are expressed from the same gene by the differential use of alternative promoters and alternative splicing of 5' exons that code for the extra 25 amino acids of the mature plasma protein and a 27 amino acid cleavable signal peptide that directs its initial translocation from the cytosol into the rough endoplasmic reticulum (Kwiatkowski et al. 1986). No functional differences have been detected between the intracellular and extracellular forms of gelsolin.

Avian epidermal growth factor receptor is another example of a protein that utilizes alternative splicing to generate membrane-bound and secreted forms (Flickinger et al. 1992). Epidermal growth factor receptor is a 170 KDa transmembrane glycoprotein which is encoded by 12 Kb or 8.6 Kb transcripts. Recently, a small transcript (2.6 Kb) was isolated. This transcript encodes a soluble, secreted and truncated receptor molecule that can act as an antagonist to inhibit full-length receptor function. Sequence analysis of corresponding cDNA and genomic DNA revealed that the 2.6 Kb epidermal growth factor receptor transcript is produced by alternative splicing. The 3' end of this transcript diverged from the sequence of the extracellular ligand-binding domain of the full length receptor are a result of the utilization of an alternative polyadenylation signal between exons 12 and 13 of the full length receptor transcript.

The gene for immunoglobulin μ (Igμ) heavy chain also
gives rise to membrane-bound and secreted forms by alternative splicing. The Igµ heavy chain is present in early B lymphocytes as a membrane-bound form. Upon B-cell maturation, following antigen stimulation, levels of the membrane-bound form progressively decrease while the production of a secreted pentameric form increases. This switch is achieved by the alternative use of 3' end exons; the early B-lymphocyte form codes for the hydrophobic membrane-binding segment (Alt et al. 1980; Early et al. 1980; Rogers et al. 1980). These examples show that alternative splicing could regulate the distributions of a protein between subcellular compartments. The targeting of protein to different locations involves the use of leader sequences that are removed from the mature protein during or after transportation, or by the use of membrane targeting sequences which are modified by attachment of membrane anchoring substituents. The distribution of proteins to more than one cell compartment could thus be regulated by alternative splicing of such targeting sequences.

2. Modulation of Protein Function

Functionally distinct protein isoforms are often produced by alternative splicing. Generation of the protein isoforms with differing functions due to alternative splicing has been found for some transcription factors (Lillycrop and Latchman, 1992; Morris et al. 1992), receptors (Mosley et al. 1989; Baumbach et al. 1989; Takaki et al. 1990; Johnson et al. 1991; Flickinger et al. 1992), ion channels (Schwarz et al. 1988;

The RNA encoding the octamer-binding transcriptional factor Oct-2 is alternatively spliced to yield multiple mRNAs encoding different isoforms of the protein. Isoforms Oct 2.4 and 2.5 differ from the other forms, Oct 2.1, 2.2, and 2.3, at their carboxyl termini. When introduced into cells lacking endogenous Oct-2, various Oct-2 isoforms have different effects on octamer-containing promoters (Lillycrop and Latchman, 1992). The Oct 2.1, 2.2, and 2.3 forms stimulate all octamer-containing promoters. However, the Oct 2.4 and 2.5 forms can repress some promoters and stimulate others, depending on the sequence of the octamer motif and its context within the promoter. Thus alternative splicing can generate different isoforms of the Oct-2 proteins with different effects on gene expression. Alternative splicing has also been observed to affect the transcription factor Pit-1 in a similar manner (Morris et al. 1992).

Changes in receptor functions can be correlated with alternative splicing of epidermal growth factor receptor (Flickinger et al. 1992), fibroblast growth factor receptor (Johnson et al. 1990; Johnson et al. 1991), interleukin receptors (Mosley et al. 1989; Takaki et al. 1990) and growth hormone receptor pre-mRNAs (Baumbach et al. 1989). In these cases, alternatively spliced mRNAs encode the soluble, secreted or truncated receptors which often lack transmembrane
domains or cytoplasmic domains, but contain extracellular ligand binding domains and retain ligand binding abilities. One can speculate that the secreted receptors can act as reservoirs to regulate levels of extracellular ligands, or inhibitors to block ligand-mediated responses by preventing interaction of the ligands with membrane-bound forms of receptors. For instance, alternatively processed mRNA from the epidermal growth factor receptor gene encodes a soluble truncated form of the receptor that can inhibit the function of the membrane-bound receptor by blocking ligand-dependent cell transformation (Flickinger et al. 1992).

Alternative splicing can also generate enzyme isoforms with differing enzymatic activities. Different glucokinase isoforms are produced by tissue-specific alternative RNA splicing in liver and pancreatic islet cells (Liang et al. 1991). L1 is major glucokinase isoform found in liver, and B1 in islets. L1 and B1 differ from each other by 15 amino acids at the NH₂ terminus due to occasional skipping of exon 1. The Km values of the B1 and L1 isoforms are similar, but the Vₘₐₓ of the B1 isoform is 2.8 fold higher than that of the L1 isoform. In addition, two minor glucokinase isoforms in liver and islet cells, L2 and B2, generated from mRNAs lacking 51 nucleotides due to the use of an alternate splice acceptor site in the exon 4, lack glucose phosphorylating activity. These results suggest that the alternative splicing of the glucokinase pre-mRNA dramatically affect the enzyme activities
of the gene products.

In a limited number of cases alternative RNA processing gives rise to mature products with totally different functions. Here the alternatively spliced region encodes the major active site of the product. This type of alternative splicing is used in the generation of a number of peptide hormones. The calcitonin gene produces both calcitonin and calcitonin gene-related peptide (CGRP) by alternative splicing and poly(A) addition site usage (Amara et al. 1982; Jonas et al. 1985). Exons 1-3 contain the 5' untranslated and common coding regions. Exon 4 contains the calcitonin coding sequence and a poly(A) addition site and is used in thyroid. In brain and other neuronal tissue exon 3 is spliced to exons 5 and 6, which contain the CGRP coding region and its 3' untranslated sequences and poly(A) addition site. Translation of these two different mRNAs produces two prehormones; proteolytic cleavages release the two different polypeptides, calcitonin and CGRP. Calcitonin functions in calcium retention while CGRP may have neuromodulatory activity (Fontaine et al. 1987).

3. Effects on RNA Stability and Translatability

Many genes produce transcripts with heterogeneity within the 5' and 3'-UTRs. The untranslated sequences are known to be involved in regulating translational efficiency (Kozak, 1991a,b) and RNA stability (Brawerman, 1987). Thus, alternative processing involving untranslated regions provides
the potential for regulated expression of the same gene product in different cells, or for regulating the expression of different products arising from alternative splicing of coding regions. A specific sequence promoting mRNA decay has been identified in the 3' noncoding region. Shaw and Kamen (1986) have shown that the colony-stimulating factor-1 (CSF-1) gene produced two transcripts that were identical in the coding region but differed only in the 3'-UTR. One 3'-UTR contains a repetitive AUUUA element found in a number of rapidly turning over mRNAs, and causes a dramatic destabilizing effect when inserted into the 3'-UTR of β-globin mRNA. This suggests that CSF-1 expression can be regulated at the level of mRNA turnover according to the selection of different 3' untranslated sequences.

The 5'-UTR has been shown to be involved in the regulation of translational efficiency. Many reports indicated that changes in the 5'-noncoding sequence of messages can make translation more or less difficult. The suspicion that complicated leader sequence can impair translation for mRNAs containing alternative 5'-UTRs has been confirmed for some gene transcripts, such as human complement protein C2 (Horiuchi et al. 1990), murine tissue inhibitor of metalloproteinases (Waterhouse et al. 1990) and rat insulin-like growth factor I (Lowe et al. 1987; Foyt et al. 1991). Features of the 5'UTRs which can influence translational capacity of the message include: 1) length; 2) the presence of
additional upstream open reading frames; and 3) the presence of secondary structure.
CHAPTER III

METHODS

A. Materials

Enzymes and chemicals were purchased from BRL (Gaithersburg, MD), Promega (Madison, WI), United States Biochemicals (Cleveland, OH) and Sigma (St. Louis, MO). $[^{32}\text{P}]dCTP$ (specific activity 3,000 Ci/m mole), $[^{32}\text{P}]UTP$ (specific activity 400 Ci/m mole) and $[^{35}\text{S}]L$-methionine (specific activity 1,000 Ci/m mole) were purchased from ICN Biochemicals (Costa Mesa, CA). Nytran filters were from Schleicher & Schull (Keene, NH). $N^\text{a}$-Benzoyloxycarbonyl-L-arginyl-L-arginine 7-amido-4-methylcoumarin (Z-Arg-Arg-AMC) was from BACHEM Bioscience Inc. (Philadelphia, PA).

Plasmids phCB79-2 (Chan et al. 1986) and phCB96-2 were gifts from Drs. S.J. Chan and D.F. Steiner (University of Chicago, Chicago, IL). The plasmid phCB79-2 (Figure 2) harbors the cDNA for a 2.3 Kb form of human cathepsin B mRNA containing 237 nt 5'-untranslated region (UTR), the coding region for preprocathepsin B, and a 798 nt 3'-UTR, cloned into the $\text{EcoRI}$ site of plasmid pGEM-2 (Promega). Adjacent to the cathepsin B cDNA is an SP6 promoter to drive mRNA synthesis.
Figure 2: Human cathepsin B cDNAs in plasmids phCB79-2 and phCB96-2. The 5' and 3' UTRs, and coding region for the prepeptide, activation propeptide and mature enzyme in the cathepsin B cDNAs are indicated. Also shown are restriction sites. Panel A shows plasmid phCB79-2 and its cathepsin B cDNA insert. Panel B shows plasmid phCB96-2 and its cathepsin B cDNA insert. The cDNA insert in phCB96-2 is missing an 88 nt segment (exon 2) present in phCB79-2.
and a T7 promoter to drive antisense or cRNA synthesis. A single internal EcoRI site interrupts the codon for glycine 328 near the carboxyl-terminus of cathepsin B. Plasmid phCB96-2 (Figure 2) contains a human cathepsin B cDNA which is missing an internal 88 nt segment (exon 2) in the 5'-UTR of phCB79-2, and is bounded at its 3'-end by the same internal EcoRI site of phCB79-2 which interrupts the glycine 328 codon. This cDNA fragment is cloned into the EcoRI site of the plasmid pGEM-4Z. Positioned at each end of the insert are SP6 and T7 promoters to drive mRNA and cRNA synthesis respectively.

**B. Solutions**

**LB Broth:** 10 g peptone (Gibco BRL), 5 g yeast extract (Gibco BRL), 6 g NaCl per liter

**SSPE:** 10 mM NaH₂PO₄, 0.15 M NaCl, 1.0 mM EDTA, pH 7.4

**TBE:** 89 mM Tris-borate, 89 mM boric acid, 2 mM EDTA

**SSC:** 15 mM sodium citrate, 0.15 M NaCl, pH 7.0

**TE:** 10 mM Tris-HCl, 1 mM EDTA, pH 7.6

**DEPC-Treated H₂O:** water was treated with 0.1% diethylpyrocarbobate for at least 12 hours and autoclaved.

**P₁ Buffer:** 100 µg/ml RNase A (Sigma), 50 mM Tris-HCl, 10 mM EDTA, pH 8.0

**P₂ Buffer:** 200 mM NaOH, 1% SDS

**P₃ Buffer:** 2.55 M KAc, pH 4.8

**QBT Buffer:** 750 mM NaCl, 50 mM MOPS, 15% ethanol, 0.15%
Triton X-100, pH 7.0

**QC Buffer:** 1.0 M NaCl, 50 mM MOPS, 15% ethanol, pH 7.0  
**QF Buffer:** 1.25 M NaCl, 50 mM MOPS, 15% ethanol, pH 8.2

**C. Cell Culture**

Tumor cells used in these studies included a cultured human colon carcinoma (from L.C. Erickson, Loyola University of Chicago, Stritch School of Medicine), human melanomas A375P and A375M (from I.J. Fidler, Anderson Hosp. and Tumor Inst.), and human prostate carcinomas PC3P and PC3M (from J. Kozlowski, Dept. of Urology, Northwestern Med. Sch.). The variant A375M and PC3M were derived from the parental lines A375 (A375P) and PC3 (PC3P) respectively by culturing cells derived from experimentally produced metastatic lung tumor nodules. The variants A375M and PC3M were more metastatic than the parent cells after i.v. injection (experimental metastasis) and after s.c. transplantation (spontaneous metastasis) into young nude mice (Kozlowski et al. 1984). The melanomas A375P and A375M were cultured in Minimum Essential Medium supplemented with Earle's salts, essential amino acids, vitamins, sodium pyruvate (110 mg/L), penicillin (100 units/ml), streptomycin (100 µg/ml), sodium bicarbonate (2.2 g/L), and fetal bovine serum (10%). The prostate carcinomas PC3P and PC3M, and colon carcinoma were cultured in RPMI 1640 supplemented with penicillin (100 units/ml), streptomycin (100 µg/ml) and fetal bovine serum (10%). 70-80% confluent cells
were subcultured by scraping into 10 ml of serum free media, pelleting at 500xg and resuspending in fresh media containing 10% fetal bovine serum. The media were changed three times weekly.

D. Human Tissues

Normal human kidney, spleen, liver, colon and breast tissues, and samples of several solid tumors were provided by the Cooperative Human Tissue Network and the National Disease Research Interchange. The tissues were reported to have been snap frozen at the time of collection and stored at -70°C. Pathological reports were provided for each tumor specimen.

E. Total RNA Isolation

Total RNAs were prepared from normal human tissues and solid tumors, and from cultured cells with a method described by Chomczynski and Sacchi (1987). Briefly, the cell pellet or tissue was crushed in liquid nitrogen and homogenized in denaturing solution containing 4M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarcosyl, and 0.1 M 2-mercaptoethanol. Sequentially, 0.1 volume of 2.0 M sodium acetate, pH 4.0, 1 volume of phenol, and 0.2 volume of chloroform-isoamyl alcohol mixture (49:1) were added to the homogenate. The aqueous phase was collected by centrifugation and the RNA in the aqueous phase was precipitated with an equal volume of isopropanol and washed with 70% ethanol. The
RNA was redissolved in diethyl pyrocarbonate (DEPC)-treated water and its integrity was examined by electrophoresis in an agarose gel (Fourney et al. 1988).

F. Quantification of Cathepsin B mRNA

The RNA dot blot procedure was used to quantify cathepsin B mRNA. Total RNA from normal human tissues and tumor cells were dissolved in a denaturing solution (50% formamide and 10x SSC) and heated at 65°C for 15 min. Varying amounts (0.1 µg to 10 µg) of denatured RNA were spotted onto Nytran filters. For a cDNA probe phCB79-2 was digested with EcoRI, and was fractionated by electrophoresis in 0.7% agarose gel. The DNA fragment containing human cathepsin B sequence was eluted from the gel slice by centrifugation through a 1.5 ml polypropylene tube which had a hole at bottom and a small amount of glass wool as a cushion. The collected DNA was then extracted with 1 volume of phenol/chloroform and precipitated by ethanol. The DNA pellet was redissolved in the TE buffer. The approximate concentration of the purified DNA fragment was determined by agarose gel electrophoresis and comparison to known amounts of DNA standards. The purified fragment was labeled with $^{32}$P-dCTP by the random primer labeling method according the Protocols and Applications Guide provided by Promega. Twenty ng of the DNA fragment in 30 µl of water was heated at 100°C for 3 min, and chilled on ice for 5 min. After addition of 10 µl of 5X labeling buffer, 2 µl of a
mixture of the unlabeled dNTPs (500 mM each of dATP, dTTP, 
dGTP), 2µl of acetylated BSA (10 mg/ml), 5 µl of [α-32P]dCTP 
(3,000 Ci/m mole, 10 mCi/ml) and 5 units of Klenow enzyme 
(Promega), the mixture was incubated at room temperature for 
1 hour. The reaction was then terminated by heating at 100°C 
for 3 min and chilling on ice for 5 min. After addition of 2 
µl of 0.5 M disodium ethylenediamine tetraacetate (EDTA) and 
48 µl of water, the labeled DNA fragment was separated from 
unincorporated [α-32P]dCTP by a Shephadex G-50 spun-column 
chromatography (Maniatis et al. 1987). The purified labeled 
probe was used in hybridization.

The Nytran filter was prehybridized for 4 hours in a 
solution containing 50% formamide, 5X Denhardt's solution 
(Maniatis et al. 1987), 0.25 mg/ml of denatured salmon sperm 
DNA, 1.0% SDS, 1.0 M NaCl, 10 mM Na2HPO4 and 0.1% tetrasodium 
pyrophosphate. Hybridization was done at 42°C overnight in a 
fresh mixture of the same solution to which 32P-labeled human 
cathepsin B cDNA was added. After hybridization, the filter 
was washed three times (30 min each time) in 0.2x SSC and 0.5% 
SDS at 65°C, and exposed to Kodak XAR film. To rehybridize 
the blot, the cathepsin B probe was removed by washing the 
filter three times in 0.01X SSPE and 0.5% SDS at 95°C for 30 
min. For the internal standard control, the stripped filter 
was similarly hybridized with 32P-labeled cDNA corresponding to 
the chicken 18s rRNA.
G. Northern Blot Analysis

Total RNAs were denatured with formamide and formaldehyde and subjected to electrophoresis in a 1.1% agarose/0.02% formaldehyde gel in 3-[N-morpholino] propanesulfonic acid (MOPS) buffer (0.02 M MOPS, 5.0 mM sodium acetate, 1.0 mM EDTA, pH 7.0) (Fournier et al. 1988). The RNAs were transferred to Nytran filter, and the filter was hybridized as described above.

H. RNase Protection Assay

Protection assays were performed using a modification of a described method (Lowe et al. 1987). The RNase protection assay, outlined in Figure 3, depends on the production of a complementary RNA probe of high specific radioactivity which hybridizes to homologous sequences in mRNA. The resulting regions of double strand RNA are resistant to digestion by RNase T1 and RNase A. Single stranded regions in the labeled RNA probe which do not complement the message are degraded. Labeled RNA fragments of characteristic size are produced which correspond to specific mRNA species present in the sample. These labeled fragments can be analyzed for size by electrophoresis in a denaturing gel.

For construction of the riboprobe vector, a 445-bp EcoRI/HindIII cDNA insert, containing sequence corresponding to the entire 5'-untranslated region of the human cathepsin B mRNA (Figure 4), was isolated from phCB79-2, and ligated into
Figure 3. Scheme for RNase protection assay. A DNA fragment containing 5'-end of human cathepsin B cDNA was cloned into pGEM-2. A radioactive antisense RNA probe was synthesized using T7 RNA polymerase and [α-32P]UTP as described in methods. The resulting antisense RNA was purified and then incubated with total RNA for annealing. RNase A and RNase T1 was added to digest single strand RNA. The protected double stranded fragments were analyzed by electrophoresis on a denaturing gel.
Figure 4: The riboprobe for RNase protection assay. The organization of a human cathepsin B mRNA, and the sizes of first six exons are shown. A 445 nt antisense RNA (riboprobe), complementary to the sequence from exon 1 to the middle of exon 5 of the cathepsin B mRNA, was synthesized for the RNase protection assay outlined in Figure 3.
a pGEM-2 vector (Promega) previously cut with EcoRI/HindIII. After linearizing the plasmid with EcoRI, the human cathepsin B riboprobe was synthesized in 20 µl reaction mixture containing 1.5 µg DNA template, 40 mM Tris-HCl (pH 7.5), 6 mM MgCl₂, 10 mM NaCl, 2 mM spermidine, 10 mM dithiothreitol, 40 units RNasin, 50 mM each of CTP, ATP and GTP, 20 µM UTP, 5 µM ³²P-UTP (50 µCi, 400 Ci/mmole), and 20 units T7 RNA polymerase according the Protocols and Application Guide provided by Promega. The mixture was then incubated at room temperature for 2 hours. After transcription, 1 µg of DNase I (BRL) and 1 µl of 200 mM vanadyl-ribonucleoside complex (BRL) were added, and the mixture was incubated for 15 min at 37°C. The mixture was then extracted with 0.5 vol of phenol and 1 vol of chloroform and subjected to purification by a polyacrylamide gel electrophoresis. After running the 8% polyacrylamide/8 M urea denaturing gel, the 445-base riboprobe was located by autoradiography and eluted from a gel slice in 400 µl buffer (2 M ammonium acetate, 1% SDS and 25 µg/ml tRNA). The labeled riboprobe was precipitated with 2 volume of 100% ethanol, dried in vacuum and resuspended in DEPC treated water.

The RNAs used in the assays were precipitated from 0.5 M NaCl with 2 volume of ethanol, recovered by centrifugation, and dried in vacuum. 20 µg of total RNA was resuspended in 29.5 µl of hybridization buffer (20 mM Tris-HCl, pH 7.6, 1 mM EDTA, 0.4 M NaCl, 0.1% SDS, and 75% deionized formamide) to which 0.5 µl of H₂O containing 500,000-700,000 cpm of the
labeled riboprobe was added. The mixture was heated to 85°C for 5 min and incubated at 45°C for about 16 hr. After hybridization, 270 µl of RNase digestion buffer (10 mM Tris-HCl, pH 7.6, 5 mM EDTA, 300 mM NaCl, 40 µg of RNase A per ml (Sigma), and 2 µg of RNase T1 per ml (BRL)) was added, and the mixture was incubated for 1 hr at 30°C. SDS was then added to a final concentration of 0.6% followed by addition 50 µg of proteinase K. The mixture was incubated at 37°C for 15 min, and then extracted with 0.5 volume of phenol and 1 volume of chloroform. Twenty µg of tRNA and 2 volume of ethanol were added to the aqueous layer. After 30 min at -20°C, the precipitate was collected by centrifugation, dried, and resuspended in formamide sample buffer (80% deionized formamide and 0.2% of bromophenol blue). The sample was heated to 95°C for 3 min, chilled on ice, and applied to an 8% polyacrylamide/8 M urea denaturing gel. After electrophoresis at 50°C in 0.5x TBE, the gel was dried and exposed to Kodak XAR film at -70°C overnight. RNA bands, corresponding to the alternative 5'-UTRs were quantitated by densitometry, and their sizes estimated by comparing their mobilities with those of DNA markers.

I. Polymerase Chain Reaction (PCR) Amplification of cDNA

The first strand cDNA was synthesized with the SuperScript™ RNase H Reverse Transcriptase (BRL) from total RNA using random primers according the method described by
Kawasaki (1991). The reaction mixture contained 1 µg total RNA, 100 pmole of random primer mixture (Pharmacia, Alameda, CA), 40 units RNasin, 1.0 mM each of dCTP, dATP, dGTP and dTTP, and 200 units of SuperScript™ RNase H" Reverse Transcriptase in a final volume 20 µl. Then the mixture was incubated at room temperature for 10 min and at 42°C for 1 hour. Four µl of this reaction mixture was used directly as template for PCR. To amplify the 5' region of cathepsin B cDNA, PCR was performed with a pair of specific oligonucleotide primers having the *HindIII* and *EcoRI* restriction sites at their termini to facilitate subcloning; the primers were 5'GGGGATCCACCTCCGCGCGC (sense) in exon 1, and 5'GGGAATTCCGGATGGATCGACGACCC (antisense) in exon 5 (Figure 5). The resulting fragment was subcloned into the *HindIII/EcoRI* site of the pBluescript vector (Stratagene, La Jolla, CA) to give phCBel.4. For amplifying the cathepsin B cDNA missing exons 2 and 3, the sense primer, 5'CAGCGCTGGGCCGGAC, was complementary to 9 nucleotides on each side of the exon 1-exon 4 junction, and the antisense primer was 5'CAGCGCTGGGCCGGAC and complementary to a sequence present in the flanking 3' untranslated region of all known, full length cathepsin B mRNAs (Figure 6). The reaction mixture contained 4 µl of the template solution (see above), 2.5 units of Taq polymerase (United States Biochemicals), 400 nM of each primer, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.1 mg/ml gelatin, 0.01% Tween 20 and 0.01% NP-40 in a
Figure 5. Scheme for PCR amplification of 5'-end of cathepsin B cDNA. Alternatively spliced forms of human cathepsin B mRNA are shown. The 5' (sense) and 3' (antisense) primers used in PCR were complementary to the indicated regions in exons 1 and 5 respectively. PCR products of 549 bp, 461 bp and 306 bp are expected after amplification of the cDNAs containing the full length 5' UTR, those missing exon 2, and those missing exons 2 and 3, respectively.
Figure 6. Scheme for the PCR amplification of cathepsin B cDNA missing exons 2 and 3. Three cathepsin B transcripts which result from alternative splicing in their 5' regions are shown. The cDNA for the transcript lacking exon 2 and 3 was amplified with a sense primer which is complementary to the sequence of the exon 1-exon 4 junction, and an antisense primer which is complementary to the sequence in the exon 11 (3'-untranslated region). A 934 bp PCR product is expected.
final volume 100 µl. Cycling was performed in a Eppendorf DNA thermal cycler with a profile of 30 sec at 95°C, 1 min at 65°C, and 2 min at 72°C for 30 cycles. Ten µl of the reaction product were analyzed by agarose gel electrophoresis, transferred onto a Nytran filter, and the filter hybridized with $^{32}$P-labeled human cathepsin B cDNA.

**J. Sequencing of PCR Fragments**

PCR reaction products were isolated, digested with HindIII and EcoRI and subcloned into the pBluescript vector for single strand sequencing by the dideoxy chain termination method (Sanger et al. 1977).

The single strand DNA was prepared with the method described by Russel et al (1986). The HB101[$F^+$] cells containing phCBel.4 plasmid were patched onto LB plates containing 100 µg ampicillin/ml and incubated overnight at 37°C. A portion of each patch was used to inoculate 1.5 ml modified LB broth (16 g tryptone, 10 g yeast extract, 5 g NaCl per liter) plus glucose (0.1%) and 100 µg ampicillin/ml in a culture tube, giving visible turbidity to the solution. Next $5 \times 10^9$-$10^{10}$ of helper phage R408 (Promega) were added. The cultures were incubated at 37°C with vigorous agitation for 4-5 h. After centrifuging the infected culture, the 1.3 ml of the supernatant was transferred to a tube containing 0.2 ml of 2.5 M NaCl and 20% polyethyleneglyco-6000 (PEG-6000) and incubated at room temperature for 15 min. The mixture was
then centrifuged in a microfuge for 5 min at room temperature. The supernatant was removed by aspiration, and the tube was briefly recentrifuged to remove the remaining trace of liquid. The pellet was resuspended in 100 µl of 10 mM Tris-HCl, pH 8.0, containing 2mM EDTA, and extracted with 50 µl of phenol. The aqueous phase was collected by centrifugation, and the single strand DNA in the aqueous phase was precipitated with 250 µl of a 25:1 mixture of ethanol and 3 M sodium acetate, washed with absolute ethanol, dried at 65°C for 10 min and dissolved in 20 µl of TE solution. The single strand DNA was then sequenced with Sequenase Version 2.0 (United States Biochemicals) according the manufacturer's instruction.

K. Construction of Expression Vectors

SP6/cathepsin B mRNA expression vectors were made by inserting cathepsin B cDNA into pGEM2 (Promega) vector adjacent to the SP6 promoter, using standard cloning procedures. The insert and vector DNAs were prepared by restriction enzyme digestion (Maniatis et al, 1987). Each digestion mixture contained 5 µg of plasmid DNA, 5 µl of appropriate restriction 10x buffer, 10-25 units of appropriate restriction enzyme in final volume of 50 µl. The mixture was incubated at 37°C for 1-3 h. The reaction end point was verified by electrophoresis of an aliquot of the sample on an agarose gel. After restriction enzyme digestion, the vectors and inserts were separated by agarose gel electrophoresis, and
the DNA fragments were eluted from the agarose gel by centrifugation (see above). The vector and insert were then ligated overnight at 14°C with 1 unit of T4 ligase (BRL) in the assay buffer provided by the manufacturer. The ligated DNA was transformed in *E.coli* strain HB101. Ampicillin resistant clones were selected, plasmid was prepared, and samples were analyzed by restriction digestion.

The construction of phCB79-2-2 is outlined in Figure 7. It contains the shorter 5′-UTR lacking exon 2 from phCB96-2, the entire preprocathepsin B coding sequence, and the shorter 3′-UTR from plasmid phCB79-2. Briefly, a 357 bp *EcoRI/HindIII* fragment from phCB96-2, containing exons 1, 3, 4 and part of exon 5, was inserted into *EcoRI/HindIII* site of pGEM-2 vector (Promega) to make plasmid phCBe1.3.4.5. phCB79-2-2 was constructed by replacing a sequence from *NheI* site to *ApaI* site of phCB79-2 with 593 bp *NheI/ApaI* fragment from phCBe1.3.4.5. phCB79-2-2,3 is identical to phCB79-2-2 except that it is also missing exon 3. Its preparation is described in Figure 8. A 87 bp *Eco47III/ApaI* fragment from phCBe1.4, containing part of exon 1 and part of exon 4, was used to replace the corresponding sequence from the *Eco47III* to the *ApaI* site in the phCB96-2 to make phCB96-2-3. A 206 bp *EcoRI/HindIII* fragment from phCB96-2-3 was isolated and inserted into the *EcoRI/HindIII* site of pGEM-2, to give phCBe1.4.5. Then phCB79-2-2.3 was constructed by replacing the sequence from the *NheI* site to the *ApaI* site of phCB79-2
Figure 7. Scheme for the construction of plasmid phCB79-2-2. phCB79-2-2, containing a cathepsin B cDNA lacking exon 2, was constructed from phCB96-2 and phCB79-2 in two steps. First, a 357 bp fragment from phCB96-2 was excised with EcoRI and HindIII, and ligated into the EcoRI/HindIII sites of a pGEM-2 vector to make phCBe1.3.4.5. To make phCB79-2-2 a 593 bp fragment was removed from phCBe1.3.4.5 with NheI and ApaI and ligated into the corresponding NheI/ApaI site of phCB79-2. Restriction enzymes used for subcloning in each step are indicated. The cathepsin B cDNA fragment in each vector is shown by double lines. Also shown are restriction sites in each vector.
Figure 8. Scheme for the construction of plasmid phCB79-2-2.3
Figure 8. Scheme for the construction of plasmid phCB79-2-2.3. phCB79-2-2.3, containing cathepsin B cDNA lacking exons 2 and 3 was constructed from phCBe1.4, phCB96-2 and phCB72-2 in three steps. First, an 87 bp fragment from phCBe1.4 was removed with Eco47III and ApaI, and ligated into the Eco47III/ApaI site of phCB96-2 in place of a corresponding region to make phCB96-2-3. Second, a 206 bp EcoRI/HindIII fragment from phCB96-2-3 was ligated into the EcoRI/HindIII sites of pGEM-2 vector to make phCBe1.4.5. Finally, a 442 bp NheI/ApaI fragment from phCBe1.4.5 was used to replace the corresponding NheI/ApaI segment of phCB79-2 to give phCB79-2-2.3. Restriction enzymes used for subcloning in each step are indicated. The cathepsin B cDNA fragment in each vector is shown by double lines. Also shown are restriction sites in each vector.
with the 442 bp NheI/ ApaI fragment from phCBel.4.5.

L. In Vitro Transcription and Translation

Plasmids were isolated by chromatography on QIAGEN columns (QIAGEN Inc., Chatsworth, CA) according to the manufacturer's instruction. Two hundred ml of LB broth containing 100 µg ampicillin/ml was inoculated with E.coli transformed by the appropriate plasmid, and incubated at 37°C overnight with vigorous agitation. The cells were pelleted by centrifugation, resuspended in 4 ml of P1 buffer to which 4 ml of P2 buffer was added, and incubated at room temperature for 5 min. After addition of 4 ml of P3 buffer, the mixture was centrifuged at 4°C for 30 min (15,000xg). The supernatant was then removed and applied onto a QIAGEN-tip 100 column, which was preequilibrated with 3 ml of QBT buffer. The column was next washed with 10 ml of QC buffer, and the plasmid DNA eluted with 5 ml of QF buffer. The collected DNA was precipitated with 0.7 volume of isopropanol, washed with 70% ethanol and dissolved in TE buffer.

For in vitro transcription, the isolated plasmid DNA was linearized with XbaI, and purified by phenol-chloroform extraction and ethanol precipitation. Each transcription reaction contained 5 µg of the appropriate linearized vector, 10 µl 5X transcription buffer (Promega), 10 mM dithiothreitol, 50 units RNasin, 0.5 mM each of ATP, CTP, and UTP, 0.05 mM m7G(5')ppp(5')G and 40 units of SP6 RNA polymerase in final
volume of 50 µl. To estimate RNA synthesis, a 3 µl aliquot of the complete reaction mixture was added to 10 µCi of dried [γ-\(^{32}\)P]UTP prior to the start of the reaction. After incubation at 37°C for 1 hour, the transcripts were purified by phenol-chloroform extraction and ethanol precipitation, dissolved in DEPC-treated water and stored at -20°C. After each purification step, portions of these samples were analyzed on a Northern blot to ensure that no loss or degradation of the RNA had occurred. Reaction mixtures containing radiolabeled UTP were treated with 10 µg of tRNA and 5% trichloroacetic acid, and the insoluble radioactivity measured with a Beckman LS7500 liquid scintillation spectrometer to determine the amount of RNA transcribed.

For *in vitro* translation, 100 ng to 300 ng of capped mRNA was heated at 65°C for 15 min, added to a reaction mixture containing 35 µl of rabbit reticulocyte lysate, 40 units of RNasin ribonuclease inhibitor, an amino acid mixture (minus methionine), 0.02 mM each, and 4 µl of \([^{35}\text{S}]\)methionine (1,100Ci/mmole) at 10 mCi/ml on a final volume of 50 µl. The reaction mixture was then incubated at 30°C for various times from 15 to 60 min. All samples were kept at 4°C before and immediately after translation.

**M. Analysis of the Translation Products**

Five µl samples from each translation mixture were diluted to 10 µl in an electrophoresis sample buffer
containing 0.125 M Tris-HCl (pH 6.8), 2% SDS, 2% 2-mercaptoethanol, 20% glycerol and 0.001% bromphenol blue, boiled for 3 min, and electrophoresed on a 10% polyacrylamide SDS gel in an electrode buffer (pH 8.3) which contained 0.025 M Tris, 0.192 M glycine and 0.1% SDS (Laemmli, 1970). Gels containing the 35S-labeled translation products were dried and autoradiographed overnight at -70°C using an intensifying screen and Kodak XAR film (Laskey, 1980).

To measure [35S]methionine incorporation, a 2 µl aliquot of each translation reaction was incubated in 250 µl of 1N NaOH, 2% H2O2 at 37°C for 10 min. After addition of 1.0 ml of ice-cold 25% trichloroacetic acid/2% casamino acids, the samples were incubated on ice for 30 min. The precipitates were collected on Whatman GF/A filters by vacuum filtration, washed with 3 ml of 5% TCA three times and once with 2 ml of acetone, dried, and the radioactivity measured with a Beckman LS7500 liquid scintillation spectrometer.

N. Folding of an In Vitro Translation Product

Plasmid phCB79-2-2.3 was used as a template for preparation of m7G(5')ppp(5')G capped synthetic RNA as described above. The in vitro translation was described in the previous section, except that the reaction mixture contained 0.6 µg of synthetic, capped mRNA and a complete mixture of amino acids (0.02 mM each) in a final volume of 109 µl.

To promote folding of the translation product, the in
vitro translation mixture (109 µl) was dialyzed against buffer (0.05 M Tris-HCl, 20% glycerol, 0.1 mg/ml bovine serum albumin, 0.15 M NaCl, 1 mM dithiothretiol, and 0.1 mM Na₂EDTA) to which had been added 6.0 M guanidinium hydrochloride. At intervals of 10 minutes or 30 minutes the dialysis buffer was removed, diluted with an equal volume of buffer to reduce the guanidinium hydrochloride by half, and the dialysis continued for two additional cycles of dilution. The final cycle was followed by dialysis against buffer only (Hager and Burgess, 1980; Vinson et al. 1988). The samples were clarified by centrifugation at 14,000 x g for 15 min, quick frozen, and store at -20°C for later analysis.

In an alternative procedure, some samples (108 µl) were treated with 12 µl of a 10X "oxido-shuffling" redox buffer (1X contains 0.15 M Tris-HCl, pH 8.0, 1 mM Na₂EDTA, 0.3 mM glutathione disulfide, and 3.0 mM reduced glutathione) and incubated at room temperature for 24 hours (Jaenicke and Rudolph). Sample were frozen and stored as described above. These procedures are summarized in Figure 9.

**O. Measurement of Cathepsin B Activity**

Cathepsin B activity was measured using the fluorogenic substrate N²-Benzyloxycarbonyl-L-arginyL-L-argine 7-amido-4-methylcoumarin (Z-Arg-Arg-AMC) (Achkar et al. 1990). The enzyme was converted to its free thiol form by preincubation with dithiothretol and EDTA. A 50 µl sample was incubated
Capped synthetic mRNA

\[ \text{in vitro translation} \]

translated product

dialysis against decreasing concentration Guanidinium-HCl

treatment with "oxido-shuffling" buffer

\[ \downarrow \]

6.0 M GmdCl

\[ \downarrow \]

3.0 M GmdCl

\[ \downarrow \]

1.5 M GmdCl

\[ \downarrow \]

0.75 M GmdCl

\[ \downarrow \]

0 GmdCl

\[ \downarrow \]

centrifugation to remove pellet

Supernatant

Cathepsin B assay

addition of 10X "oxido shuffling" buffer (1X contains 0.3 mM glutathione disulfide and 3 mM reduced glutathione

incubation at room temperature for 24 hr

Cathepsin B assay

Figure 9. Procedure for folding of translation products in vitro. In vitro translated products were treated in one of two ways to promote proper folding: with an "oxido-shuffling" buffer to facilitate disulfide bond formation; and with decreasing concentration of guanidinium hydrochloride with changes at 10 and 30 min intervals.
with an equal volume of activation buffer containing 30 mM dithiothreitol and 15 mM EDTA, pH 5.2, for 30 min at 25°C. Then 37.5 µl of 0.6 M citrate/phosphate buffer, pH 6.2 was added, and the reaction was initiated by the addition of 12.5 µl of the substrate (1.128 mM in dimethyl sulfoxide (DMSO)) at a final concentration of 94 µM and 8.3% DMSO. The AMC product was quantitated with an SLM/Aminco SPF-500C spectrofluorometer at excitation and emission wavelengths of 370 and 460 nm, respectively. Some samples were also pretreated for 30 min at 37°C with pepsin (0.83 mg/ml) to remove the remainder of the activation propeptide prior to assay (Achkar et al. 1990; Rowan et al. 1992).
CHAPTER IV

RESULTS

A. Quantification of Total Cathepsin B mRNA in Human Normal Tissues and Tumors by Slot Blot Analysis

Cathepsin B mRNA in different normal human tissues and human tumors was quantified by hybridization of RNA slot blots with labeled human cathepsin B cDNA from phCB79-2 as the probe. Total RNAs were isolated from five kinds of normal human tissues, skeletal muscle, kidney, liver, spleen and colon, and from two cultured human melanoma cells, A375P and A375M (Chomezynski and Sacchi, 1987). Agarose gel electrophoresis of the total RNAs was performed following denaturation with formamide and formaldehyde. The integrity of the 18S and 28S ribosomal RNA indicated that the isolated RNAs were largely intact.

Figure 10A shows a representative slot blot result for cathepsin B mRNA in five normal human tissues and two melanoma cells. The specificity of the cDNA probe was verified by a Northern blot (Figure 11), which shows that the probe only hybridized to cathepsin B mRNAs. Cathepsin B mRNA levels in A375 melanoma variants are higher than those in all the normal tissues examined. A375M, which is highly metastatic, shows a
Figure 10. Slot blot hybridization analysis of human normal tissues and cultured melanoma cells for levels of cathepsin B mRNA. Panel A. Total RNAs were isolated from normal human tissues, skeletal muscle, kidney, liver, spleen and colon, and two cultured human melanoma cells, A375F and A375M. 10 µg, 1 µg or 0.1 µg of total RNA was loaded onto Nytran filter, and the filter was hybridized with 32P-labeled human cathepsin B cDNA. Panel B. To ensure equal loading of RNA in all the slots, the blot was stripped and rehybridized to a control probe, the 32P-labeled cDNA of chicken 185 RNA.
Figure 10. Slot blot hybridization analysis of human normal tissues and cultured melanoma cells for levels of cathepsin B mRNA. Panel A, Total RNAs were isolated from normal human tissues, skeletal muscle, kidney, liver, spleen and colon, and two cultured human melanoma cells, A375P and A375M. 10 µg, 1 µg or 0.1 µg of total RNA was loaded onto Nytran filter, and the filter was hybridized with $^{32}$P-labeled human cathepsin B cDNA. Panel B, To ensure equal loading of RNA in all the slots, the blot was stripped and rehybridized to a control probe, the $^{32}$P-labeled cDNA of chicken 18S RNA.
higher expression of cathepsin B mRNA than A375P which is poorly metastatic, even after adjustments for differences in loading. To ensure that similar amounts of RNAs were present in each sample, the filter was stripped of the cathepsin B cDNA probe, and was rehybridized to a cDNA probe for the chicken 18S rRNA (Figure 10B). Intensities of the hybridization signals were measured by densitometry.

Table 1 shows the relative levels of cathepsin B mRNAs in five normal tissues and two cultured melanoma cells normalized to the levels of 18S rRNA. Compared to the spleen, which has the highest level of cathepsin B mRNA among the five normal tissues examined, the cathepsin B mRNA levels in A375P and A375M were 1.5 and 2.3 fold higher respectively.

B. Quantification of Variant Cathepsin B mRNAs by Northern Blot and RNase Protection Assay

Cathepsin B mRNAs are heterogeneous due to differences in processing of 5'- and 3'-ends of the pre-mRNA (Gong et al. 1993). The variable use of a cryptic 5' splice site in exon 11 with splicing to exon 12 causes the occasional loss of a polyadenylation signal in exon 11 and results in messages which differ by about 2 Kb in the lengths of their 3'-UTR. Variable removal of exon 2 produces messages which differ by 88 nt in their 5'-UTR. The discovery of multiple cathepsin B mRNAs generated by the alternative processing of the pre-mRNA opens up the possibility that it may be a mechanism for regulating cathepsin B expression in human tissues.
Table 1. Relative level of cathepsin B mRNA in human tissues and tumor cells

<table>
<thead>
<tr>
<th>Normal Tissues</th>
<th>Ratio of density (CB/18s RNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon</td>
<td>0.01</td>
</tr>
<tr>
<td>Breast</td>
<td>0.21</td>
</tr>
<tr>
<td>Skeletal Muscle</td>
<td>0.15</td>
</tr>
<tr>
<td>Liver</td>
<td>0.18</td>
</tr>
<tr>
<td>Spleen</td>
<td>0.30</td>
</tr>
<tr>
<td>Kidney</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Cultured Melanomas

<table>
<thead>
<tr>
<th>Cultured Melanomas</th>
<th>Ratio of density (CB/18s RNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A375P, poorly metastatic</td>
<td>0.46</td>
</tr>
<tr>
<td>A375M, highly metastatic</td>
<td>0.68</td>
</tr>
</tbody>
</table>
To investigate alternative splicing of the 3'-UTR, Northern blot analysis was performed. Figure 11 is a representative blot which reveals the presence of the 2.3 Kb and 4.0 Kb transcripts in normal human skeletal muscle, kidney, liver, spleen and colon, and in the two human melanoma variants, A375P and A375M. Densitometric scan of the Northern blot (Table 2) revealed that the ratio of the 2.3 Kb transcript to the 4.0 Kb transcript only varied by about a factor 2, from 2.8 to 6.2. These differences do not appear large enough to suggest that variant 3'-UTRs in cathepsin B mRNA have a regulatory significance.

An RNase protection assay was used to detect and quantify human cathepsin B mRNAs with differing 5'-ends. The riboprobe used is protected by the first 445 bases of mRNA containing the longer 5'-UTR (Figure 4). Based on the sequences of the two 5'-UTRs found in human cathepsin B cDNA (Chan et al. 1986; Gong et al. 1993), hybridization of this riboprobe with human cathepsin B transcripts should result in three protected fragments, a 445 nt fragment corresponding to mRNA containing the full length 5'-UTR, and 276 and 81 nt fragments corresponding mRNA lacking exon 2. These fragments were separated on a denaturing gel, and their sizes were estimated by comparing their mobilities with those of φX174 RF DNA/HaeIII molecular size markers. These three bands were seen in Figure 12. The densities of the 276 nt bands were much higher than those of 445 nt bands in all tumor cells
Figure 11. Northern blot of cathepsin B mRNA from human normal tissues and cultured melanoma cells. Total RNAs were isolated from normal human tissues, skeletal muscle, kidney, liver, spleen and colon, and two cultured human melanoma cells, A375P and A375M. 15 µg of total RNA from each sample was denatured, electrophoresed on agarose gel and transferred onto Nytran filter, and the filter was hybridized with $^{32}$P-labeled human cathepsin B cDNA. The presence of 2.3 Kb and 4.0 Kb transcripts are indicated.
Table 2. Relative proportion of 2.3 Kb and 4.0 Kb cathepsin B mRNAs

<table>
<thead>
<tr>
<th>Normal Tissues</th>
<th>2.3Kb/4.0Kb ratio</th>
<th>2.3Kb/4.0Kb ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skeletal muscle</td>
<td>6.2</td>
<td>skeletal muscle</td>
</tr>
<tr>
<td>Kidney</td>
<td>2.8</td>
<td>skeletal muscle</td>
</tr>
<tr>
<td>Liver</td>
<td>6.1</td>
<td>skeletal muscle</td>
</tr>
<tr>
<td>Spleen</td>
<td>3.5</td>
<td>skeletal muscle</td>
</tr>
<tr>
<td>Colon</td>
<td>2.7</td>
<td>skeletal muscle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cultured Tumors</th>
<th>2.3Kb/4.0Kb ratio</th>
<th>2.3Kb/4.0Kb ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>A375P, poorly metastatic</td>
<td>5.1</td>
<td>poorly metastatic</td>
</tr>
<tr>
<td>A375M, highly metastatic</td>
<td>6.0</td>
<td>highly metastatic</td>
</tr>
</tbody>
</table>

* The signals of 2.3 Kb and 4.0 Kb bands from Northern blot were measured by densitometry. Ratios of the signals of these two bands were calculated.
examined, including five cultured tumor cells, A375M, A375P, PC3M, PC3P and colon carcinoma, and three solid tumors, two colon carcinomas (one a liver metastasis) and a breast cancer. The densities of these two bands were similar in most of normal tissues examined, such as liver, kidney, breast and colon. In contrast, in skeletal muscle the density of the 445 nt band was higher than that of the 276 nt band. The ratio of the intensity of the 276 nt band to that of the 445 nt band is a measurement of the relative proportions of the two transcripts in human tissues and tumors (Table 3). The ratios varied from less than 0.1 to 3.2 for normal tissues, and from 6.0 to 20.1 for the tumors. From the results in Table 3, it can be seen that the transcripts lacking exon 2 are much more abundant in human tumors than in normal human tissues. Figure 12 also reveals the presence of an unexpected fragment of about 140 nt which was found in significant amount in three solid tumors and two cultured human melanoma cells, A375M and A375P. This fragment arises from transcripts lacking exon 3 (see below). When the RNase protection assay was performed with probe alone (without RNA), no protected bands were seen. When the RNase protection assay was performed without RNase digestion, only the 445 base fragment was seen (Figure 13).

C. Amplification of the 5'-Ends of Cathepsin B cDNA from Human Normal Tissues and Tumors by PCR

RNase protection studies revealed a fourth major protected band (140 +/- 10 bases) with mRNAs from human breast
Figure 12. RNase protection of variant 5' -ends of human cathepsin B mRNAs. After hybridization to total RNA, the 445 nt riboprobe is completely protected from RNase digestion by transcripts which contain the complete 5' -UTR. Transcript lacking exon 2 produce 276 and 81 nt fragments. Panel A, RNA from normal tissues and two cultured melanoma cell lines. Panel B, RNA from three different human tumors, two colon carcinomas (one a liver metastasis) and breast cancer. Panel C, RNA from cultured human tumor cells. RNA from some human tumors give rise to a fourth prominent band (*) which appears to be about 140 nt in length. This fragment arises from transcript missing exons 2 and 3.
Table 3. Relative proportion of cathepsin B mRNA transcripts with and without exon 2

<table>
<thead>
<tr>
<th>Normal Tissues</th>
<th>276/445 ratio^*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon</td>
<td>0.8</td>
</tr>
<tr>
<td>Breast</td>
<td>2.4</td>
</tr>
<tr>
<td>Skeletal muscle</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>Liver</td>
<td>2.6</td>
</tr>
<tr>
<td>Spleen</td>
<td>3.2</td>
</tr>
<tr>
<td>Kidney</td>
<td>2.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solid Tumors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon</td>
<td>20.1</td>
</tr>
<tr>
<td>Colon (metastasis to liver)</td>
<td>18.6</td>
</tr>
<tr>
<td>Breast</td>
<td>23.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cultured Tumors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A375P, poorly metastatic</td>
<td>6.0</td>
</tr>
<tr>
<td>A375M, highly metastatic</td>
<td>9.6</td>
</tr>
<tr>
<td>PC3P, poorly metastatic</td>
<td>8.9</td>
</tr>
<tr>
<td>PC3M, highly metastatic</td>
<td>12.0</td>
</tr>
<tr>
<td>Colon, unknown</td>
<td>14.4</td>
</tr>
</tbody>
</table>

^ The signals of 276 base and 445 nt bands from RNase protection assays were measured by autoradiography and densitometry. Ratios of the signal intensities were calculated after correcting for differences in the [α-32P]UMP content of each band.
Figure 13. Evaluation of the riboprobes used in the RNase protection assay. RNase protection assay was performed with probe alone (without RNA), no protected bands were seen. RNase protection assay was performed without RNase digestion, only 445 nt fragment was seen (Riboprobe).
and colon carcinomas and human melanomas. One possibility was that this fragment originated from a cathepsin B transcript in which exon 1 was spliced directly to exon 4, deleting exons 2 and 3 (125 base fragment expected in RNase protection assay). To test this, the 5'-ends of the cathepsin B mRNAs from skeletal muscle, solid colon carcinoma, cultured tumor cells A375M and PC3M were reverse transcribed into cDNA and amplified by PCR. After electrophoresis, the PCR products were transferred onto a Nytran filter and hybridized with human cathepsin B cDNA. This hybridization revealed 549 bp and 461 bp fragments which were derived from cDNAs which contain or lack exon 2 (Figure 14). The 461 bp band was predominant with RNA from the solid colon carcinoma and the cultured tumor cells, A375M and PC3PM, but was barely detectable with RNA from muscle. These results confirmed the findings obtained with the RNase protection assay. The 306 bp band is the expected product for cDNA lacking both exons 2 and 3. In agreement with RNase protection studies, 306 bp band was seen with A375M and solid colon carcinoma, but not with skeletal muscle and PC3M.

D. Molecular Cloning of Human Cathepsin B cDNA Lacking Exons 2 and 3

In order to characterize the transcript lacking exons 2 and 3, a cDNA corresponding to this transcript was cloned by PCR. Total RNA isolated from A375M cells was reverse transcribed and the single stranded cDNA was amplified by PCR.
Figure 14. PCR amplification of the 5'-ends of cathepsin B cDNA from human normal tissues and tumors. This figure shows a Southern blot of products obtained after reverse transcription and PCR amplification of the 5'-ends of cathepsin B cDNA from two human tumor cell lines, A375M (melanoma) and PC3P (prostate carcinoma), a solid colon carcinoma, and normal skeletal muscle. The pair of primers used for the PCR were complementary to regions in exon 1 and exon 5. The 549 and 461 bp bands are from cDNAs which contain and lack exon 2 respectively. The 306 bp fragment is the expected product of cDNA which lacks both exons 2 and 3. The 306 bp band was only seen with cDNAs from tissues which yielded detectable levels of the 140 nt band in RNase protection studies (Figure 12).
The primers used in PCR have the *HindIII* and *EcoRI* restriction sites at their termini, and were designed to amplify the 5'-end of cathepsin B cDNA from exon 1 to the middle of exon 5 as described in the Methods. Because the level of cathepsin B cDNA missing exons 2 and 3 was low, the PCR products were separated by agarose gel electrophoresis, and fragments smaller than 461 bp were eluted and reamplified by PCR. Figure 15 shows the 306 bp product which has the expected size of a cathepsin B cDNA fragment containing exon 1, exon 4 and part of exon 5, and missing exons 2 and 3. This PCR fragment was cloned into the pBluescript vector (Stratagene) to make the plasmid phCBel.4 for single strand sequencing. The complete sequencing of several positive clones confirmed that the 306 bp PCR product was derived from cathepsin B mRNA which lacked exons 2 and 3. Figure 16 shows a portion of a representative sequence which contains an exon 1 - exon 4 junction. In order to further characterize the transcript lacking exons 2 and 3, a fragment from phCBel.4, containing part of exon 1 and part of exon 4, was cloned into phCB79-2 to replace the corresponding region containing exons 2 and 3. The resulting plasmid, phCB79-2-2.3, harbored a cathepsin B cDNA lacking exons 2 and 3 as described in Methods (Figure 8).

E. Detection of Cathepsin B mRNA That Is Missing Exons 2 and 3 by PCR

Total RNA from tumor cells was used to clone a cathepsin B 5' cDNA fragment which is missing exon 2, noncoding, and
Figure 15. PCR amplification of 5'-end of cathepsin B cDNA missing exons 2 and 3. 5'-ends of cathepsin B cDNAs were amplified from A375M cells as described in Figure 15. Products smaller than 461 bp were isolated by agarose gel electrophoresis and used as templates for reamplification by PCR. A 306 bp fragment (indicated by arrow) was obtained, which is expected product of cathepsin B cDNA missing exons 2 and 3.
Figure 16. Sequence of the 306 bp PCR product amplified from cathepsin B cDNA in A375M cells. The 306 bp PCR product, obtained as described in Figure 15, was cloned into EcoRI and HindIII sites of pBluescript vector, and sequenced. The complete sequence of several positive clones revealed that they arose from cathepsin B transcript which lacked exons 2 and 3. A portion of a representative sequence corresponding to the exon 1 - exon 4 junction is shown.
exon 3, encoding the signal prepeptide and part of the activation propeptide. Reverse transcription and PCR amplification was next used to confirm that human tumors contain a corresponding mRNA which contains the remainder of the procathepsin B coding sequence and the 3'-UTR present in other cathepsin B messages (Figure 17). For the amplification, the 5' primer (sense primer) was complementary to 9 nucleotides on each side of the exon 1 - exon 4 junction, and the 3' primer (antisense primer) was complementary to a 20 nucleotide sequence present in the flanking 3' untranslated region of all known, full length cathepsin B mRNAs (Chan et al. 1986). cDNA templates prepared from PC3M cell RNA, and the plasmids phCB79-2 and phCB79-2-2.3, were included to verify the specificity of the sense primer for the exon 1-exon 4 junction. Previous RNase protection studies were unable to detect any cathepsin B transcripts in PC3M that were missing exon 3; plasmid phCB79-2 harbors a full length cathepsin B cDNA, including the complete 3'-UTR of the 2.3 kb human cathepsin B mRNA; and plasmid phCB79-2-2.3 is identical to phCB79-2 except it is missing exons 2 and 3. The 934 bp PCR product, obtained with phCB79-2-2.3 as a template, is indistinguishable in size from the PCR products obtained with the cDNAs from the two human melanomas variants, A375M and A375P. The identity of the PCR products which were amplified from cathepsin B cDNA was confirmed by hybridization with human cathepsin B cDNA.
In Vitro Translation of Variant Cathepsin B mRNA.

RNAse protection assays have shown striking differences in the relative abundances of cathepsin B mRNAs which differ in their 5’-ends in normal human tissues and in human tumors. Exon 2, which is more frequently absent from the 5’-UTR of cathepsin B mRNAs from human tumors, has features which can affect the translational efficiency of the mRNA, the length of the 5’-UTR, and contains adjacent stop codons which prevent protein synthesis. The potential role of second codons in the 5’-UTR of cathepsin B mRNAs which differ in their ability to direct protein synthesis, has been the subject of recent studies. Since exon 3 contains the translation start site, and provides an additional 3’-UTR, cathepsin B mRNAs expressing exon 3 have also been found in normal tissues, and solid colon and breast carcinomas. Since exon 3 contains the translation start site, and provides an additional 3’-UTR, cathepsin B mRNAs expressing exon 3 have also been found in normal tissues, and solid colon and breast carcinomas.

Figure 17. Detection of cathepsin B mRNA that lacks exons 2 and 3 by PCR. Total RNA was reverse transcribed and cDNA amplified by PCR. Plasmid DNA was amplified directly after linearization with XbaI. The sense primer was directed against the exon 1-exon 4 junction and the antisense primer was directed against a 3’ untranslated sequence present in all full length cathepsin B mRNAs. Lane M, molecular size marker prepared by digestion of lambda phage with EcoRI and HindIII; PC3M, human prostate carcinoma; A375M, human melanoma (metastatic); A375P, human melanoma (parental, poorly metastatic); phCB79-2, harbors a cDNA clone containing all 11 exons of cathepsin B including the 798 nucleotide 3’-UTR; phCB79-2-2.3, identical to phCB79-2, except missing exons 2 and 3. Panel A, after electrophoresis of PCR products on a 0.7% agarose gel and staining with ethidium bromide. Panel B, after transfer of the PCR products to a Nytran filter, hybridization to human cathepsin B cDNA and autoradiography.
F. In Vitro Translation of Variant Cathepsin B mRNA

RNase protection assays have shown striking differences in the relative abundances of cathepsin B mRNAs which differ in their 5'-ends in normal human tissues and in human tumors. Exon 2, which is more frequently absent from the 5'-UTRs of cathepsin B mRNAs from human tumors, has features which can affect the translational efficiency of the mRNA. It extends the length of the 5'-UTR, contains adjacent start and stop codons upstream of the normal start site, it is GC rich, and computer analysis indicates the potential for the formation of secondary structure between a region of exon 2 and a nearby downstream sequence. Thus the cathepsin B mRNAs which differ in their 5'-UTRs may vary in their ability to direct protein synthesis. A cathepsin B transcript lacking exon 3 has also been found in melanoma cells, and solid colon and breast carcinomas. Since exon 3 contains the translation start site, and encodes the signal prepeptide and 7 residues of propeptide, it was unclear whether such transcripts containing an exon 1 - exon 4 splice junction would be translationally active. However, there is an in-frame methionine codon, corresponding to amino acid 52 in preprocathepsin B, which potentially could be used for translation initiation. To determine the translational activities and translatability of the various human cathepsin B mRNAs which differ in their exon composition, an in vitro transcription and reticulocyte lysate translation system was used as described in Methods.
The SP6 promoter in pGEM-2 vector has been used to generate specific transcripts with biological activity (Melton et al. 1984). Cathepsin B cDNAs with different exon composition in their 5'-ends were cloned into the EcoRI site, adjacent to the SP6 promoter as described in Methods. phCB79-2, phCB79-2-2 and phCB79-2-2.3 were used to generate the corresponding cathepsin B mRNA species. Cathepsin B mRNA from phCB79-2 contains the longer 5'-UTR, the entire preprocathepsin coding sequence, and the shorter 3'-UTR. Cathepsin B mRNA from phCB79-2-2 is the same as that from phCB79-2 except that it has shorter 5'-UTR lacking exon 2. Cathepsin B mRNA from phCB79-2-2.3 is the same as that from phCB79-2 except that it contains the shorter 5'-UTR, and is also missing exon 3 encoding the prepeptide and 25 residues of the propeptide. Translation from the next in-frame methionine codon should produce a protein lacking the signal peptide and 34 of the 62 residue propeptide. Capped mRNAs transcribed in vitro were examined by electrophoresis on an agarose gel, transferred onto a Nytran filer, and the filter was hybridized with human cathepsin B cDNA. The autoradiogram in Figure 18 shows that essentially all of the transcripts were full length and of the expected size. Phenol/chloroform extraction and ethanol precipitation did not cause the degradation and loss of the transcripts. These transcripts were used for in vitro translation in the reticulocyte lysate in the presence of 35S-methionine.
Figure 18. Analysis of SP6 directed variant cathepsin B transcripts. The capped variant cathepsin B mRNAs were synthesized as described under the Methods. Before (-) or after (+) phenol/chloroform extraction and ethanol precipitation, 0.5 μg of the synthetic RNA was denatured with formamide/formaldehyde, electrophoresed in a 1.1% agarose /0.02% formaldehyde gel and transferred onto a Nytran filter. The filter was hybridized with 32P-labeled human cathepsin B cDNA. Panel A: phCB79-2, RNA transcribed from phCB79-2 which harbors a cDNA clone containing all 11 exons of cathepsin B including the shorter 3'-UTR; phCB79-2-2, RNA transcribed from phCB79-2-2 which is missing exon 2. Panel B: phCB79-2-2.3, RNA was transcribed from phCB79-2-2.3, which is missing exons 2 and 3, and was extracted with phenol/chloroform.
Translation activities were determined by varying both RNA template concentration and incubation time for the translation reaction. The results were expressed both as the rate of labeled methionine incorporation at constant template concentration and as the specific activity of the template (incorporated methionine/ng RNA) at constant time. $^{35}$S-methionine labeled protein was precipitated with trichloroacetic acid, and radioactivity measured. Figure 19 shows the translational activities of two cathepsin B mRNAs with or without exon 2 with respect to the amount of RNA and time of incubation at 30°C. From this figure it can be seen that cathepsin mRNA with the shorter 5'-UTR, the major isoform from tumors (phCB79-2-2) is about twice as active as the mRNA with the longer 5'-UTR, the major isoform from normal tissues (phCB79-2). $^{35}$S-methionine labeled protein was also analyzed by SDS electrophoresis and autoradiography. The protein yield from cathepsin B mRNA with the shorter 5'-UTR (phCB79-2-2) was higher than that from the mRNA with the longer 5'-UTR (phCB79-2) (Figure 20). Figure 20 also shows a single major band of the size expected (approx 37 kDa) for nonglycosalated form of preprocathepsin B (339 amino acids) (Chan et al. 1986). The results showed in Panel B were same as those in Panel A, except varying amount of RNA template was used, and the incubation time was 1 hour.

In order to study the translation potential of cathepsin B mRNA lacking exon 3, the synthetic RNA missing exon 2 and
Figure 19. Rates of *in vitro* translation of synthetic cathepsin B mRNAs containing variant 5'-UTR

A

![Graph A](image)

B

![Graph B](image)
Figure 19. Rates of \textit{in vitro} translation of synthetic cathepsin B mRNAs containing variant 5'-UTRs. Panel A, the specific activities of the two variant templates were determined by incubating reticulocyte lysates for 60 min with the indicated concentrations of capped cathepsin B synthetic RNAs as described in Methods. The amount of $[^{35}\text{S}]\text{L-methionine}$ incorporated into trichloroacetic acid-insoluble products was measured with a liquid scintillation spectrometer. Panel B, the effect of time on $[^{35}\text{S}]\text{L-methionine}$ incorporation into trichloroacetic acid-insoluble products was measured as described for panel A using 300 ng of template in a final volume of 50 µl. Closed circles, mRNA template used in the translation was transcribed from clone phCB79-2-2 lacking exon 2; open circles, from clone phCB79-2 containing exon 2.
Fig 20. SDS-PAGE electrophoresis analysis of in vitro translation products from cathepsin B mRNAs with and without exon 2. Synthetic capped cathepsin B mRNAs were incubated with a reticulocyte lysate at 30°C as indicated. Five µl of the in vitro translation reaction mixture were then electrophoresed on a 10% polyacrylamide-SDS gel and autoradiographed. phCB79-2-2, RNA template transcribed from phCB79-2-2 lacking exon 2; phCB79-2, RNA template from phCB79-2 containing exon 2. In control, no RNA template added to the translation reaction. Panel A, 300 µl of synthetic capped cathepsin B mRNAs and 50 µl of a reticulocyte lysate were incubated for the times indicated. Panel B, the indicated amounts of synthetic capped cathepsin B mRNAs in 50 µl of reticulocyte lysate were incubated for one hour.
Figure 20. SDS-PAGE electrophoresis analysis of in vitro translation products from cathepsin B mRNAs with and without exon 2. Synthetic capped cathepsin B mRNAs were incubated with a reticulocyte lysate at 30°C as indicated. Five µl of the in vitro translation reaction mixture were then electrophoresed on a 10% polyacrylamide-SDS gel and autoradiographed. phCB79-2-2, RNA template transcribed from phCB79-2-2 lacking exon 2; phCB79-2, RNA template from phCB79-2 containing exon 2; control, no RNA template added to the translation reaction. Panel A, 300 ng of synthetic capped cathepsin B mRNAs in 50 µl of a reticulocyte lysate were incubated for the time indicated. Panel B, the indicated amounts of synthetic capped cathepsin B mRNAs in 50 µl of reticulocyte lysate were incubated for 1 hour.
3 from phCB79-2-2.3 was also translated \textit{in vitro} in reticulocyte lysate, and analyzed by SDS-electrophoresis and autoradiography. Figure 21 compares the translation products of all three cathepsin B transcripts containing variant 5'-ends. Not only can the transcript missing exons 2 and 3 (phCB-2-2.3) be translated \textit{in vitro}, it is seen to be much more active than the other two isoforms. A densitometric scan of the autoradiogram in Figure 21 revealed that mRNA obtained from phCB79-2-2.3 was translated nearly 4 times more efficiently than phCB79-2-2 and nearly 8 times more efficiently than phCB79-2. Second, the resulting product is of the expected size (about 32 KDa) for translation initiation from the first methionine codon in exon 4 (288 amino acids).

\textbf{G. Refolding of a Truncated Cathepsin B Produced by the In vitro Translation of mRNA Lacking Exons 2 and 3}

Reticulocyte lysates appear to be able to translate message lacking exons 2 and 3 from a second in-frame methionine codon (residue 52). The expected 32,000 dalton protein product is obtained which should lack the signal peptide and first 34 residues of the 62 amino acid propeptide region. To determine if this product can be folded to produce a native structure, it was treated in one of two ways: with an "oxido-shuffling" buffer containing glutathione disulfide and reduced glutathione to facilitate disulfide bond formation (Jaenicke and Rudolph, 1989); and with decreasing concentration of guanidinium hydrochloride as described in
Figure 21. *In vitro* translation of synthetic cathepsin B mRNAs with variant 5'-ends. 300 ng of synthetic capped cathepsin B mRNAs containing variant 5'-ends in 50 µl of a reticulocyte lysate were incubated for 1 hour at 30°C. Two µl of the mixture from each sample were electrophoresed on a 10% polyacrylamide-SDS gel and autoradiographed. phCB79-2-2, the RNA template transcribed from phCB79-2-2 lacking exon 2; phCB79-2, the RNA template from phCB79-2 containing exon 2; phCB79-2-2.3, the RNA template from phCB79-2-2.3 lacking exons 2 and 3; control, no RNA template added to the translation reaction.
Methods (Hager and Burgess, 1980; Vinson et al. 1988). The ability to acquire enzymatic activity is a highly stringent test of protein folding into a native conformation. The results in Figure 22 demonstrate that the missing propeptide region of the 32,000 dalton protein product is not essential for correct folding and acquisition of enzymatic activity. In vitro translation reactions which lacked the synthetic mRNA templates gave no activity under any of the refolding conditions tried (data not shown).

No systematic study was made of the conditions necessary to optimize the in vitro folding of the 32,000 dalton translation product. Thus, the maximum recovery of activity, obtained by dialysis against decreasing concentrations of guanidinium hydrochloride at intervals of 30 minutes, is still likely to be a lower limit. Using kinetic constants reported for Z-Arg-Arg-NMec hydrolysis by rat cathepsin B (Tchoupe et al. 1991), and the amount of $^{35}S$-methionine incorporated into TCA precipitable products, we estimate that as much as 2% of the synthetic truncated protein may have acquired a native structure. This level was easily measured by the standard cathepsin B assay. Also noteworthy is our finding that the guanidinium hydrochloride treated product was active (33%) prior to treatment with pepsin to remove the remainder of the activation propeptide. Either the loss of the entire activation propeptide is not needed for activation, or some of the enzyme underwent autolysis during folding which removed
Figure 22. Refolding of a truncated cathepsin B produced by the in vitro translation of mRNA missing exon 2 and 3. Capped synthetic mRNA, prepared with plasmid phCB79-2-2.3 as a template, was translated by a reticulocyte lysate as described in Methods. Refolding was initiated by sequential dialysis against decreasing concentrations of guanidinium hydrochloride with buffer changes at 10 or 30 min intervals, or by treatment with a glutathione disulfide/reduced glutathione "oxido-shuffling" buffer. Enzymatic activity is expressed as fmol of 7-amino-4-methylcoumarin (AMC) released from Z-Arg-Arg-NMec per min per µg of synthetic mRNA translated. Activity was measured before and after treatment with pepsin to remove the remainder of amino-terminal propeptide. No activity was observed when synthetic mRNA was withheld from the translation reaction.
the remainder of the propeptide. The "oxido-shuffling" treatment of the in vitro translated product did not yield an active enzyme species.
A. Multiple Forms of Cathepsin B Transcripts in Human Tissues

Cathepsin B cDNA cloned from normal human kidney and the hepatoma cell line, HepG2, are heterogeneous (Chan et al. 1986; Gong et al. 1993). Northern blotting revealed the presence of 2.3 and 4.0 Kb transcripts (Figure 11). Analysis of the cloned cDNAs from both sources has further identified four cathepsin B transcript types as shown in Figure 23. In this present work, I have also identified a fifth type which is present in some human tumors, by using a combination of PCR and RNase protection assays. Transcript type 1 (4.0 Kb) contains a 237 nt 5' untranslated region (5'UTR), 1017 nt encoding preprocathepsin B, and about a 2800 nt 3'UTR, while type 2 is identical to type 1 except for the deletion of an 88 nt segment in the 5'UTR (Figure 23). Transcript types 3 and 4 (2.3 Kb) contain a 789 nt 3'UTR. The types 3 and 4 differ in that type 4 has the same 88 nt deletion in the 5'UTR found in the type 2 transcript. The transcript type 5 has the 88 nt deletion which is found in the transcripts type 2 and 4, and also has an additional 151 nt deletion corresponding to
Figure 23. Variant cathepsin B transcripts in human tissues. Overlapping cDNA clones corresponding to types 1, 2, 3, and 4 were isolated from human kidney and HepG2 phage libraries. The type 5 transcript was cloned from human melanoma cells and characterized by PCR as described in the text. The small open box indicates the exon 1. The solid box indicates the 88 nt 5' untranslated segment which was deleted in transcript types 2, 4, and 5. The type 5 transcript also has an additional 151 nt deletion corresponding to the signal peptide (S) and 7 amino acid residues of the propeptide (P). The hatched box indicates the alternative 3' untranslated region. The part of the 3' untranslated region in the transcript types 1 through 5, which is common to all full length cathepsin B mRNAs, is also shown. The remainder of the 3' untranslated region in the type 5 transcript has not been characterized.
the signal peptide and 25 amino acid residues of the propeptide. The 3'-UTR of the type 5 transcript has not yet been identified. The preprocathepsin B coding sequence was identical in all transcripts except for type 5 which lacked the exon encoding signal peptide and 25 residues of the propetide.

Cloning and sequencing the human cathepsin B gene demonstrated that the multiple cathepsin B transcripts were derived from alternative processing of the pre-mRNA (Gong et al. 1993). Figure 24 shows the organization of the human cathepsin B gene. Phage \( \lambda hCB107 \) contained a 11,500 bp insert which included the first two exons, which corresponded to the 5' untranslated sequences. Phage \( \lambda hCB102 \) contained a 13,500 bp insert which included the entire coding sequence of preprocathepsin B divided into 9 exons numbered 3-11, and contained alternative 3'-UTRs located in exon 11 and exon 12. Thus, RNA processing at a cryptic intron donor site located in exon 11 and splicing to exon 12 deleted the first polyadenylation signal and generated the longer alternative 3'-UTR found in transcript types 1 and 2. In contrast, the 2.3 Kb transcripts (types 3 and 4) are terminated after the polyadenylation signal in exon 11. The deleted 88 nt 5' untranslated sequence found in transcript types 2 and 4 is due to exon skipping in which exon 1 is spliced directly to exon 3. The 239 nucleotide deletion from the 5' end of transcript type 5 is also due to exon skipping in which exon 1 is spliced
Figure 24. Organization of human cathepsin B gene. The introns and exons are shown. The solid areas are the regions which encode preprocathepsin B. Phage \(\lambda hCB107\) contains a 11,500 bp insert which include first two exons containing the 5' untranslated region. Phage \(\lambda hCB102\) contains a noncontiguous 13,500 bp insert which includes the remaining 9 exons numbered 3-11. Also shown are restriction sites.
directly to exon 4.

The nucleotide sequences of the cathepsin B gene at all the exon-intron junctions conform to the canonical GT/AG rule (Table 4) (Breathnach and Chambon, 1981), so that the generation of cathepsin B transcripts with variant 5'-ends cannot be explained simply as the result of a poor splice donor-acceptor pair between exons 1 and 2, and between exons 2 and 3. Similarly, the generation of cathepsin B mRNAs with variant 3'-UTRs cannot be explained as being due to a poor splice donor-acceptor pair between exons 11 and 12. Factors which regulate alternative splice selection will be described later.

In addition to these five types of cathepsin B transcripts in Figure 23, other transcripts have been reported. Human colorectal tumors contain low levels of two additional messages of 1.5 Kb and 3 Kb (Murnane et al. 1991). Human osteoclastomas, which also consist of large numbers of apparently normal osteoclasts, contain 2.4, 1.9 and 1.2 Kb cathepsin B-related mRNAs (Page et al. 1992). The origins and sequences of these transcripts are unclear. In my studies, transcripts smaller than 2.3 Kb were not detected by Northern blot analysis.

Multiple forms of cathepsin B transcripts have also been found in murine melanomas (Qian et al. 1989). Murine B16 melanoma cells contain three transcripts for cathepsin B, a 2.2 Kb form also seen in normal tissues, and two larger
Table 4. Exon-intron organization of the human cathepsin B gene

<table>
<thead>
<tr>
<th>exon size</th>
<th>5' splice donor</th>
<th>intron size</th>
<th>3' splice acceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 124 bp</td>
<td>5' UT CGCTGGgtgag...</td>
<td>3500 bp</td>
<td>...tcccagGCTGGT</td>
</tr>
<tr>
<td>2  88</td>
<td>5' UT CGAGTAgtaag...</td>
<td>5000</td>
<td>...ctccagGTGGAT</td>
</tr>
<tr>
<td>3 151</td>
<td>TrpGln TGGCAGGttagg...</td>
<td>820</td>
<td>AlaGly tcccagGCCGGG</td>
</tr>
<tr>
<td>4  86</td>
<td>GlnArg CCAGAGgtgag...</td>
<td>1450</td>
<td>ValMet atccagAGTTATG</td>
</tr>
<tr>
<td>5 115</td>
<td>CysTrp TGCTGGgttaag...</td>
<td>1750</td>
<td>AlaPhe ctgcagGCCTTC</td>
</tr>
<tr>
<td>6 119</td>
<td>AspGly GGACGGgttagg...</td>
<td>750</td>
<td>CysAsn tttcagCTGAAT</td>
</tr>
<tr>
<td>7  86</td>
<td>HisVal CATGTAGgttaag...</td>
<td>230</td>
<td>GlyCys catcagGGTGCA</td>
</tr>
<tr>
<td>8 114</td>
<td>HisTyr CACTACGgttaag...</td>
<td>470</td>
<td>GlyTyr ttccagGATAAC</td>
</tr>
<tr>
<td>9 117</td>
<td>LysSer AAGTCAGgtgct...</td>
<td>1380</td>
<td>GlyVal tttcagGAGTTG</td>
</tr>
<tr>
<td>10 129</td>
<td>AspAsn GACAATGgtgag...</td>
<td>470</td>
<td>GlyPhe tcccagGCTTCT</td>
</tr>
<tr>
<td>11 142*</td>
<td>3' UT ATCGGGGgtagna...</td>
<td>823</td>
<td>...ctgcagAGAATC</td>
</tr>
<tr>
<td>12 2700</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Exon 11 = 886 bp for 2.3 Kb cathepsin B mRNAs (types 3 and 4) in which the alternative splice site donor site is not utilized. UT stands for untranslated.
transcripts of 4 Kb and 5 Kb only observed in the melanomas. Recently, a 4 Kb message for cathepsin B has been reported in metastatic ras-transformed NIH 3T3 cells (Chambers et al. 1992). Moin et al (1989), employing a human cathepsin B cDNA probe, have also observed a 4 Kb transcript in normal mouse liver. However, Qian et al (1989; 1991c) were unable to detect 4 and 5 Kb messages for cathepsin B in normal mouse liver, kidney, and spleen using either a cathepsin B cDNA probe or a fragment from the 3'-end of the mouse cathepsin B gene specific for the larger transcripts from melanoma cells. These variant transcripts were found to encode the same cathepsin B precursor protein but to differ in the lengths of their 3'-untranslated regions (Qian et al. 1991b; Qian et al. 1991c). The use of alternative polyadenylation signals, which is also seen with the human cathepsin B gene, explains the differences between these variant cathepsin B transcripts in mouse melanomas.

B. Levels of Expression of Variant Cathepsin B mRNAs in Human Normal Tissues and Tumors

Many reports have indicated an association between cathepsin B and cancer. Elevated levels of cathepsin B have been correlated with increased metastatic potential or invasiveness in a variety of tumors from humans (Pietras and Roberts, 1981; Recklies et al. 1982; Dufek et al. 1984; Watanabe et al. 1989; Sheahan et al. 1989) and rodents (Sloane et al. 1982; Koppel et al. 1984; Sloane et al. 1986). Indeed,
elevated cathepsin B expression has been thought to be a marker of the tumorigenic and/or malignant state (Sloane et al. 1982; Koppel et al. 1984). Most of these studies used enzymatic or immunochemical methods to detect and quantitate the enzyme. However, using activity measurements to estimate enzyme levels suffers from uncertainties concerning the specificity of the substrates, and the presence of different amounts of endogenous inhibitors and activators in different tissues which may lead to an under- or over-estimation of the enzyme. Immunochemical methods may be compromised by the cross-reactivity between closely related cysteine proteinases (Wiederander and Kirschke, 1986). In this present study, I have employed cDNA for human cathepsin B to examine its expression in normal human tissues and cultured melanoma cells of different metastatic properties. The mRNA levels for melanomas A375P (parental, poorly metastatic) and A375M (metastatic) are 1.5 and 2.3 fold higher respectively than that for kidney and spleen, the tissues with the highest expression of cathepsin B among the five normal tissues studied. This result is in agreement with the reports of enhanced cathepsin B activities observed in a variety of tumors such as murine fibrosarcomas (Keren and LeGrue, 1988), human liver cancer (Dufek et al. 1984), human ovarian carcinoma (Mort et al. 1981), human genital tract cancer (Pietras and Roberts, 1981) and human breast carcinoma (Recklies et al. 1982). The elevated levels of cathepsin B
mRNA in human melanoma cells suggested that in these cells cathepsin B expression is regulated, at least in part, at the level of transcription and/or mRNA stability.

Post-transcriptional mechanisms may also be involved in the regulation of cathepsin B expression. As mentioned above, cathepsin B mRNA is heterogeneous due to differences in processing of the 5'- and 3'-ends. The selection of variant 5' splice donor sites in exon 11 produces messages which differ in the lengths of their 3'-UTRs. Although the functional significance of mRNAs which vary in their 3'-UTRs is not known in most cases, modification at the 3'-ends of messages have been shown to affect rates of mRNA turnover (Brawerman, 1987; Ross, 1988) and ability to be translated into protein (Kruys et al. 1987). We have shown that both 4.0 Kb (types 1 and 2 in Figure 23) and 2.3 Kb transcripts (types 3 and 4 in Figure 23) were expressed in the normal tissues we examined and in cultured melanoma cells. There may be subtle differences in the expression of the 2.3 Kb transcript versus 4.0 Kb transcript in certain tissues (Table 2). As large differences in the relative abundances of the 2.3 and 4.0 Kb transcripts were not observed between normal tissues and tumor cells, the regulatory significance of these two cathepsin B mRNA forms is unclear at present. The significance of the longer 3'-UTRs in the 4.0 kb and 5.0 kb cathepsin B mRNAs in murine tumors also remains unknown.

Human cathepsin B transcript types 2 and 4 (Figure 23)
were found to result from the occasional skipping of exon 2 during pre-mRNA processing, resulting in an 88 nt deletion from their 5'-UTRs. Differences in the length of the 5'-UTR of the messages has also been shown to affect mRNA stability (Brawerman, 1987) and translation efficiency (Young et al. 1981; Waterhouse et al. 1990; Horiuchi et al. 1990; Manzella and Blackshear, 1990; Foyt et al. 1991; Bares et al. 1991; Kim et al. 1992). In this study, I have observed a greater than 30 fold difference in the relative levels of the two types of messages in normal human tissues (Table 3). Thus, cathepsin B mRNAs which contain the truncated 5'-UTR were barely detectable in skeletal muscle, but were the majority in liver and spleen. Human tumors not only contained higher levels of mRNA for cathepsin B, but the preponderant messages in tumors also contained the shorter 5'-UTR (Table 3). These observations suggest that alternative processing of the 5'-UTR of cathepsin B pre-messages may contribute to the regulation of cathepsin B expression. The differential expression of mRNA with variant 5'-UTR in different tissues has been demonstrated for some other genes, for example, human insulin-like growth factor II gene (Shen et al. 1988), human complement protein C2 (Horiuchi et al. 1990), mouse complement factor B gene (Nonaka et al. 1989), mouse α-amylase gene (Shaw et al. 1985) and Drosophila alcohol dehydrogenase gene (Benyajati et al. 1983; Bond and Davidson, 1986).

Another potentially significant finding in this study is
the presence in human breast and colon carcinomas and in human melanoma of a fifth type of cathepsin B message which lacks exon 2 (non-coding) and also exon 3 (encoding the 17 residue signal peptide and 25 residues of the activation propeptide) (Figure 12 and 15). The 5'-end of this transcript was cloned by PCR techniques (Figure 15) and its identity was confirmed by sequencing (Figure 16). Reverse transcription and PCR amplification confirmed that this transcript contained the remainder of the procathepsin B coding sequence and the 3'-UTR characteristic of normal cathepsin B transcript (Figure 17). Further characteristic of this transcript and its significance in the regulation of cathepsin B expression are discussed below.

C. Mechanisms of Alternative Splicing

The mechanisms involved in the regulation of alternative splicing are the subject of intense study. Though not presently well understood, several themes are emerging. Alternative splicing often appears to be achieved by subtle variations of the basic mechanisms of RNA processing. Many studies have shown that differential splice site utilization involves the selection of a variant of a cis element, such as the consensus sequence around the 5' and 3' splice-site and around the branch point (Ruskin et al. 1985; Aebi et al. 1986; Smith and Nadal-Ginard, 1989). In the case of human cathepsin B, alternative splicing at the 5' end is not due to the
selection of a variant consensus sequence at the 5' and 3' splicing sites, as discussed above.

Regulation of alternative splicing also involves differences in cellular trans-acting factors that usually assemble with the transcript into a splicing complex. Cases of alternative splicing, where identical pre-mRNA transcripts give rise to different mRNAs in a cell-type or developmental stage-specific manner, clearly demonstrate that the splicing environment of different cells varies for these transcripts (Breitbart and Nadal-Ginard, 1987; Wieczorek et al. 1988). The different distributions of variant cathepsin B mRNAs resulting from alternative splicing in human tissues and tumors are more likely due to differences in the cellular trans-acting factors which are involved in pre-mRNA processing.

The trans-acting factors which have been identified to be associated with pre-mRNA processing include snRNP, hnRNP proteins and several other protein factors. U1, U2, U4/U6, and U5 snRNPs are the most extensively characterized functional components of the spliceosome where pre-mRNA splicing takes place (Padgett et al. 1983; Krainer et al. 1984; Kramer et al. 1984; Black et al. 1985; Black et al. 1985; Chabot et al. 1985; Black and Steitz, 1986). Some snRNAs show sequence heterogeneity (Lund, 1988) and are regulated in a developmental and tissue-specific manner (Kato and Harada, 1985; Lund et al. 1985; Lund and Dahlberg, 1987;
Fu et al. 1988; Krof et al. 1988). Some studies of snRNP proteins indicated than these could be modified by phosphorylation (Woppman et al. 1988). Thus, differentially modified snRNP protein components, as well as snRNA-sequence heterogeneity may allow for functionally distinct, tissue-specific or developmentally regulated snRNPs which could permit cells to tune the existing splicing machinery and facilitate differential splice-site utilization. Alterations in U1 snRNP levels have also been shown to affect splice site selection in both artificial and natural pre-mRNAs (Green, 1991).

In addition to the above snRNPs, several protein factors have been shown to be required for splicing, such as hnRNP proteins, A1, A2, B1, B2, C1, C2, and D, which were associated with hnRNA (Dreyfuss, 1986; Chung and Wooley, 1986; Ruby and Abelson, 1991; Conway et al. 1988; Gil et al. 1991; Patton et al. 1991), and splicing factor SF2 (Krainer et al. 1990a,b). Recently, Mayeda and Krainer (1992) reported that hnRNP A1 and splicing factor SF2 could regulate alternative pre-mRNA splicing in vitro. They demonstrated that changes in the relative concentrations of SF2 and hnRNP A1 could modulate alternative 5' splice site selection in vitro in some natural pre-mRNAs, such as the adenovirus E1A pre-mRNA. These splicing factors did not appear to operate in a substrate-specific manner, but rather they had a general effect on the polarity of alternative 5' splice-site selection. Thus, an
excess of SF2 over hnRNP A1 consistently resulted in activation of proximal 5' splice sites, whereas an excess of hnRNP A1 over SF2 generally favors distal 5' splice sites. If regulation of alternative 5' splice-site selection by hnRNPA1 and SF2 takes place in vivo, the concentration or the activities of one or both of these factors is predicted to be subject to tissue-specific, developmental, or physiological control. Interestingly, the intracellular concentration of hnRNP A1 has been reported to increase in proliferating versus quiescent cells (LeStourgeon et al. 1977; de Koch et al. 1981; Celis et al. 1986; Planck et al. 1988). In addition, specific transformation-related changes in the alternative splicing patterns of several cellular genes have been reported (Boris et al. 1987; Zardi et al. 1987; Goodwin et al. 1991). This present study also demonstrates that the human cathepsin B gene can undergo changes in splicing patterns in tumor cells. These results together with the role of hnRNPA1 and SF2 in the alternative splicing suggested that quantitative differences in hnRNPA1 and SF2 might contribute to these transformation-related changes in gene expression. The effect of hnRNPA1 and SF2 on the changes of alternative splicing in human cathepsin B pre-mRNA among different normal tissues and tumors can be tested by quantification of hnRNPA1 and SF2 in these tissues.

Another possibility for different distribution of variant human cathepsin B mRNA in normal tissues and tumors may be
differences in cellular factors which affect stability of the respective mRNAs. This alternative hypothesis can also be tested directly by measuring the rates of decay of the variant messages.

D. Comparison of Translational Activities of Cathepsin B mRNAs with Variant 5'-UTRs

As discussed above, multiple forms of cathepsin B cDNAs which differ in their lengths of 5'-UTRs have been cloned. The results of RNase protection assays confirmed the presence of cathepsin B mRNAs with different 5'-UTRs in human normal tissues and tumors, and also revealed that the ratio of transcripts with a full length 5'-UTR to truncated forms varied widely among normal human tissues, and between normal tissues and tumors. Forms containing a truncated 5'-UTR (lacking exon 2) were much more abundant in human tumors than that in normal tissues. Many studies have demonstrated that 5'-UTR can affect translational capacity of a transcript. To investigate the possible role of the 5'-UTR of cathepsin B transcripts on their translation, the translation efficiencies of cathepsin B mRNAs differing in their 5'-UTRs were studied in this present work. Transcripts lacking exon 2 were twice as active as those containing exon 2 in an in vitro translation assay employing synthetic capped RNAs and a reticulocyte lysate (Figure 19 and 21), suggesting that the in vitro translation of cathepsin B is under the control of the 5'-UTR. These results together with those from the RNase
protection assays suggest that alternative splicing of the 5'-UTR of cathepsin B may play a role in the regulation of cathepsin B expression.

The mechanisms involved in regulation of translation by 5'-UTR have been studied, and three properties of the 5'-UTR have been shown to influence the translatability of an mRNA (Kozak, 1991a,b): (1) the presence of GC rich sequences which can modify secondary structure; (2) the presence of short open reading frames; (3) the length of the 5'-UTR. Rao et al. (1988) demonstrated that 5'-UTR of c-sis/platelet-derived growth factor 2 (PDGF-2) transcript exerted an inhibitory effect on translation. This transcript contains a long 5'-UTR that is highly GC rich and can form stable secondary structures. Deletion of the 5'-UTR resulted in as much as a 10 fold increase in c-sis PDGF-2 expression in vivo. There is growing evidence that expression of a gene may be negatively modulated at the translational level by stable secondary structures in the 5'-UTR of mRNA. It has been observed that the translation rate of ornithine decarboxylase mRNA containing full length 5'-UTR was much lower than that of mRNA containing no 5' lead sequence in both in vivo and in vitro assays (Grens and Scheffler, 1990; Manzella and Blackshear, 1990). This effect was shown to be due to the GC-rich segment of the UTR which has the potential to form a very stable hairpin structure. The effect of secondary structure in the 5'-UTR on the translation of a transcript has also been
studied by insertion of a synthetic oligonucleotide designed to create a hairpin structure upstream from the ATG initiation codon of preproinsulin transcripts (Kozak, 1986; Kozak, 1989). The results revealed that the hairpin structures in the range of stability of $\Delta G = -50 \text{ kcal/mol}$ could reduce the efficiency of translation 85 to 95%, whereas weaker secondary structures ($\Delta G = -30 \text{ kcal/mol}$) did not affect the rate of translation. It was thought that the weaker secondary structure could be readily melted by functional 40S ribosomal subunits which are presumed to bind to the capped 5' end of an mRNA and migrate linearly until reaching the first AUG codon where they are joined by a 60S subunit. In contrast, stable secondary structure could block the migration of the 40S subunits.

Another factor which can affect translational efficiency is the presence of an open reading frame in the 5'-UTR. The occurrence of an upstream AUG start codon which is followed shortly by a termination codon to create a small open reading frame at the 5'-end of the mRNA nearly always reduces the efficiency of translation initiation from a downstream AUG codon. Impairment of translation by an open reading frame burdened 5' untranslated leader sequence has been reported for many mRNAs, such as rat insulin-like growth factor I (Foyt et al. 1991; Lowe et al. 1987), rat farnesyl pyrophosphate synthetase (Teruya et al. 1990), rat thyroid hormone receptor (Murray et al. 1988), human interleukin-7 (Young et al. 1991), human and mouse interleukin-2 receptor (Weinberg and Swain,
118

human fibroblast growth factor-5 (Bares et al. 1991) and transforming growth factor-β3 (Arrick et al. 1991). The simplest explanation is that, after a ribosome translates the upstream open reading frame, the 40S ribosomal subunit may reinitiate at another AUG codon downstream. The inhibitory effect of the upstream AUG codon may be attributable to the inefficiency of this reinitiation. Reinitiation could be inefficient for a variety reasons. One possibility is that ribosomes frequently dissociate from mRNA at the upstream termination codon. A related possibility is that ribosomes remain on the mRNA but are unable to recruit other components of translation initiation before reaching the second downstream open reading frame. A third mechanism of inhibition can be imagined in which ribosomes become stalled at the upstream open reading frame, thereby blocking further scanning downstream (Mueller and Hinnebusch, 1986).

In addition to secondary structures and open reading frames in the 5'-UTR, the length of 5'-UTR has also been shown to affect the translation of mRNA. The transcript of the human complement protein C2 has an unusually long 5'-UTR (388 nt). Deletion of this region resulted in a 10 fold enhancement in the translational efficiency of C2 in transient eukaryotic cellular assays (Horiuchi et al. 1990). A decreased translation rate due to the lengthening of the 5' noncoding sequence was also found for the genes of murine tissue inhibitor of metalloproteinases (Waterhouse et al.
1990), rat guanylate cyclase (Chinkers et al. 1989) and human $\theta$ globin (Leung et al. 1989). It has been reported that RNAs containing a 5'-UTR greater than 200 nt were translated much less efficiently than their truncated counterparts by eukaryotic ribosomes in vitro (Young et al. 1981). Because in most cases the originally long 5'-UTRs contain sequences which can form secondary structures, there is argument that the effect of this kind of 5'-UTR on the translation might be due to both length and secondary structure. Recently Kozak (1991c,d) has used synthetic 5' noncoding sequences which were designed to increase the length of the 5'-UTR without introducing secondary structure to study the effect of the length of 5'-UTR on translation. In contrast to the negative effect on translation reported previously, she found that lengthening the 5' noncoding sequence of transcripts could increase their translational efficiency under conditions of translation in reticulocyte lysates. In these cases translation efficiency was proportional to leader length in the range of 17 to about 80 nucleotide. It seems likely that long 5' leader sequences appear to accumulate extra 40S ribosomal subunits, which may account for their translational advantage. However, the effect on translation of inserting more than 80 nucleotides of synthetic sequence prior to the AUG codon remains unknown.

The 88 nt exon 2 of human cathepsin B transcripts, which is sometimes spliced out of the 5'-UTR to generate transcript
with truncated 5'-UTR, has features which can affect the translation efficiency of the transcript. It extends the 5'-UTR to 237 nucleotides for full length 5'-UTR, contains an adjacent start and stop codon upstream from the normal translation start site and is GC rich. In addition, a preliminary computer analysis of the secondary structure of cathepsin B cDNA indicates that there is a potential for formation of a stable stem-loop structure between part of exon 2 and a nearby downstream sequence. These features may explain the decreased translational activity of the cathepsin B transcript with the full length 5'-UTR in the in vitro assay compared to that of the cathepsin B transcript missing exon 2.

E. Characteristics of Cathepsin B Transcript Missing Exons 2 and 3

A significant new finding in this present work is the detection in human breast and colon carcinomas and in human melanomas of a fifth type of cathepsin B transcript which lacks exon 2, noncoding, and exon 3 encoding the 17 residue signal peptide and 25 residues of the activation peptide (Figure 12, 15, 17, and 24). I have shown that this transcript is complete in that it contains the remainder of the cathepsin B coding sequence and part of the 3'-UTR which is common to all full length cathepsin B mRNAs (Figure 17). I have also demonstrated by in vitro transcription and translation that the corresponding synthetic message lacking exons 2 and 3 can be translated, possibly from an in-frame
internal methionine (residue 52) within the propeptide region. The resulting 32 KDa protein product (Figure 22) would lack the signal peptide which is required for directing cathepsin B to the lysosomes, and the first 34 residues of the 62 amino acid propeptide which inhibits enzyme activity. Remarkably, this RNA is 4 times more active in the reticulocyte lysate than an RNA lacking exon 2, and 8 times more active than the RNA containing the full length 5'-UTR. Furthermore, I determined that this truncated product can be folded to produce a native structure.

After treatment of an in vitro translation mixture with decreasing concentrations of guanidinium hydrochloride to promote refolding, cathepsin B enzyme activity could be detected. In contrast, treatment with an "oxido-shuffling" buffer to promote disulfide bond formation failed to generate an active enzyme. The ability to acquire enzymatic activity is a highly stringent test of folding into a native conformation. Results in Figure 22 show that the activity obtained after dialysis against decreasing concentrations of guanidinium hydrochloride is easily detected by the standard cathepsin B assay. This demonstrates that the 32,000 dalton truncated form of cathepsin B can be refolded to produce an active species, and that the missing propeptide region of this protein product is not essential for correct folding. I also found that treatment of the refolded product with pepsin, to remove the remainder of the activation propeptide, increased
its activity three fold. We postulated that either the removal of the entire activation propeptide may not be necessary for activation, or some of the enzyme may undergo autolysis during folding which removed the remainder of the propeptide.

These results indicated that cathepsin B transcripts from human tumors which lacks exons 2 and 3 are likely to be functional mRNAs. Our ability to refold the translated product of the corresponding synthetic mRNA into an active enzyme \textit{in vitro} increases the likelihood that such message would produce a stable, active protein \textit{in vivo}. Further, we can speculate that this molecular form of cathepsin B, missing the N-terminal signal peptide, may no longer be targeted to the lysosome.

Cellular transformation and other pathological changes can be accompanied by a redistribution of cathepsin B to secretory vesicles (Docherty et al. 1984; Saluja et al. 1989; Achkar et al. 1990), the plasma membrane (Pietras and Roberts, 1981; Sloane et al. 1987; Keren and LeGrue, 1988; Weiss et al. 1990) and the nucleus (Pietras and Roberts, 1981). Transformed cells also contain and secrete several different molecular forms of cathepsin B including higher molecular weight latent and active forms (Recklies et al. 1980; Mort et al. 1980; Dufek et al. 1984; Biaci and Knopeel, 1986; Petrova-Skalkova et al. 1987; Keppler et al. 1988b; Docherty and Phillips, 1988; Maciewicz et al. 1989). Some of these undoubtedly reflect alterations in the linked targeting and
posttranslational processing of the cathepsin B which result in changes in the amount and composition of oligosaccharide side chain (Keppler et al. 1988b; Nishimura et al. 1988) and the extent of proteolytic modifications (Chan et al. 1986; Hara et al. 1988). Our present results also raise the possibility that in some human tumors a part of this diversity of form and location may be due to alternative splicing of cathepsin B pre-mRNA. One can speculate that the putative product of a cathepsin B mRNA missing exon 3 could have a cytosolic and/or nuclear location and could participate in the modulation of cellular activity by the limited proteolytic modification of key regulatory proteins. Interestingly, it has recently been shown that the product of the c-Ha ras gene is structurally homologous to some members of the cystatin type II gene family, and can bind to cathepsin B and related enzymes with high affinity (Hiwasa et al. 1987; Rawlings and Barrett, 1990).
1. Multiple forms of cathepsin B transcripts, which result from alternative processing in the 5' and 3' untranslated regions (UTRs), have been identified in normal human tissues and human tumors by Northern blot and RNase protection assays.

2. The relative levels of the cathepsin B transcripts (2.3 Kb and 4.0 Kb) differing in the lengths of their 3'-UTRs has been measured in normal human skeletal muscle, kidney, liver, spleen and colon, and in the two human melanoma variants A375P and A375M by Northern blot. The ratios of the 2.3 Kb transcript to the 4.0 Kb transcript only varied by about a factor 2, from 2.8 to 6.2. Thus, the regulatory significance of these two cathepsin B mRNA forms is unclear at present.

3. The relative levels of the cathepsin B mRNAs which differ in the lengths of their 5'-UTRs were also measured by RNase protection assays in normal human skeletal muscle,
kidney, liver, spleen and colon, in samples from human colon and breast carcinomas, and in several established cultured human tumor cell lines, two prostate carcinoma variants, PC3P and PC3M, two melanoma variants A375P and A375M, and colon carcinoma. The ratios of the transcript lacking exon 2 to that containing exon 2 varied widely from less than 0.1 to 3 for the normal tissues, and from 6.0 to 24 for the tumors and tumor derived cell lines. Thus, transcripts with the shorter 5'-UTR were much more abundant in tumors than in normal tissues.

3. An in vitro transcription/translation assay was used to demonstrate that the cathepsin B transcripts with variant 5'-UTRs differ in their rates of translation. The relative rates are about 2:1 for mRNAs lacking exon 2 to mRNAs containing the full length 5' end.

4. Human breast and colon carcinomas and human melanomas, A375P and A375M, contain a cathepsin B transcript which is also missing exon 3 encoding the signal peptide and 7 residues of the activation propeptide. This transcript is complete, and contains the remainder of the cathepsin B coding sequence and part of the 3'-UTR which is common to all full length cathepsin B mRNAs.

5. An in vitro transcription/translation assay was used
to demonstrate that cathepsin B transcripts missing exon 3 could be translated from an internal methionine codon (residue 52), producing a 32 KDa product lacking the signal peptide and more than half the propeptide. This truncated form of cathepsin B should no longer be targeted to the lysosome.

6. Finally, the translation product from the message missing exons 2 and 3 was able to acquire a native-like conformation after refolding in vitro. This refolded, truncated protein was active against the specific substrate, N-benzyloxycarbonyl-L-arginyll-L-arginine-7-amido-4-methylcoumarin (Z-Arg-Arg-AMC). Treatment with pepsin, to remove the remainder of the activation propeptide, increases catalytic activity 3 fold. These results indicate that cathepsin B mRNA lacking exons 2 and 3 is likely to be a functional mRNA.

These findings suggest that alternative splicing of cathepsin B pre-mRNA may be involved in regulating the expression and subcellular distribution of cathepsin B in human tissues and tumors.


Aebi, M., Hornig, H., and Weissman, C. (1987). 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* **50:**237-246.


Benyajati, C., Spoerel, N., Haymerle, H., and Ashburner, M.


of transformed fibroblasts is an activatable acid-protease. J. Biol. Chem. 261:1780-1785.


cathepsin H and expression of five lysosomal cathepsins in normal tissues and in a rat carcinosarcoma. *Int. J. Biochem.* **22**:1457-1462.


Weinberg, A.D. and Swain, S.L. (1990). IL-2 receptor (Tac antigen) protein expression is down-regulated by the 5'-untranslated region of the mRNA. *J. Immunol.* **144**:4712-4720.


Zucker, S., Lysik, R.M., Wieman, J., Wilkie, P.D., and Lane,
The dissertation submitted by Qiuming Gong has been read and approved by the following committee:

Dr. Allen Frankfater (Dissertation Director)  
Associate Professor, Biochemistry, Loyola University.

Dr. Richard M. Schultz  
Professor, Biochemistry, Loyola University.

Dr. Mark R. Kelley  
Assistant Professor, Biochemistry, Loyola University.

Dr. Sally A. Amero  
Assistant Professor, Biochemistry, Loyola University.

Dr. Gary L. Engelmann  
Associate Professor, Medicine, Loyola University.

Dr. Russell Pieper  
Assistant Professor, Hematology/Oncology, Loyola University.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Sept 9, 1993  
Date

Allen Frankfater  
Director's signature