

THE CHARACTERIZATION AND EXPRESSION OF MOUSE
BILIARY GLYCOPROTEINS IN NORMAL AND MALIGNANT TISSUES

by Madelaine Rosenberg

A thesis submitted to the Faculty of Graduate Studies and Research of McGill University
in partial fulfillment of the requirements of the degree of Master of Science.

(c) Madelaine Rosenberg

August 1993

Department of Biochemistry
McGill University, Montreal, Quebec, Canada

Shortened version of the thesis title (for M.Sc. Thesis of Madelaine Rosenberg):

The characterization and expression of Mouse Biliary Glycoproteins.

(66 characters including spaces)

ABSTRACT

Mouse biliary glycoproteins (Bgp) are members of the carcinoembryonic antigen (CEA) family, a subfamily of the immunoglobulin superfamily. Nine similar but distinct mouse Bgp cDNA clones have been isolated and characterized. These Bgp isoforms encode polypeptides that contain multiple Ig domains, a single transmembrane domain and a short or long intracytoplasmic domain. Sequence analyses have demonstrated that Bgp isoforms are generated through complex alternative splicing of a single gene and allelic variation. Mouse Bgp proteins are highly expressed in the epithelial cells of the liver and the intestine and are implicated in diverse cellular functions including cell adhesion and bile acid transport, and are postulated to be involved in signal transduction. Investigation into the expression of mouse Bgp isoforms in normal and malignant tissues revealed that Bgp expression is downregulated in tumors. Furthermore, Bgp expression is influenced by transcriptional control mechanisms involving DNA methylation of the *Bgp* gene upstream regulatory regions.

RÉSUMÉ

Les glycoprotéines biliaires (Bgp) sont des membres de la famille des antigènes carcinoembryonnaires (CEA) et font partie de la superfamille des immunoglobulines. Neuf ADNc distincts ont été identifiés chez la souris. Ces ADNc codent pour des protéines Bgp contenant de multiples domaines immunoglobulines, un seul domaine transmembranaire et un domaine cytoplasmique, court ou long. Les analyses de séquences démontrent que ces neuf isoformes sont générés par l'épissage alternatif d'un seul ARN messenger précurseur ou par variation allélique. Les protéines Bgp murines sont exprimées abondamment dans les cellules épithéliales du colon et du foie et sont impliquées dans diverses fonctions incluant l'adhésion cellulaire et le transport des acides biliaires. Il est aussi postulé que ces protéines sont impliquées dans la transmission des signaux moléculaires. L'étude de l'expression des isoformes Bgp dans les tissus normaux et cancéreux chez la souris révèlent que l'expression des protéines Bgp est diminuée dans les tumeurs. L'expression de Bgp est influencée par des mécanismes de régulation transcriptionnelle impliquant la méthylation de l'ADN dans les régions régulatrices du gène *Bgp*.

ACKNOWLEDGMENTS

I would like to take the opportunity to thank the members of my laboratory for their contributions to the results presented in this thesis and for their guidance and participation in helpful discussions. In particular, I would like to thank my supervisor, Dr. Nicole Beauchemin, for her technical help, for the preparation of the manuscript presented in Chapter Two, and for the critical reading of the manuscript presented in Chapter Three and of this dissertation. I would like to thank Patrick Nédélec for the genomic analyses of the mouse Bgp gene, for his contributions to both manuscripts presented and for enduring my teasing. I wish to thank Claire Turbide for sharing her technical expertise in Western analyses and in numerous other techniques and for her contributions to both manuscripts presented. I would also like to thank Andrew Munk for his advice and patience during the preparation of this thesis.

I would also like to thank the Cancer Research Society Inc. and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche for their financial support.

PREFACE

The manuscript entitled Expression of the *Bgp* Gene and Characterization of the Mouse Colon Biliary Glycoprotein Isoforms by *Kimberly McCuaig, Madelaine Rosenberg, Patrick Nédellec, Claire Turbide and Nicole Beauchemin*, presented in Chapter Two of this thesis, was published in the journal *Gene* 127: 173-183, 1993. This paper was a cooperative effort by the members of Dr. Nicole Beauchemin's laboratory. The cloning and sequencing of several of *Bgp* cDNA clones, the PCR amplifications and the Southern analysis presented were done by Kimberly McCuaig. The Northern analyses and the quantifications were done by Dr. Nicole Beauchemin and Claire Turbide. The numerous detailed figures were prepared by Patrick Nédellec and Dr. Nicole Beauchemin. My contributions to this paper include the screening of a mouse colon library and subsequent cloning and characterization of the mouse *BgpD* cDNA isoform. This cDNA clone was used in the reconstruction/cloning of several other isoforms bearing long intracytoplasmic domains. Furthermore, I sequenced on both strands the cDNA isoforms *BgpD*, *BgpE* and *BgpF*.

The manuscript entitled The Expression of Mouse Biliary Glycoprotein, a CEA-related Protein, is Downregulated in Malignant Tissues by *Madelaine Rosenberg, Patrick Nédellec, Serge Jothy, David Fleiszer, Claire Turbide and Nicole Beauchemin*, presented in Chapter Three, was submitted for publication to *Cancer Research* on April 21st, 1993 and was accepted for review by the journal on May 10th, 1993. After the submission of this thesis to the Faculty of Graduate Studies and Research, this manuscript was

accepted for publication in the journal *Cancer Research*. My contribution to this paper include the preparation of the manuscript, the Northern analyses and the preparation of several RNA samples. I also generated transfectant cells expressing BgpD proteins and evaluated the expression of Bgp proteins in transfected cells expressing BgpD using immunoprecipitation and Western analyses. I also assessed the expression of Bgp isoforms in various carcinoma cell lines and in several normal tissue samples by immunoprecipitation and western analyses. Patrick Nédellec, who is also co-first author, performed the genomic analyses of the *Bgp* gene presented in the paper. Dr. Serge Jothy conducted the immunostaining experiments. The antibodies used in Western analyses and several of the RNA samples were prepared by Claire Turbide and Dr. Nicole Beauchemin. Claire Turbide also prepared the colon and liver proteins samples. The primary mouse colon carcinomas were provided by Dr. David Fleischer.

The bibliographic references of the two manuscripts as well as of the introduction and the discussion of this thesis have been amalgamated in the interest of brevity and are found at the end of this dissertation.

This thesis includes the text of original papers accepted for publication or submitted for publication and therefore in accordance to the "Guidelines Concerning Thesis Preparation" from the Faculty of Graduate Studies and Research, the following must be cited:

"Manuscript and Authorship:

The Candidate has the option, subject to the approval of their Department, of including as part of the thesis the text, or duplicated published text, of an original paper or papers.

-Manuscript-style theses must still conform to all other requirements explained in the Guidelines Concerning Thesis Preparation.

-Additional material (procedural design data as well as descriptions of equipment) must be provided in sufficient detail (ie. in appendices) to allow clear and precise judgment to be made of the importance and originality of the research reported.

-The thesis should be more than a mere collection of **manuscripts** published or to be published. It must include a general abstract, a full introduction and literature review and a final overall conclusion. Connecting texts which provide logical bridges between different manuscripts are usually desirable in the interest of cohesion.

It is acceptable for theses to include, as chapters, **authentic copies** of papers already published, provided these are duplicated clearly and bound as an integral part of the thesis. In such instances, connecting texts are mandatory and supplementary explanatory material is always necessary.

-Photographs or other material which do not duplicate well must be included in their original form.

-While the inclusion of manuscripts co-authored by the candidate and others is acceptable, the candidate is required to make an explicit statement in the thesis of who contributed to such work and to what extent, and supervisors must attest to the accuracy of the claims at the Ph.D. Oral Defense. Since the task of the Examiners is made more difficult in these cases, it is in the candidate's interest to make the responsibilities of authors perfectly clear."

TABLE OF CONTENTS

	Page
TITLE PAGE	I
ABSTRACT	II
RÉSUMÉ	III
ACKNOWLEDGMENT	IV
PREFACE	V
TABLE OF CONTENTS	VIII
ABBREVIATIONS	XII

CHAPTER ONE: INTRODUCTION

Carcinoembryonic Antigens	1
The CEA Family of Related Protein	2
Chromosomal Localization and Genomic Organization of the CEA Gene Family	3
The CEA Subgroup	5
The Structure of CEA	5
The Structure of NCA	7
The Structure of Biliary Glycoprotein Isoforms	8
Expression of CEA subgroup members	10
Functions Associated with Members of the CEA Subgroup	12
The PSG Subgroup	13
The Conservation of CEA-related Molecules in Many Species	15
The Characterization of Mouse CEA-related Antigens	15
Mouse Biliary Glycoprotein Expression Patterns	18

**CHAPTER TWO: Expression of the *Bgp* gene and Characterization of
Mouse Biliary Glycoprotein Isoforms**

Summary	20
Introduction	21
Results and Discussion	23
Library cDNA clones	23
Reverse Transcription-PCR Amplification	24
The Nucleotide Sequence of the cDNAs and the Amino Acid Sequences	27
Northern Analysis and Relative Abundance of Transcript Isoforms	30
Conclusions	31
Acknowledgments	35
Figure Legends, Methods and Figures	
Figure 1: Topology of Library cDNA Clones	36
Figure 2: PCR Amplification Reactions and Hybridization of Amplified Fragments	37
Figure 3: Topology of cDNA Sequences of Mouse <i>Bgp</i> Isoforms and Deduced aa Sequences	39
Figure 4: Northern Analyses and Quantifications	41

**CHAPTER THREE: The Expression of Mouse Biliary Glycoprotein,
a Carcinoembryonic Antigen-related Protein,
is Downregulated in Malignant Tissues**

Abstract	42
Introduction	42
Materials and Methods	46
Cell Culture and Transfections	46
Generation of Antibodies	47
Immunodetection Analyses	48
Immunoprecipitation	49
Immunocytochemistry of Tumors	49
[³² P]-Labeled Probes	50
RNA Preparation and Northern Analyses	50
Genomic DNA Analyses	51
Results	52
Transcriptional Activity	52
Posttranscriptional Activity	54
Cellular Expression	57
Regulation of Expression	58
Discussion	61
Acknowledgments	67
Figure Legends and Figures	
Figure 1: Northern Analyses of Mouse <i>Bgp</i> mRNA	68
Figure 2: <i>Bgp</i> protein Isoform Expression	69

Figure 3: Immunostaining	71
Figure 4: Genomic Southern Analyses	72

CHAPTER FOUR: DISCUSSION AND CONCLUSIONS

The Characterization and Expression of Mouse Biliary Glycoproteins in Normal and Malignant Tissues	73
The Role of Alternative Splicing and Allelic Variation in the Generation of Multiple Bgp Isoforms	74
A Role in Cell Adhesion	79
Ecto-ATPase Activity	82
A Potential Role in Signal Transduction	82
Potential Interactions with Cytosolic Proteins	85
A Role in Hepatic Bile Acid Transport	86
Mouse Hepatitis Virus Receptors	88
Functional Implications of the Existence of Multiple Bgp Isoforms	90
The Role of Mouse Bgp Proteins in Malignant and Regenerating Tissues	92
Transcriptional Regulation of the Mouse <i>Bgp</i> Gene	94
Conclusions	97
REFERENCES	100

ABBREVIATIONS

A domain	C-2 set Ig domain
aa	amino acid(s)
B domain	C-2 set Ig domain
BGP	biliary glycoprotein
Bgp	murine biliary glycoprotein
<i>BgpX</i>	cDNA encoding mouse Bgps (replaces mmCGM to conform with mouse genome nomenclature; <i>X</i> is assigned by order of characterization)
bp	base pair(s)
C	carboxyl
Ca ²⁺	calcium
CAM	cell adhesion molecule
CD	complement determining
cDNA	complementary DNA
CEA	carcinoembryonic antigen
CGM	CEA-related gene family member
cys	cysteine
Cyt	intracytoplasmic tail
CytL	long cytoplasmic domain
CytS	short cytoplasmic domain
DNA	deoxyribonucleic acid
EGF	epidermal growth factor

GPI	glycanphosphatidylinositol
ICAM-1	intercellular adhesion molecule-1
Ig	immunoglobulin
IgC	Ig constant
IGF-I	Insulin-like growth factor-I
IgV	Ig variable
IT	intracytoplasmic
kb(s)	kilobase(s)
kD	kilodaltons
LIPE	hormone-sensitive lipase
MHV	mouse hepatitis virus
MHVR	MHV receptor
mmCGM5	<i>Mus musculus</i> CGM 5
MMTV	mouse mammary tumor virus
MOPS	4-morpholinopropanesulfonic acid
~ <i>Mr</i>	approximate molecular weight in daltons
mRNA	messenger ribonucleic acid
N	amino
NCA	non-specific cross-reacting antigen
N-CAM	neural cell adhesion molecule
N-linked	linked to amino group of asparagine
nt	nucleotide(s)
oligo	oligodeoxyribonucleotide
ORF	open reading frame
P	phosphorus

PBS	phosphate-buffered saline
p.c.	post coitum
PCR	polymerase chain reaction
PDGFR	platelet-derived growth factor receptor
pI	isoelectric point
PKC	protein kinase C
PSG	pregnancy-specific glycoprotein
Psg	mouse or rat pregnancy-specific glycoprotein
RIT	oligonucleotide specific for IT in the antisense orientation
RNA	ribonucleic acid
RT	reverse transcription
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis.
TCR	T-cell receptor
TGF- β 1	transforming growth factor- β 1
TM	transmembrane
TPA	12- <i>O</i> -tetradecanoylphorol 13-acetate
5' or 3'UTR	5' or 3' untranslated region

CHAPTER ONE

INTRODUCTION

Carcinoembryonic Antigens

Cancer is one of the leading causes of death in the industrialized world and is thought to arise from a series of alterations that result in the progressive loss of cellular growth control mechanisms (Nowell, 1986). Improved clinical methods which allow malignancies to be detected earlier and treated with higher specificity and efficiency are being developed. These methods exploit, in part, the differences between cancerous and normal tissues. One area of intense investigation has been the study of proteins which are differentially expressed in tumor and normal cells. Carcinoembryonic antigen (CEA), a widely used human tumor marker, is one of the most extensively characterized tumor-associated proteins (Shively and Beatty, 1985).

The overexpression of CEA is a common phenotypic change in variety of malignancies including colon, breast, lung and pancreatic carcinomas (Thompson et al., 1991). CEA was originally described by Gold and Freedman as a tumor specific-antigen found specifically in colorectal cancers and fetal gut (Gold and Freedman, 1965). Subsequent studies, however, revealed the presence of CEA in normal body fluids and normal colonic mucosa, although in significantly lower amounts (Chu et al., 1972; Kupchik and Zamcheck, 1972; Egan et al., 1977; Fritsche and Mach, 1977). The clinical applications of CEA as a tumor marker were further complicated by the discovery of several antigens expressed in a variety of normal and malignant tissues which cross-reacted immunologically with CEA (Thompson and Zimmermann, 1988). Nevertheless, the measurement of serum CEA levels remains a widely used diagnostic tool for detecting post-operative recurrences of colorectal breast and lung cancers

(Shively and Beatty, 1985).

CEA is the prototype of a large family of structurally and immunologically-related proteins which are expressed in a variety of normal and malignant tissues. The roles of these proteins in tumor progression as well as in normal cellular functions are of considerable interest but are presently ill-defined. The molecular cloning and biochemical characterization of CEA family members has given some insight into their possible functions.

The CEA Family of Related Antigens

The human CEA family encompasses numerous membrane-anchored and secreted glycoproteins that share a number of structural features which place them within the immunoglobulin (Ig) superfamily (Williams and Barclay, 1988). All CEA family members possess an N-terminal domain of 108-110 amino acids that resembles an Ig variable (IgV)-like domain followed by a varying number of C-2 set Ig constant (IgC)-like domains denoted as A domains (92-96 amino acids) or B domains (86 amino acids) (Thompson et al., 1991). The deduced secondary structures of CEA gene family members reveal that each C2-set domain forms an array of anti-parallel beta strands which fold into two beta sheets between which a hydrophobic interior is formed, a common feature of Ig-like folds (Williams and Barclay, 1988; Bates et al., 1992). Each C2-set domain possesses two highly conserved cysteine residues which are presumed to form a disulphide bridge stabilizing the Ig-like fold (Williams and Barclay, 1988). Although the N-terminal IgV-like domains of CEA family members lack these conserved

cysteine residues, computer analyses predict that they are also assembled in typical Ig-like folds which are believed to be stabilized by hydrophobic interactions and conserved salt bridges (Williams and Barclay, 1988; Thompson et al., 1991). All CEA-related molecules possess a highly conserved 34 amino acid leader sequence which serves to guide the nascent polypeptide chain through the rough endoplasmic reticulum and the Golgi complex where posttranslational modifications occur (Thompson et al., 1991). The leader peptide is removed and does not form part of the mature protein; the processed polypeptide is then directed to the plasma membrane where it either becomes anchored or is secreted (Thompson et al., 1991).

The CEA gene family presently includes 22 distinct genes that can be divided into two subgroups based on sequence comparisons and expression patterns: the CEA subgroup and the pregnancy-specific glycoprotein (PSG) subgroup (Thompson et al., 1991). The distinguishing features of the CEA family members lie in their number of IgC-like domains and in their C-terminal structures. Members of the CEA subgroup contain either a transmembrane domain or a hydrophobic C-terminal sequence that direct covalent membrane attachment. The PSG subgroup members display hydrophilic C-terminal domains or lack C-terminal structures altogether and are secreted.

Chromosomal Localization and Genomic Organization of the CEA Gene Family

The entire human CEA gene family is encompassed within a 1.2 megabase region on chromosome 19 that maps to 19q13.1-19q13.2 in the vicinity of the genes encoding the cytochrome P₄₅₀ subfamily IIA, transforming growth factor (TGF)- β 1 and hormone-

sensitive lipase (LIPE) (Thompson et al., 1991). Members of the CEA gene family are further divided: the CEA subgroup genes are clustered together, as are the PSG subgroup genes which map telomerically to the CEA subgroup genes (Thompson et al., 1991). The CEA genes are also very closely linked; some CEA subgroup members are separated by as little as 16kb (Thompson et al., 1991) while several *PSGs* genes are only 6kb apart (Thompson et al., 1990).

Genomic analyses of CEA gene family members have revealed a direct correlation between exons and protein domain structures, a feature found in other Ig superfamily members (Thompson et al., 1991). In general, the first exon encodes the 5' *UTR* and the first two thirds of the leader domain while the second exon codes for the last third of the leader domain and the entire N-terminal domain. In rodent *Psgs*, the second exon is repeated up to 5 times (Rebstock et al., 1990). Beyond the N-terminal domain exon(s), a variable number of C-2 set (A or B domain) encoding exons are found and generally occur in the order A1 B1, A2 B2 etc. The number of the C-2 set domain exons ranges from one (*mmCGM5*, a mouse *Psg* gene) to six (the human *CEA* gene) (Thompson et al., 1991). Variability is observed in the structure and number of the exons that follow the C-2 set Ig domains. They may code for hydrophilic C-terminal domains, membrane attachment or transmembrane and cytoplasmic domains; these exons are always followed by a 3' *UTR* (Thompson et al., 1991). However, as alternative splicing has been demonstrated for several CEA gene family members, not all putative exons are found in the mature transcripts and their corresponding proteins (summarized in Thompson et al., 1991).

The CEA Subgroup

The CEA subgroup includes the *CEA* gene itself, the biliary glycoprotein (*BGP*) gene, the non-specific cross-reacting antigen (*NCA*) gene, the CEA gene family member 6 (*CGM6*) gene, and five additional *CGM* genes (1, 2, 7-9) for which the natural gene products have not yet been identified (Thompson et al., 1991). Antigens in this subgroup are generally expressed at the surface of epithelial cells in both normal and malignant tissues (Berling et al., 1990; Thompson et al., 1991). cDNA and protein sequence analyses indicate that most members of the CEA subgroup are membrane-anchored either by glycanphosphatidylinositol (GPI)-linkages, as is the case for CEA, NCA and CGM6, or by a transmembrane and cytoplasmic domain as is seen in BGP isoforms (Thompson et al., 1991). The structures, expression patterns and deduced functions of CEA, NCA and BGP have been investigated and are described below.

The Structure of CEA

The biochemical features, tissue distribution and clinical relevance of CEA have been thoroughly characterized (Shively and Beatty, 1985). As deduced from cDNA clones, the mature *CEA* mRNA codes for a polypeptide with the following features: a leader peptide which is removed during posttranslational processing, a 108 amino acid N-terminal domain, six C-2 set Ig domains denoted as A1 B1, A2 B2 and A3 B3, and a hydrophobic 26 amino acid C-terminal domain (Beauchemin et al., 1987; Oikawa et al., 1987a,b; Kamarck et al., 1987; Zimmermann et al., 1987). Protein sequencing confirmed the identity of the N-terminal domain and the C-2 set domains (Hefta et al.,

1988). However, protein sequencing also indicated that the 26 amino acid C-terminal domain previously postulated to be the CEA membrane-anchoring component, was not found in the mature protein (Hefta et al., 1988). This domain was subsequently shown to be removed during posttranslational processing and replaced by a complex ethanolamine-GPI moiety which anchors the protein to the membrane (Hefta et al., 1988; Takami et al., 1988).

While the exact mechanisms of this type of posttranslational processing are not fully understood, GPI-linkages are found in a large number of eukaryotic proteins (Cross, 1990), including in the Ig superfamily members neural cell adhesion molecule (N-CAM) (Hemperly, 1986) and Thy-1 (Tse et al., 1985). GPI-membrane bound proteins are believed to possess a larger mobility in the membrane plane than integral membrane proteins which may have consequences on protein interactions and functions (Cross, 1990). CEA can be freed from the membrane surface by digestion with phosphatidylinositol specific phospholipase C or D which cleave GPI-linkages (Hefta et al., 1988). It has been suggested that the amount of CEA bound to the membrane may be regulated by the release of similar digestion enzymes into the serum (Thompson et al., 1991). Such mechanisms could explain the rapid turnover of CEA in the mucosa of the normal intestine and possibly the increased levels of CEA in the serum of colorectal cancer patients (Thompson et al., 1991).

CEA is a heavily glycosylated protein that is 50-60% carbohydrate by weight (Thompson and Zimmermann, 1988). The monosaccharide content of CEA consists primarily of glucosamine and negligible amounts of galactosamine indicating that the

carbohydrates are primarily asparagine (N)-linked (Thompson et al., 1991). CEA possesses 28 potential N-linked glycosylation sites; it is predicted that each carbohydrate chain would have to contain, on average, 20 monosaccharides for CEA to possess its observed molecular weight of 180kD (Thompson et al., 1991). Differences in CEA N-linked glycosylation has been detected in samples obtained from different sources such as normal colonic mucosa and colorectal adenocarcinomas, suggesting that these difference may be due to different glycosylation systems in normal and cancerous tissues (Garcia et al., 1991). These variations may provide specific determinants found only in tumor-associated CEA which could facilitate the design of toxins which could be targeted to tumors expressing CEA (Jothy et al., 1986).

The Structure of NCA

NCA contains an N-terminal IgV-like domain of 108 amino acids followed by two C-2 set Ig domains, denoted as A1 and B1, and a hydrophobic 24 residue C-terminal domain which, as in CEA, is replaced during posttranslational processing by a GPI moiety anchoring the mature protein to the membrane (Kolbinger et al., 1989). Variations in the sugar content of NCA which has 12 potential N-linked glycosylation sites, have also been observed (Thompson et al., 1991). NCA (45kD) and TEX (75kD), two NCA isoforms initially characterized as distinct proteins, were shown by fast bombardment mass spectrometry and protein sequencing to have the same polypeptide backbone (Hefta et al., 1990). These two NCA isoforms correspond to NCA-50 and NCA-90 which have been shown independently to be encoded by the *NCA* gene

(Kolbinger et al., 1989).

The Structure of Biliary Glycoprotein Isoforms

The *BGP* gene is a distinct member of the CEA gene family since it codes for multiple isoforms which contain a transmembrane domain and a short or long intracytoplasmic (IT) domain. BGP was first identified by Svenberg in 1976 as a normal component of human bile which was immunologically-related to CEA and NCA (Svenberg, 1976). Eleven different *BGP* gene isoforms have been identified through molecular cloning and PCR amplification of reverse transcribed RNA (Hinoda et al., 1988; Barnett, et al., 1989; Kuroki et al., 1991; Barnett et al., 1993). These *BGP* isoforms are derived through complex alternative splicing of a single genomic transcriptional unit which results in the exclusion of specific exons from the mature mRNA (Barnett et al., 1993).

BGP_a, the largest BGP isoform to be characterized at the cDNA level, is composed of a 5'UTR, a leader peptide, a N-terminal domain, three C2-set domains denoted as A1, B1, A2, followed by a membrane spanning domain, a 71 amino acid IT domain and 3'UTR (Barnett et al., 1989). In three similar isoforms, variations are seen in the inclusion or exclusion of the A2 domain and in the length of the IT domains. *BGP_b* is very similar to *BGP_a* in all aspects except it lacks the A2 domain (Barnett et al., 1989). *BGP_c* is identical to *BGP_a* in its extracellular portions but possesses a short IT domain of only 10 amino acids (Barnett, et al., 1989) while *BGP_d* lacks the A2 domain and contains a short IT domain (Barnett et al., 1989).

BGP isoforms bearing three C-2 set domains possess 19 potential N-linked glycosylation sites (Hinoda et al., 1988; Barnett et al., 1989) while BGP proteins have been shown to be ~40% carbohydrate in weight (Svenberg et al., 1979a). The amino acid sequence of the A2 domain of BGP is only 45% homologous at the amino acid levels to all other IgC-like domains of CEA, NCA and BGP, which are over 75% homologous among themselves (Barnett et al., 1989). The B1 domain contains an additional unpaired cysteine residue near the junction with the A2 domain or transmembrane domain (depending on the isoform) which has prompted the suggestion that BGP isoantigens may form homodimers or heterodimers linked together by disulphide bridges (Barnett et al., 1989).

Four novel *BGP* cDNAs which differ substantially from the previously described isoforms, have recently been identified by PCR amplification of reverse transcribed mRNA (Barnett et al., 1993). The two most intriguing isoforms, *BGP_y* and *BGP_z*, code for antigens identical to *BGP_a* and *BGP_c* respectively, except that their A2 domain is replaced by one of two 31 amino acid domains which bear no homology to other CEA family members or Ig superfamily members. The two 93 base pair exons which code for these domains show significant homology to human *Alu* repeat structures (Barnett et al., 1993). The *Alu*-like domains of *BGP_y* and *BGP_z* are 78% homologous to one another, as is expected of *Alu* repeats. The biological significance of these *Alu*-like domains is unknown. The other two novel isoforms, *BGP_x* and *BGP_{x'}*, code for polypeptides in which the N-terminal domain is linked to the transmembrane domain and possess either a long and short IT domain respectively. These two isoforms lack the A1,

B1 and A2 IgC-like domains which are found in all other CEA-related gene products suggesting that these isoforms may have particular functions (Barnett et al., 1993). Initial studies indicate that these newly identified isoforms are more abundant in tumors than in normal tissues, suggesting that the expression of these BGP isoforms may be associated with the transformed phenotype (Barnett et al., 1993).

Three additional *BGP* cDNA isoforms designated *BGPg*, *BGP_h* and *BGP_i*, have been cloned from a human leukocyte library (Kuroki et al., 1991). The deduced proteins have the following structures: Leader-N-A1-B1-A2, Leader-N-A1-B1, and Leader-N-A1-B1-C-terminal, respectively. All three proteins lack transmembrane and IT domains and are presumably secreted. While the putative proteins encoded by these cDNAs have not been detected in leukocytes or other hematopoietic cells, they are postulated to correspond to the BGP species detected in human bile or serum (Kuroki et al., 1991).

Expression of CEA Subgroup Members

CEA subgroup members show a great deal of variability in their patterns of expression in both normal and malignant tissues. Strong CEA expression is found in the majority of human colonic and gastric carcinomas, in adenocarcinomas of the breast and the pancreas, and in non-small cell lung cancers (Thompson et al., 1991). CEA is weakly expressed in normal colonic epithelia while large amounts are present in the developing fetal gut (Thompson et al., 1991). CEA expression is not detected in leukocytes (Berling et al., 1990). In contrast, NCA is expressed in cancers of hematopoietic and epithelial origin and is also present in a variety of normal fetal and

adults tissues including spleen, lung, liver, colon, and several types of blood cells (Thompson et al., 1991). The expression of CGM6 appears to be restricted to leukocytes (Berling et al., 1990).

BGP isoforms are also found in a variety of cells of epithelial and hematopoietic origin. Initial immunodetection experiments demonstrated that BGP proteins were present in normal biliary tracts (Svenberg et al., 1979b). Subsequent experiments showed that BGP proteins are expressed in many normal tissues including colonic epithelial cells (Hinoda et al., 1988; Kuroki et al., 1991), hepatocytes (Hinoda et al., 1990), leukocytes (Kuroki et al., 1991), human fetal liver (Barnett et al., 1989) and a variety of other normal tissues (Barnett et al., 1993).

While BGP proteins are found in the serum of normal individuals, elevated BGP serum levels are found in patients suffering from non-malignant liver or biliary disorders (Svenberg et al., 1976; Hinoda et al., 1988). The increase in BGP serum concentrations has been correlated with biliary obstructions involving post-hepatic jaundice or primary biliary cirrhosis (Svenberg et al., 1979c). BGP expression is also associated with malignant diseases: *BGP* transcripts are present in cell lines derived from several colonic adenocarcinomas, squamous bladder carcinomas, pancreatic cancer and acute myelogeneous leukemias (Barnett et al., 1989; Barnett et al., 1993). Hinoda et al., have reported that there is no significant difference in *BGPI* (*BGP α*) mRNA expression in normal and malignant human liver tissues and human colon tissues, suggesting that the *BGP* gene expression is governed by different mechanisms than those which regulate *CEA* and *NCA* gene expression (Hinoda et al., 1990).

Functions Associated with Members of the CEA Subgroup

The physiological functions of CEA-related antigens are not well understood. The functional characterization of CEA subgroup members has been facilitated by the availability of cDNA constructs. *In vitro* aggregation assays using cDNA-transfected cells have demonstrated that CEA, BGP, NCA and CGM6 function as intercellular adhesion molecules (Benchimol et al., 1989; Oikawa et al., 1989; Rojas et al., 1990; Zhou et al., 1990; Oikawa et al., 1991; Oikawa et al., 1992). CEA and NCA mediate cell adhesion in a Ca^{2+} -independent manner, a feature found in other Ig-related adhesion molecules (Benchimol et al., 1989; Oikawa et al., 1989; Zhou et al., 1990). Some discrepancies in the properties of BGP adhesion exist: BGPa and BGPb were reported by Rojas et al., 1990 to mediate cell adhesion in a temperature- and calcium-dependent fashion while Oikawa et al., 1992 reported that BGPa is involved in Ca^{2+} -independent adhesion activities. These differences may arise from different levels of BGP expression at the cell surface or the fact that each laboratory used different cellular backgrounds (Rojas et al., personal communication; Oikawa et al., 1992). CEA, BGP and NCA mediate homotypic adhesion while CEA is capable of mediating heterotypic adhesion with both NCA and BGP (Zhou et al., 1990; Oikawa et al., 1992). In contrast, CMG6 mediates heterotypic cell adhesion specifically with NCA, indicating that cellular adhesion by CEA family members is not just a fortuitous activity (Oikawa et al., 1991).

Intercellular adhesion plays an important role during embryonic development and morphogenesis (Edelman, 1988). The expression pattern of CEA proteins during embryonic development correlates with an *in vivo* adhesive function (Benchimol et al.,

1989). CEA is implicated in three of the four stages of colorectal carcinogenesis (Jessup and Thomas, 1989) and is postulated to play a role in intestinal tissue organization during development (Benchimol et al., 1989).

Other functions have been associated with members of the CEA subgroup. BGP, CEA and NCA have been shown to bind through their carbohydrate moieties *Escherichia coli* isolated from human gut and various strains of *Salmonella* (Leusch et al., 1991). CEA-bacterial binding is only possible in the intestine where CEA is principally expressed. Such a process could favour the bacterial colonization of the normal colonic mucosa (Thompson et al. 1991). The rapid shedding of CEA from colonic epithelia may modulate the steady-state levels of the intestinal flora (Thompson et al., 1991). Alternatively, the presence of CEA subgroup members on the surface of leukocytes could favour the binding of bacteria and ultimately phagocytosis (Thompson et al., 1991).

The Pregnancy-Specific Glycoprotein Subgroup

The PSG subgroup is composed of numerous closely related glycoproteins encoded by at least 11 genes (Thompson et al., 1991). PSGs are on average 30% carbohydrate by weight (Watanabe and Chou, 1988) and are produced in large quantities in the placenta during pregnancy forming the major placental products found in maternal serum (Lin et al., 1974). All *PSG* genes and cDNAs characterized so far possess similar domain organizations with a leader, an N-terminal domain and multiple C-2 set domains (Chan, 1990). In contrast to CEA subgroup members, the majority of *PSG* cDNAs contain short hydrophilic or degenerate C-terminal domains which do not allow for

membrane anchorage. These structural features combined with the fact that placental PSGs are found in the maternal serum (Lin et al., 1974) strongly suggest that most PSGs are not bound to the plasma membrane and are secreted (Chan, 1990).

PSGs have also been found in various extraplacental sources, albeit in lower amounts. The expression of PSGs in fetal liver, uterus, testis, intestine and various hematopoietic cell types is well documented (reviewed in Chan, 1990). The expression of PSGs in gestational trophoblastic tumors including choriocarcinomas and chorioepitheliomas has also been described and suggests a potential use of PSGs as tumor markers for certain forms of cancer (Chan, 1990).

Although human PSGs have been extensively characterized chemically and structurally, little is known about their functions. PSGs have been shown to possess *in vitro* immunosuppressive properties while tentative *in vivo* experiments indicate that PSGs may function as immunomodulators during pregnancy (Cerni, 1977; Chan, 1990). The mechanism of this inhibitory function, however, remains to be elucidated (Chan, 1990). Other clues into the possible function(s) of PSGs may arise from amino acid sequence comparisons with other known proteins. For example, RGD tripeptide sequences which have been shown to modulate cell adhesion to fibronectin and a variety of other adhesive proteins (Hynes, 1992) are found in the N-terminal domain of several PSG members. This may suggest that some PSGs may function as receptors and/or adhesion molecules (Thompson et al., 1991). While the physiological functions of this CEA-subgroup of proteins remain unclear, they are believed to be significant in light of the large quantity of PSGs produced during pregnancy and their identification as oncofetal proteins (Ogilvie

et al., 1989).

The Conservation of CEA-related Molecules in Many Species

CEA-like proteins have been identified in a number of non-human mammalian species. Antigens immunologically related to human NCA and CEA have been detected in *Cynomolgus* monkey spleen and lung tissue samples (Engvall et al., 1976). CEA-related antigens have also been detected in the blood of chimpanzees, baboons, gorillas and orangutans (Haagensen et al., 1982). Antigens similar to but serologically distinct from CEA were detected in the rat intestine as early as 1976 (Abeyounis and Milgrom, 1976). Both rat (Lin and Guidotti, 1989; Aurivillius et al., 1990; Margolis et al., 1990) and mouse (Turbide et al., 1991; McCuaig et al., 1992; this thesis) *Bgp* homologs have been identified and extensively characterized. Bovine and non-human primate PSGs have also been characterized (Chan, 1990) while antigens which exhibit similar expression patterns but low sequence homology to human PSGs are presently being characterized in the rat and the mouse (Kodelja et al., 1989; Rudert et al., 1992; Rebstock et al., 1990). The broad species conservation of CEA-related proteins through evolution indicates that the functions of CEA-related antigens may be essential and further suggests that CEA-related molecules may exist in the lower orders of the animal kingdom.

The Characterization of Mouse CEA-related Antigens

In depth investigations into the function(s) of CEA family members in normal and malignant tissues as well as their role in development requires the use of animal model

systems. Rodents, mice in particular, are appropriate animal models since they are well characterized developmentally and genetically. Functional analyses in mice are facilitated by techniques such as controlled carcinogenesis, transgenesis and gene targeting. As a first approach, the laboratory of Dr. N. Beauchemin has investigated the mouse CEA-related genes. Initial analyses of mouse genomic DNA revealed the presence of multiple CEA-like genes while Northern analyses showed that discrete CEA-like transcripts were found in mouse colon and liver RNA (Beauchemin et al., 1989a,b). Moreover, mouse CEA-like proteins were detected with polyclonal anti-CEA antibodies in protein samples prepared from normal adult colon (Beauchemin et al., 1989a). Together, these results demonstrated the existence of mouse CEA analogues.

Since then, nine similar but distinct cDNA isoforms of one mouse CEA-related gene have been identified and characterized from mouse colon and liver cDNA libraries (Turbide et al., 1991; Dveksler et al., 1991; McCuaig et al., 1992; Chapter 2, McCuaig et al., 1993). These CEA-related isoforms are the mouse homologs of human BGP as defined by their deduced domain organization, their expression patterns and their mode of membrane anchorage (Rojas et al., 1990; Barnett et al., 1993; McCuaig et al., 1993). All the mouse *Bgp* cDNAs code for polypeptides which contain a leader peptide of 34 amino acid, a 108 amino acid N-terminal domain, one or three C2-set Ig domains, followed by a transmembrane domain and either a short (10 amino acid) or a long (73 amino acid) IT domain. The cloning of several of these cDNA isoforms as well as the structures and expression patterns of mouse *Bgp* cDNA isoforms are described in detail in Chapter Two. These data strongly suggest that the nine distinct *Bgp* cDNA isoforms

are generated through complex alternative splicing of a single *Bgp* gene and through allelic variation.

We have previously demonstrated that two of the known isoforms, BgpA and BgpB, function as intercellular adhesion molecules *in vitro* (Turbide et al., 1991; McCuaig et al., 1992), much like human CEA, NCA and BGP (Benchimol et al., 1989; Zhou et al., 1990; Rojas et al., 1990; Oikawa et al., 1992). Mouse NIH 3T3 transfectant cells expressing BgpA proteins on their surface mediated homophilic intercellular adhesion in a Ca^{2+} - and temperature-independent manner; this adhesion was specifically inhibited by anti-mouse Bgp antibodies (McCuaig et al., 1992). Transfectant L-cells expressing BgpB proteins were shown to exhibit Ca^{2+} - and temperature-dependent intercellular adhesion properties (Turbide et al., 1991).

The deduced amino acid sequences of mouse Bgp isoforms are also highly homologous to several well characterized rat hepatocyte proteins including the cell-CAM 105/ecto-ATPase/pp120/HA4 proteins (Aurivillius et al., 1990; Margolis et al., 1990; Lin and Guidotti, 1989) which have been shown by antibody cross-reactivity assays and by sequence analyses of their cDNAs to be equivalent to one another (Hixson and McEntire, 1989; Margolis et al., 1990; Mowery and Hixson, 1991). Cell-CAM 105 has been characterized as a Ca^{2+} -independent cell adhesion molecule that mediates hepatocyte aggregation *in vitro* (Ocklind and Obrink, 1982). The amino acid sequences of the long IT domains of the human, rat and mouse BGP proteins are highly homologous (Hinoda et al., 1988; Lin and Guidotti, 1989; McCuaig et al., 1993) suggesting that this domain may be essential for certain function(s). The human and rat

BGP long IT domain have been reported to be phosphorylated on tyrosine residues (Rees-Jones and Taylor, 1985; Afar et al., 1992). The rat Bgp long IT domain has been implicated in the transport of bile salts through the canalicular domain of hepatocytes (Sippel et al., 1993) and in phosphorylation events activated by the insulin receptor tyrosine kinase (Margolis et al., 1990), and in binding calmodulin (Blikstad et al., 1992). Rat Bgp has also been shown to exhibit ecto-ATPase activity (Lin and Guidotti, 1989). While the exact role of mouse Bgp proteins in normal cellular physiology is unknown, a number of different functions are associated with human and murine BGP isoforms which suggest that these proteins may be involved in a variety of cellular functions.

Mouse Biliary Glycoprotein Expression Patterns

Mouse biliary glycoproteins are actively transcribed and translated in many normal adult tissues. Bgp proteins are expressed, like other CEA subgroup members, in cells of epithelial origin (Huang et al., 1990). Immunostaining experiments revealed that Bgp proteins are specifically localized to the luminal membrane of glandular epithelia in a variety of tissues (Huang et al., 1990). Other experiments demonstrated that *Bgp* transcripts are highly expressed in colon and liver and to a lesser extent in uterus, ovary, gallbladder, pancreas, and kidney (Turbide et al., 1991; McCuaig et al., 1992). These results were confirmed by western analyses using polyclonal anti-mouse Bgp antibodies which demonstrated that Bgp proteins are abundantly expressed in normal mouse colon and liver (McCuaig et al., 1992).

The expression of mouse *Bgp* during embryonic development has been assessed

by Northern analyses and *in situ* hybridization (Huang et al., 1990). *Bgp* transcripts are expressed in the mouse embryo from day 10.5 post coitum (p.c.) to birth in a spatial and temporal fashion. Northern analyses of RNA samples prepared from intestine of mouse fetuses of 16.5 days p.c. to birth and from adult mice showed high expression of *Bgp* transcripts. *Bgp* is also actively transcribed in tissues derived from mesenchymal and migratory neural cells (neuroectodermal cells) such as meninges, cartilage, bone, blood vessel walls, and bronchioles at various stages during the development of the mouse fetus (Huang et al., 1990). The pattern of *Bgp* expression *in vivo* is consistent with its role as an intercellular adhesion molecule and further suggests that *Bgp* may play an important role during active morphogenesis.

Because CEA expression is upregulated in numerous human cancers, it was important in the development of our mouse model to evaluate the expression of the mouse *Bgp* gene, a CEA-related gene, in normal and malignant tissues. The results of these experiments are described in detail in Chapter Three. Our data reveal that mouse *Bgp* transcript and protein expression is markedly decreased in primary colon carcinomas and colon and liver carcinoma cell lines in comparison to their normal tissue counterparts. Furthermore, this decrease in expression has been correlated with alterations in DNA methylation of the mouse *Bgp* gene upstream regulatory regions.

CHAPTER TWO

EXPRESSION OF THE *Bgp* GENE AND CHARACTERIZATION OF MOUSE COLON BILIARY GLYCOPROTEIN ISOFORMS

A manuscript by Kimberly McCuaig, **Madelaine Rosenberg**, Patrick Nédellec, Claire Turbide and Nicole Beauchemin, published in the journal *Gene* 127: 173-183, 1993.

SUMMARY

The biliary glycoprotein (BGP)-encoding gene is a member of the human carcinoembryonic antigen (CEA) gene family. We have now cloned several mouse *Bgp* cDNAs from an outbred CD^R-1 mouse colon cDNA library as well as by reverse-transcription-PCR amplification of colon RNA. The distinguishing features of the deduced Bgp protein isoforms are found in the two divergent N-terminal domains, the highly conserved internal C2-set immunoglobulin domains, and an intracytoplasmic domain of either 10 or 73 amino acids (aa). The cDNA structures suggest that these mRNAs are produced through alternative splicing of a *Bgp* gene and the usage of multiple transcriptional terminators. The Bgp deduced aa sequences are highly homologous to several well characterized rat hepatocyte proteins such as the cell-CAM 105/ecto-ATPase/pp120/HA4 proteins. Oligodeoxyribonucleotide probes representing the various cDNA isoform domains revealed predominant transcripts of 1.8, 3.1 and 4.0 kb on Northern analyses of mouse colon RNA; some of these bands are actually composed of several co-migrating transcripts. The transcripts encoding the long intracytoplasmic-tailed Bgp proteins are expressed at one-tenth the relative abundance of the shorter-tailed species. We have previously demonstrated that several mouse *Bgp* cDNAs, when transfected into eukaryotic cells, express Bgp proteins at the cell surface and function *in vitro* as cell adhesion molecules, much like their human and rat counterparts. The expression of the many Bgp isoforms at the surface of epithelial cells, such as colon, suggests that these proteins play a determinant role, through self- or heterologous

contact, in renewal and/or differentiation of their epithelia.

INTRODUCTION

Carcinoembryonic antigen (CEA) (Gold and Freedman, 1965) is a human tumor marker used to evaluate recurrences of gastrointestinal, breast and lung cancers (Shuster et al., 1980). The CEA gene family, composed of at least 22 genes clustered on human chromosome 19q13.1-q13.2, is divided into two subgroups based on sequence comparisons and expression patterns (Thompson et al., 1992; Khan et al., 1992): the CEA subgroup (*CEA*, *NCA*, *BGP*, *CGM6*) and a number of genes with as yet undefined gene products (designated *CGMs*; Thompson et al., 1992) and the pregnancy-specific (*PSG*) subgroup (Khan et al., 1992). The *CEA*-related gene products exhibit structural features of the immunoglobulin (Ig) supergene family with N-terminal domains resembling variable regions and internal domains homologous to C2-set constant domains (Williams and Barclay, 1988). BGPs (Svenberg, 1976; Svenberg et al., 1979; Hinoda et al., 1988; Barnett et al., 1989) are unique in this family in that all isoforms, produced through alternative splicing of a single gene (Barnett et al., 1989), bear either short (10 aa) or long (71 aa) intracytoplasmic tails (Cyts) (Hinoda et al., 1988; Barnett et al., 1989). In contrast, *CEA*, *NCA* and *CGM6* are linked to the membrane by a glycopospholipid anchor (Hefta et al., 1988; Kolbinger et al., 1989; Berling et al., 1990).

CEA, *NCA* and *BGP* are postulated to participate in intestinal tissue organization

during embryonic development and function *in vitro* as intercellular adhesion molecules (Benchimol et al., 1989; Oikawa et al., 1989; Rojas et al., 1990; Zhou et al., 1990). The CEA and NCA glycoproteins are also involved in bacterial recognition and in colonization (Leusch et al., 1991). PSGs are postulated to act as immunomodulators during pregnancy (Cerni et al., 1977).

To pursue functional investigations, we have begun characterization of the mouse CEA-related gene family. In previous work, we have shown that mouse Bgp proteins (the cDNAs had previously been called mmCGM1a and mmCGM2 and have been renamed *BgpA* and *BgpB*) expressed at the surface of transfectant cells function *in vitro* as adhesion molecules (Turbide et al., 1991; McCuaig et al., 1992). Moreover, Dveksler et al. (1991) have demonstrated that one of these mouse Bgp proteins (MHVR with identical coding sequence to the deduced aa sequence of BgpA) acts as the mouse hepatitis virus (MHV) receptor (Dveksler et al., 1991).

Contrary to human BGP isoforms which contain the same Ig-like N-terminal domain (Barnett et al., 1989), the previously reported mouse *Bgp*-like cDNAs each encode a distinct N-terminal domain, and internal C2-set domains that are conserved but not identical (Turbide et al., 1991; McCuaig et al., 1992). This finding implied that either a *Bgp* gene subfamily exists, or that complex transcriptional mechanisms are operative to account for these *Bgp* cDNAs. It therefore became important to define the number of *Bgp* isoforms and to study their expression in mouse tissues for future functional studies. The data presented in this paper describes the characterization, the relative abundance and the differential expression of nine *Bgp* cDNA isoforms expressed

in colon. The results suggest that a unique mouse *Bgp* gene undergoes alternative splicing, and that different polyadenylation signals are used to produce a number of related proteins with possible complementary functions. However, the two classes of related *Bgp* cDNAs cloned from the outbred CD-1 mouse may represent allelic variants of this *Bgp* gene (Dveksler et al., 1993).

RESULTS AND DISCUSSION

Library cDNA Clones

To characterize the mouse CEA-related gene family, we screened a CD-1 mouse colon cDNA library (Beauchemin et al., 1989). This yielded a great number of cDNA clones (i.e., 392), some of which presented different restriction patterns or different cDNA sequences. To verify if these cDNA clones were representative of other mouse CEA-related genes, clones 23, 32, 37, 58, 64 were completely sequenced while clone 132 was partially sequenced. As shown in Fig. 1, some of these clones (clones 23 and 132) overlapped perfectly within sequences encoding either the signal sequence, the N-terminal or the A2 domain of the *BgpA* cDNA (formerly mmCGM1a) (McCuaig et al., 1992), while others (clones 23, 32, 37, 58, 64) were aligned with DNA coding for either the leader, the N-terminal, the A2, the TM domains, the Cyt or the 3'UTR sequence of another published cDNA clone, *BgpB* (formerly mmCGM2) (Turbide et al., 1991). Clone 23 extended the published *BgpB* cDNA 3'UTR (Turbide et al., 1991) by 119 nt and ended in a poly(A) tail. Another cDNA clone (clone 58) encoded a longer Cyt (73

aa) than the previously reported 10 aa cytoplasmic domain (Turbide et al., 1991; McCuaig et al., 1992); this new mouse Cyt domain, however, resembled closely the cDNA sequence of human *BGP* (Barnett et al., 1989) and of a rat ecto-ATPase cDNA (Lin and Guidotti, 1989). Clones 32 and 64 also encoded conserved but not identical A1 and B1 domains when compared to similar domains of the *BgpA* cDNA. cDNA clone 37 demonstrated high homology to the reported 3'UTR of the MHVR isolated from Balb/c mice (Dveksler et al., 1991), but extended this sequence by 199 nt to another poly(A) tail.

Reverse Transcription-PCR Amplification

The information from the cDNA sequences and computer-translated aa sequences compiled from the cDNA clones were indicative of multiple *Bgp* cDNA isoforms present in mouse colon. We therefore conducted a series of experiments using a RT-PCR amplification technology (Frohman et al., 1988) to identify the type of *Bgp* isoforms present in this tissue. Total RNA from CD-1 mouse colon was reverse-transcribed using either a (dT)₁₇ adaptor primer or KM5, a primer located in the 3'UTR of *BgpB* and *BgpD* (see Fig. 3b). PCR amplifications were then performed on single stranded cDNAs with combinations of primers located in the coding and non-coding regions and the products were cloned and sequenced (Fig. 2a). Since we had observed that major sequence differences were found between the N-terminal domains (designated N1 and N2; Turbide et al., 1991; McCuaig et al., 1992), divergent primers within these two N-terminal domains (46N1, R46N1, CGM2N and RCGM2N) were synthesized, tested at

medium and high hybridization stringency and demonstrated no cross-reactivity (data not shown). The coding sequence of the two A1 domains (A1a, A1b), B1 domains (B1a, B1b) and the two A2 domains (A2a, A2b) differed by only 4, 1 and 3 nt, respectively (Fig. 3b). Common primers specific for each of these domains (KM7 and KM8 for A1 and 33-35 for A2) as well as a primer specific to the region encoding the long Cyt (RIT) were synthesized.

The use of primers specific to the 5' *UTR* of *BgpA* (KM2), and to the N-terminal domains (46N1 or RCGM2N) resulted in the amplification of a 360-bp fragment only when KM2 was combined with a primer from the N1-terminal domain (46N1). This result indicates that the 5' *UTR* of the transcripts containing the N2 domain is different in the region of KM2 (Fig. 2a). Varying the Mg^{2+} concentrations of the amplification reaction for the KM2-RCGM2N primer pair did not produce a fragment. The identity of the KM2-46N1 fragment was confirmed by sequencing of the fragment.

The sequences of the cDNA clones suggested that splicing events occurred on the primary *Bgp* transcript(s), resulting in the deletion of the A1 and B1 domains (McCuaig et al., 1992), or in the inclusion of an exon to generate a long Cyt protein. To ascertain the numerous splicing possibilities downstream of the N-terminal domains, amplifications were carried out between N-terminal primers (R46N1 or CGM2N) and oligonucleotides corresponding to the A2 domain (33-35), the long Cyt (RIT) or one of the 3' *UTRs* (KM6). As is summarized in Fig. 2a, amplification between primers R46N1 and 33-35 or CGM2N and 33-35 resulted in the synthesis of two different-sized fragments in each case. Cloning and sequencing revealed that the shorter-sized fragments were derived

from cDNAs where the N-terminal coding domains, N1 or N2, were followed by their respective A2 domains (N1 domain with A2a, N2 domain with A2b). None of the sequenced clones spliced either the N1 domain to the A2b domain, or the N2 domain to the A2a domain. Sequencing of the longer fragments cloned from each amplification demonstrated similar specificity; the N1 domain only associated with the A1a and B1a domains followed by the A2a domain while the N2 domain was joined to the A1b, B1b and A2b domains.

Similarly, amplifications using the N-terminal domain and the long Cyt-specific primers produced two DNA fragments. Sequencing confirmed that the shortest ones carried a deletion of the A1 and B1 domains, while the longer fragments included these two domains. Amplifications conducted with primers located in the N-terminal coding domains and in the 3' *UTR* (KM6) confirmed the previous results. A greater number of clones (seven) representing these fragments were sequenced, one of which was shown to code for a long Cyt domain. Results from these amplifications were confirmed by Southern analyses. As can be seen in Fig. 2b, the 1350 and 1450-bp fragments amplified by either pairs (R46N1-KM6 or CGM2N-KM6) hybridized to an oligo specific to the A1 domains (KM7) (Fig. 2b, panel A), while all the fragments hybridized to the 33-35 oligo defining the A2 domain (Fig. 2b, panel B). The bands hybridizing to oligo RIT (Fig. 2b, panel C) confirmed that the short-tailed cDNA isoforms have a counterpart encoding a long Cyt. These hybridizations also revealed a new 3.0-kb DNA fragment specifically amplified in primer combinations containing the N1 primer. This novel cDNA fragment will be the subject of future reports. This fragment is not due to contamination with

genomic DNA since it was not present in PCR amplifications when the RT reaction was omitted (data not shown).

The Nucleotide Sequences of the cDNAs and the Amino Acid Sequences

The structure of the deduced isoforms are depicted in Fig. 3a. A minimum of nine cDNA isoforms are generated from the mouse *Bgp* gene(s). Two of these cDNA isoforms (*BgpA* and *MHVR*) have an identical coding region but somewhat different 5' and 3' *UTRs* (the first 41 nt upstream of the ATG start codon are identical) (Dveksler et al., 1991; McCuaig et al., 1992). As was demonstrated by Southern analysis (Fig. 2b, panel C) and sequencing of the PCR products, cDNAs which lack a region encoding two C2-set domains such as *BgpB* and *BgpC* (clones 1 and 23 respectively) can also encode a long Cyt to generate the *BgpG* and *BgpH* cDNA isoforms. Likewise, the Southern and sequencing analyses indicate that the regions encoding the long Cyt as well as three internal repeats are present in the two longest cDNA isoforms designated *BgpD* and *BgpF* (Fig. 3b).

The *Bgp* isoforms exhibit two variants of the signal sequence: four-nt substitutions (Fig. 3b) lead to changes in two aa (Fig. 3c). The leader region is followed by an N-terminal domain. There are two possible N-terminal domains (N1 and N2) which exhibit both common and divergent areas: the nt sequences encoding the first 37 aa are identical except for one nt substitution, while the last two thirds of these domains exhibit numerous nt differences, many of which are located in the third position of a codon (Fig. 3b and 3c). As has been noted in the human *CEA* gene (Schrewe et al., 1990), the last

nt of the N-terminal domain exon is paired with the first two nts of the following exon to form the first codon of the next domain. Because of this splicing position, different aa are encoded by the various *Bgp* isoforms, depending on whether the N-terminal domain is followed by the A1 or the A2 domain. If there is an A1 domain, the first aa in A1 is Pro (*BgpA/D/E/F*) and in A2, Glu. If there is no A1 domain, the first aa in A2 is a Gln (*BgpB/C/G/H*) (Fig. 3c). There are two variants for each of the internal C2-set Ig domains; however, there are few nt changes (Fig. 3b) and those that do occur lead to aa substitutions (Fig. 3c), and in one case, a silent mutation (indicated by arrow in Fig. 3c). A linker region (Li) joining the A2 domain to the TM region is predicted by hydropathy computer analysis.

Two Cyt domains are derived from essentially the same sequence of DNA. As reported for human BGP (Barnett et al., 1989), the mouse long Cyt is generated by the inclusion of a 53 bp exon (nts 1464-1517 in Fig. 3b) which shifts the ORF from what would otherwise encode a short cytoplasmic tail. The translated proteins with a short Cyt are produced by using the first in frame stop codon (1533-1535) while proteins bearing a long Cyt are generated by the use a stop codon at nts 1667-1669 due to a the shift in reading frame. These stop codons are followed by a 3'UTR of 1.25 (CytL) or 1.39 (CytS) kb (Figs. 1 and 3).

The UTRs of these cDNAs are also variable. cDNA clone 23 contains a 103-bp 5'UTR. However, a longer 5'UTR has been demonstrated by primer extension analysis (Nédellec and Beauchemin, unpublished results). As indicated above, *BgpA* and *MIIVR* share the first 41 nts upstream of the ATG start codon but diverge further upstream of

this point, suggesting that there are a minimum of two 5' *UTR* exons, or that a more complicated exon pattern lies upstream of the signal sequence. Sequencing of cDNA library clones and PCR-amplified cDNA clones have indicated a polymorphic region in the 3' *UTR* (indicated at the bottom of Fig. 3b). Since these sequence differences occur in a highly repetitive region and are from outbred CD-1 mice, it is not known presently if this variable region is due to stuttering during the RT reactions, errors during PCR amplifications, or if it represents a *bona fide* genetic polymorphism. Further downstream, cDNA library clones terminate at three different positions: clone 23 exhibits the shortest 3' *UTR* (position 2114 in Fig. 3b), the *MHVR* cDNA (Dveksler et al., 1991) ends at position 2710 (Fig. 3b) and clone 37 contains a further 199 nts ending with a poly(A) tail at nt position 2922. Several poly(A) consensus signals (double underlines in Fig. 3b) are located upstream of the various poly(A) tails, only one of which fits the perfect consensus sequence (AATAAA) (Proudfoot and Brownlee, 1976).

The predicted aa sequence of the Bgp isoforms are presented in Fig. 3c and may be sorted into two classes, defined by their N-terminal domains. The first 37 aa of these domains are identical; however, in the following stretch of 19 aa, nine residues are not conserved; charged structures replace non-polar aa and *vice versa* (K→A, M→K, F→Q). The cysteine residues of the C2-set domains, thought to be involved in intrachain disulfide bonding (Williams and Barclay, 1988), are well conserved in all isoforms (shading in Fig. 3c).

Northern Analysis and Relative Abundance of Transcript Isoforms

Northern analyses were performed on mouse colon transcripts with oligos representing different domains of the cDNAs (Fig. 4a and b) to quantify the relative abundance of the *Bgp* isoforms (Fig. 4c). Three major transcripts were identified with these probes. These were measured as 1.8, 3.1 and 4.0 kb relative to rRNA markers. In comparison, human *BGP* restriction fragments has been shown to hybridize to 2.2 and 3.9-kb transcripts from a variety of cell lines (Hinoda et al., 1988; Barnett et al., 1989). The nt sequences of *BgpD* and *BgpF* isoforms reported in Fig. 3b are 2922 nts long, while the *BgpB* and *BgpC* or *BgpG* and *BgpH* isoforms are 1518 or 1569 nts long, respectively. The size discrepancy between the cDNA sequences and the transcript lengths measured by Northern analysis may be due to an additional 300 nts at the 5' end (Nédellec and Beauchemin, unpublished results). As well, the length of the poly(A) tails are thought to be 400-bp longer than the DNA sequences of the human *CEA* gene family members (Beauchemin et al., 1987). This adjustment would be consistent with a size of 3622-bp for the longer cDNAs. The existence of an even longer 3' *UTR* has not been ruled out. The 4.0-kb transcripts and a strong hybridization signal at 3.1-kb with a relative abundance of 1.5 are detected with oligo **KM6**; the 3.1-kb length is compatible with cDNAs expressing L-N-A2-TM-CytS or CytL carrying a complete 3' *UTR* region.

To quantify the number of transcripts co-migrating in the revealed bands, equal pmoles of the labelled oligos were added to hybridization mixtures to ensure that the intensity of the bands would reflect the relative abundance of the transcripts. The most striking observation was that, although the cDNAs encoding the isoforms containing only

two Ig domains (N and A2) could be readily cloned from a cDNA library or amplified by PCR, their abundance as 1.8-kb transcripts was minimal (Fig. 4c: **46N1**, **RCGMN2** and **KM7**). The N1 and N2-terminal domains are preferentially expressed with three Ig C2-set constant domains as full-length transcripts. But, the transcripts encoding proteins with long Cyt are preferentially expressed without A1 and B1 domains and migrate as 1.8-kb rather than 4.0-kb. These are 1/10 to 1/3 as abundant as the short Cyt-encoding isoforms. As expected, hybridization with an oligo (3' *UTR*-2) located in the distal region of the 3' *UTR* detects preferentially 4.0-kb transcripts.

An oligo specific for the A2 domain (**33-35**) binds to both the 1.8 and 4.0-kb transcripts; however, quantification of the autoradiograms at different exposure times indicated exceedingly high values for this domain. Although the migration of these transcripts is similar to the 18S rRNA, the signal seen does not represent cross-hybridization to rRNA; when this oligo was used to probe non *Bgp*-expressing cell lines, no signal was encountered. However, this A2 domain is very conserved when compared to a similar domain expressed in the mouse *Psg* cDNAs (Kodelja et al., 1989; Rebstock et al., 1990; Rudert et al., 1992). Thus, the oligo **33-35** may be detecting some of these *Psg*-encoding transcripts.

CONCLUSIONS

(1) The data presented in this report emphasize the multiplicity of the mouse *Bgp* isoforms and suggest mechanisms responsible for their generation. So far, nine similar,

but not identical cDNA isoforms have been cloned from CD-1 mouse colon (Turbide et al., 1991; McCuaig et al., 1992; this report) and from Balb/c liver (Dveksler et al., 1991). Alternative splicing is active in the generation of isoforms exhibiting different Cyt domains as well as isoforms lacking internal A1 and B1 domains. Northern analysis performed on mouse colon RNA with specific oligo probes and quantification of the autoradiograms have confirmed the diversity of the *Bgp* isoforms; three major transcripts (1.8, 3.1 and 4.0-kb) encode these isoforms, albeit in different relative amounts. The relative abundance of the *Bgp* long Cyt-containing transcripts are approximately tenfold less abundant than the short Cyt-tailed isoforms. This finding may reflect on the function of the long-tailed *Bgp* proteins since they contain consensus Tyr phosphorylation sites that have been conserved throughout evolution and are present in the mouse, the rat and the human BGP counterparts (Lin et al., 1989; Afar et al., 1992; Culic et al., 1992). Overexpression of these isoforms may be detrimental to cell growth and/or differentiation.

(2) In contrast to what has been described for human *BGP* isoforms (Barnett et al., 1989) there are, in fact, two classes of mouse *Bgp* isoforms. Sequence analyses performed on the cDNA isoforms from CD-1 mice have demonstrated that the N1-terminal domain is only associated with either the A1a or A2a domain, while the N2-terminal domain is joined to the A1b or A2b domain. These C2-set Ig domains differ from each other by only a few nucleotides while the N-terminal domains exhibit a greater number of changes. The A2 domain is also a required structural feature of the mouse

colon *Bgp* isoforms contrary to the human *BGPb* and *BGPd* isoforms (Barnett et al., 1989) which lack this domain as do *BGPg*, *BGPb*, *BGPx* and *BGPp* (T. Barnett, personal communication). However, all isoforms examined in this report carry identical TM and long or short Cyt domains. The length of the 3'UTR varies but the sequence within remains invariable (except for a few nt modifications in a highly repetitive region), strongly suggesting that a single *Bgp* gene encompasses all exons.

However, the generation of isoforms bearing a few nt modifications could also be explained by allelic variation. The present characterization was done on the outbred CD-1 mouse. A recent report suggests that expression of the two classes of isoforms (defined by their N-terminal domains) are due to allelic variation of the same gene. Balb/c mice do not express the N2-containing cDNA isoforms and conversely, SJL/J mice lack the N1-bearing isoform transcripts and proteins (Dveksler et al., 1993). The splicing versus allelic variation issue will be resolved with the cloning of the complete Balb/c *Bgp* gene (P. Nédellec and N. Beauchemin, in progress).

(3) The feature which renders the two isoform classes distinct lies in their N-terminal domains. The first 37 aa of these domains are identical, while the last third encloses many modifications mostly concentrated in a central core. Recently, Bates et al. have constructed a three-dimensional model for human CEA based on the NMR structure of the rat CD2 as well as X-ray crystal structures of human CD4 and Ig variable domains (Bates et al., 1992). The first two domains of CEA are predicted to have a rod-like appearance with numerous tightly packed β sheets and a few exposed loops on either side

of the molecule. Interestingly, the most divergent aa residues (positions 72-90) within the mouse Bgp N-terminal domains would be localized in the C-C' loop which is predicted to be exposed. On the other hand, some of the conserved 37 aa are positioned within another exposed loop. These loops are the most likely targets for interactions with other cells or ligands. In fact, two described properties of these mouse glycoproteins are dependent upon such interactions; two protein products of the *Bgp* cDNA isoforms (BgpA and BgpB) can function *in vitro* as cell adhesion molecules (Turbide et al., 1991; McCuaig et al., 1992) and one (MHVR or BgpA) has also been shown to bind to MHV spike glycoprotein (Williams et al., 1990; Dveksler et al., 1991). Although other domains of the CEA/BGP proteins are necessary for intercellular adhesion, the N-terminal domain appears to be critical for this function (Oikawa et al., 1991). Similarly, other members of the Ig supergene family (CD4, ICAM-1, CD2, poliovirus receptor) require that their N-terminal domains enter into contact with either viral envelope glycoproteins (Greve et al., 1989; Mendelsohn et al., 1989; Staunton et al., 1989; White and Littman, 1989; Koike et al., 1990; Pollard et al., 1991) or their respective ligands to be functionally activated (Staunton et al., 1990; Koike et al., 1991; Register et al., 1991).

Human *BGP* mRNAs and proteins appear to be less abundant than CEA in colon but more abundant in liver (Hinoda et al., 1990; Drzeniek et al., 1991). Extensive expression analyses of these two protein entities have been hampered by the lack of specific anti-BGP antibodies which have only recently been developed (Drzeniek et al., 1991). However, the mouse *Bgp* mRNAs (Turbide et al., 1991; McCuaig et al., 1992)

and proteins (Rosenberg et al., in press) are more abundant in colon than liver. Since a true CEA homolog or any of the phosphatidylinositol-linked CEA family members have yet to be identified in the mouse or rat models, it is tempting to speculate that the functions associated with CEA in the human may represent a recent evolutionary event (Stanners et al., 1992) and that these functions may be ensured in the mouse by other *Bgp*-related genes (P. Nédellec and N. Beauchemin, in preparation). The significance of the concurrent expression of the many Bgp isoforms on the surface of mouse tissues such as colon, intestine, liver and uterus (Turbide et al., 1991; McCuaig et al., 1992) may be that these proteins play a determinant role, through self- or heterologous contact, in renewal and/or differentiation of their epithelia.

Acknowledgments

The authors wish to thank Dr. Annette Herscovics and Dr. Clifford Stanners for critical reading of the manuscript and Dr. André Veillette for the gift of the NIH3T3 cells.

Figure 1. Topology of library cDNA clones. *5'UTR1* and *5'UTR*: 5' untranslated regions, *L1* and *L2*: signal sequences, *N1* and *N2*: N-terminal domains, *A1a*, *A1b*, *B1a*, *B1b*, *A2a*, *A2b*: C2-type Ig domains, *Li*: linker, *TM*: transmembrane domain, *CytS*: putative short Cyt domain, *CytL*: long Cyt domain, *3'UTR1* and *3'UTR2*: 3' untranslated regions, *An*: positions of poly(A) tails.

Methods: cDNA was synthesized as previously described (Beauchemin et al., 1989). The CD-1 mouse colon cDNA library was screened with cDNA restriction fragments corresponding to the *EcoRI-SstI* and *AccI-EcoRI* restriction fragment of clone 46 (McCuaig et al., 1992) and with the oligo RIT (antisense: 5'-TGAGGGTTTGTGCTCTGTGAGATC) representing a region of the long Cyt conserved between human BGP (Barnett et al., 1989) and rat ecto-ATPase (Lin and Guidotti, 1989). *Bgp* cDNA restriction fragments were subcloned into unique sites of the Bluescript SK+ plasmid (Stratagene, La Jolla, CA) and overlapping DNA restriction fragments were sequenced on both strands with either the T7 or the T3 promoter primers or with internal primers by the dideoxy chain termination method (Sanger et al., 1977) using T7 DNA polymerase (Pharmacia, Montreal, CANADA). Sequences were analyzed using the DNAsis, Prosis (Pharmacia) and the GCG sequence analysis system (Devereux et al., 1984) programs.

Figure 1

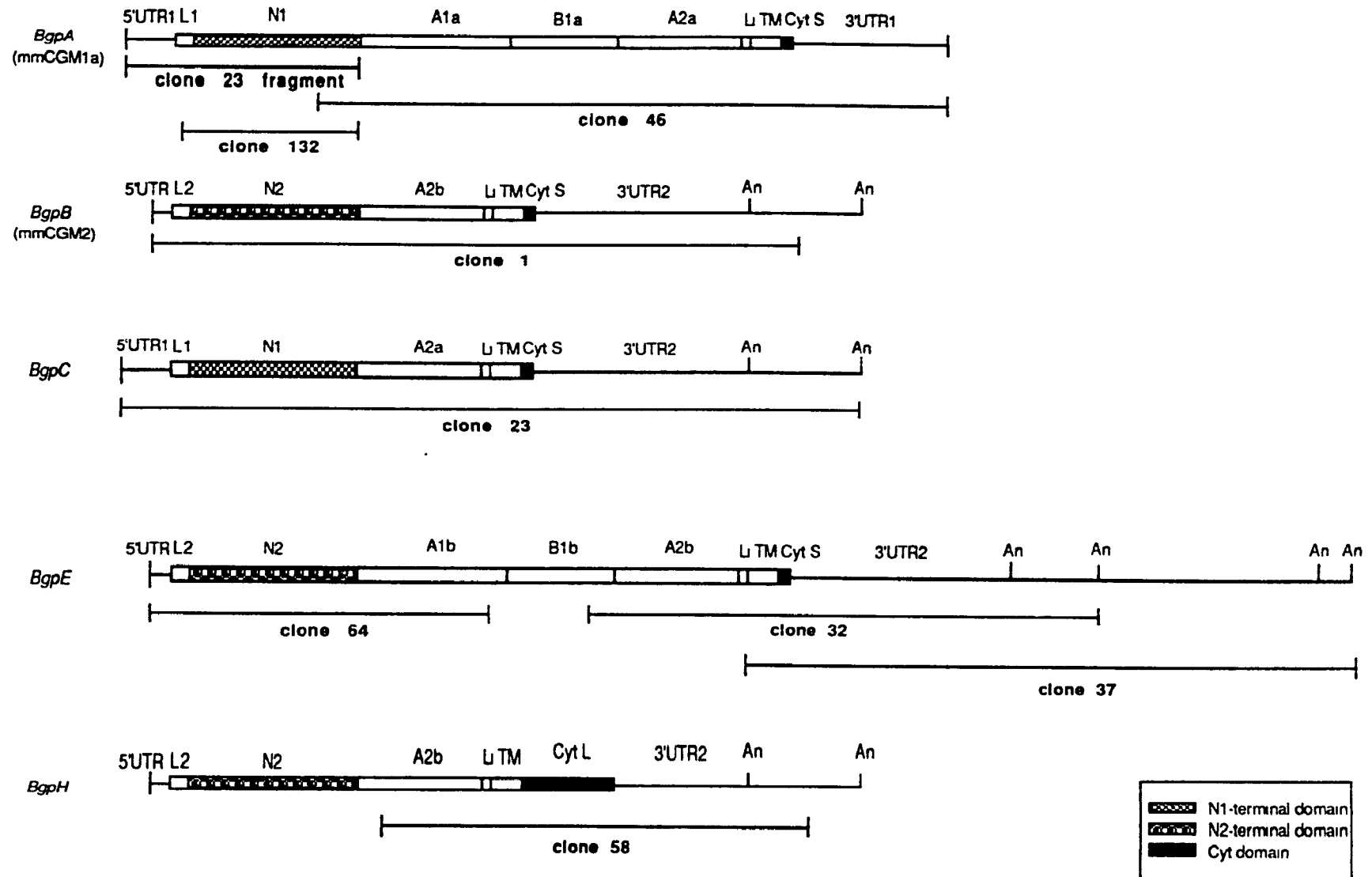


Figure 2. PCR Amplification Reactions and Hybridizations of Amplified Fragments.

(a) PCR amplification reactions: The names and positions of the various oligo primers used in the PCR amplification reactions are indicated. The numbers in parentheses represent the approximate length of the fragments obtained from each reaction.

(b) Hybridization of PCR fragments: PCR fragments were obtained from amplifications with oligos **R46N1-KM6** (lane 1) and **CGM2N-KM6** (lane 2). They were fractionated on a 1% agarose gel and hybridized with ^{32}P -labelled the oligo **KM7** (panel **A**), with oligo **33-35** (panel **B**) and with oligo **RIT** (panel **C**).

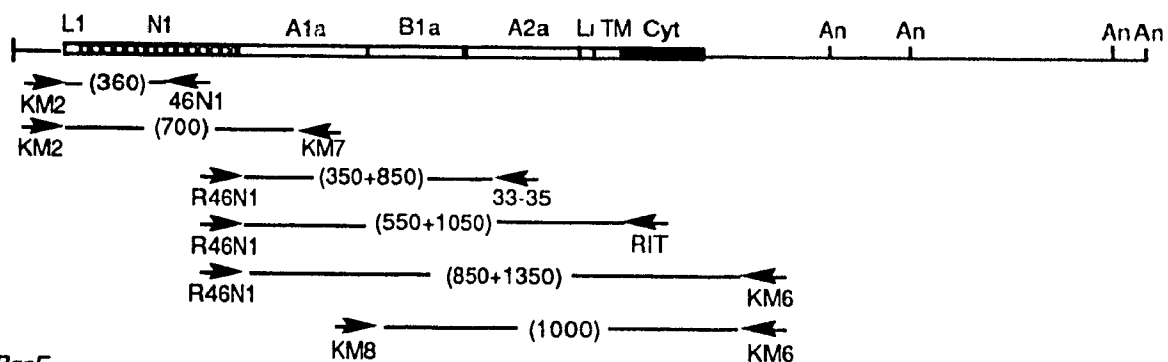
Methods: Total RNA from CD-1 mouse colon was prepared by guanidium isothiocyanate extraction and centrifugation (Beauchemin et al., 1989). RT was performed using AMV reverse transcriptase (Pharmacia, Montreal, CANADA) and 10 μg of total mouse colon RNA as template essentially as described previously (Turbide et al., 1992). The oligo primers used in this reaction were either a $(\text{dT})_{17}$ containing restriction sites (5'-GACTCGAGTCGACGGTACCCT₁₇) or oligo **KM5** (antisense: 5'-TTGATACCTCACTCTCAGCCA). The amplification reactions were incubated at an annealing temperature of 44°C for 2 min, at an elongation temperature of 72°C for 3 min and at a denaturing temperature of 94°C for 1 min for 40 cycles in a total volume of 100 μl containing 20 mM Tris.Cl pH 8.8 (at 24°C)/10 mM KCl/2 mM MgSO_4 / 10 mM ammonium sulfate/0.1% Triton X-100/0.1 μg BSA/0.2 mM of each dNTP/40 pmoles of phosphorylated primers, using Vent^R DNA polymerase (New England Biolabs), to decrease possible proofreading errors (Frohman et al., 1988; Mattila et al., 1991). The oligo primers used were: **KM2**: sense 5'-CCAAGTCCCGACAAGTAGTG;

RCGM2N: antisense 5'-GTCTTATTAGTGCCTGTTAC; CGM2N: sense 5'-GTAACAGGCACTAATAAGAC; 46N1: antisense 5'-CTTCATGGTGATCATTG; R46N1: sense 5'-CCAAATGATCACCATGAAG; KM7: antisense 5'-GGGTCAC TTCGGTTGACACT; KM8: sense 5'-CCAGTGAGTGTCAACCGAAG; 33-35: antisense 5'-CCGGCATCTTCCCTCTTAATAGGGTCTATTCTG; RIT: antisense 5'-TGAGGGTTTGTGCTCTGTGAGATC; KM6: antisense 5'-GGCTCCAGGATCCACCTTTTCTTC. The PCR amplification products were blunt-ended (Sambrook et al., 1989) with T4 DNA polymerase (Boehringer-Mannheim) to degrade single-stranded DNA resulting from non-symmetrical amplification (Innis et al., 1990), separated by electrophoresis and cloned into the *Sma*I site of the Bluescript SK+ plasmid (Stratagene) for sequencing. A minimum of three independent clones were sequenced for each fragment resulting from every amplification reaction with different pairs of primers. Control samples omitting reverse transcription were included to verify that the RNA was not contaminated by genomic DNA. These fragments were also electrophoresed through 1% agarose gels and transferred to GeneScreen Plus membranes (NEN Dupont). Hybridizations were carried out as previously described (McCuaig et al., 1992) with 2×10^6 cpm/ml of [32 P]oligos (Sambrook et al., 1989). The filters were washed to a final stringency corresponding to 5°C below the predicted melting temperature of each oligo.

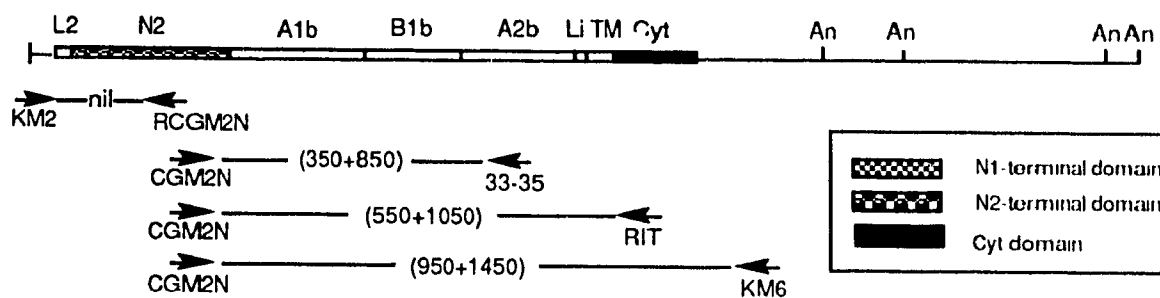
Figure 2

a

BgpD



BgpF



b

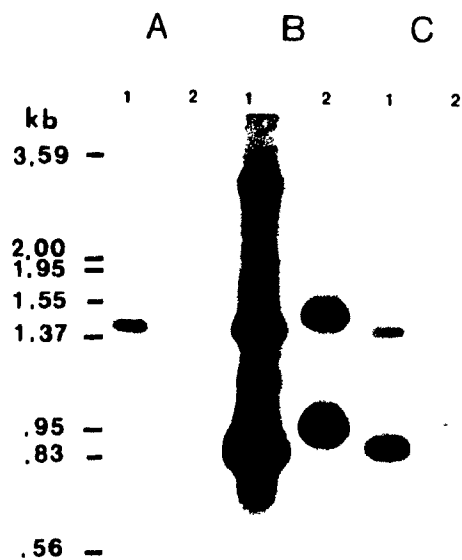


Figure 3. Topology and cDNA Sequences of Mouse *Bgp* Isoforms and Deduced aa Sequences.

(a) Topology of *Bgp* isoforms: Structures of the various cDNA isoforms. Abbreviations used have been described in Legend to Fig. 1, **MHVR**: mouse hepatitis virus receptor.

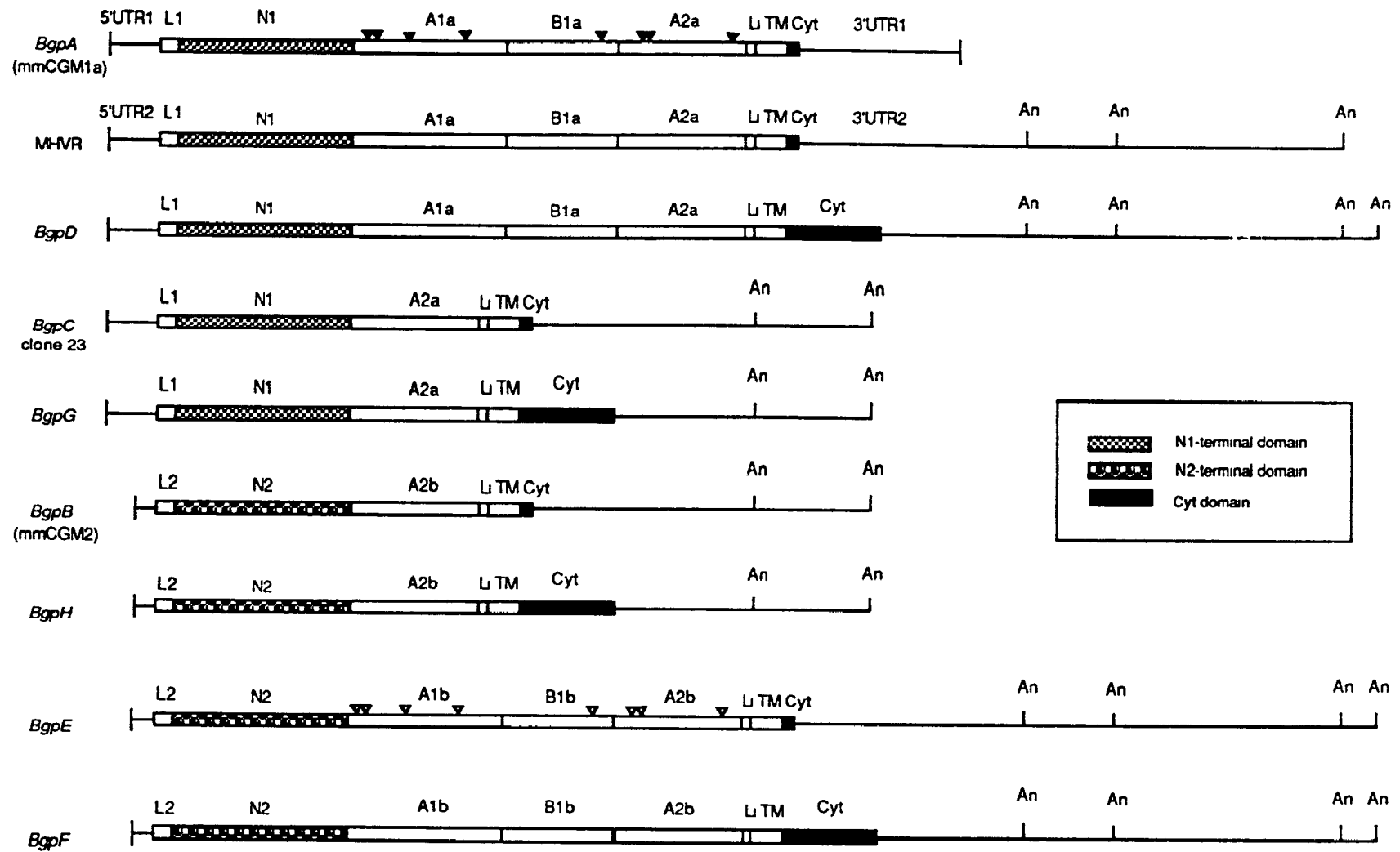
(b) cDNA sequences: The nt sequences of the longest isoforms (*BgpD* and *BgpF*) are presented. Dots indicate identical nts. The sequences of the oligos are indicated by lines above or below the nt sequences and their orientation by arrowheads on these lines. The **ATG** start and **TGA** stop codons are boxed. The domains are identified by thick lines and captions over and under the sequences. **Open arrowheads** delineate the 53 bp exon inserted to generate the long Cyt domain. Nt 1534-1550 are common to 3'UTR1 and 3'UTR2. The **asterisk** (at nt 1550) represents the position where the 3'UTR1 (McCuaig et al., 1992) replaces the 3'UTR2. **Double underlines** indicate putative or actual polyadenylation consensus signals. $\uparrow A_{11}$ at position 2114 represents the poly(A) tail of cDNA clone 23. $\uparrow A_{22}$ at nt 2710 represents the poly(A) tail of the *MHVR* cDNA (Dveksler et al., 1991). The poly(A) tail at nt 2908 is that of cDNA clone 37. Sequence differences between nt 1747-1829 in the 3'UTR between various clones are indicated at the bottom of the figure. GenBank accession numbers: *BgpC*: X67278, *BgpD*: X67279, *BgpE*: X67280, *BgpF*: X67281, *BgpG*: X67282, *BgpH*: X67283.

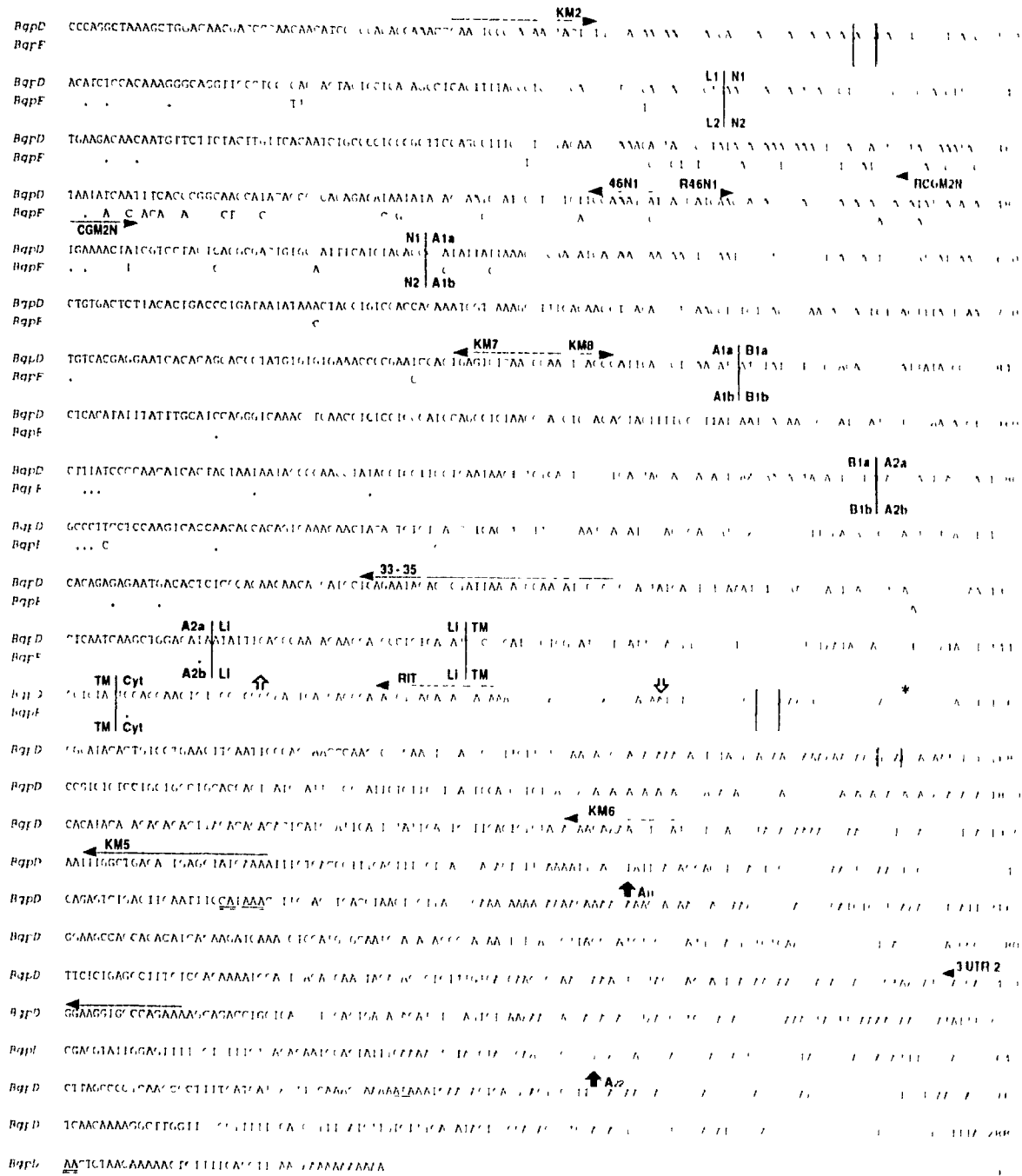
(c) Deduced aa sequences of *Bgp*: The nt sequences were computer-translated and the deduced aa sequences were aligned. **Double arrowheads** indicate delineation of the domains. The protein sequences are written in the one letter aa code. Aa positions are indicated above the signal sequences and N-terminal domains. Glycosylation sites are

underlined. Shaded C represent the conserved Cys residues. **Arrow** indicates position of the nt modification leading to a silent mutation.

Methods: The full coding cDNAs for *BgpD*, *BgpE*, *BgpF*, *BgpG*, *BgpH* were reconstructed by ligating overlapping cDNA clones at various restriction sites. Reconstructed cDNA clones were completely resequenced to ensure accuracy of sequence.

Figure 3a





	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

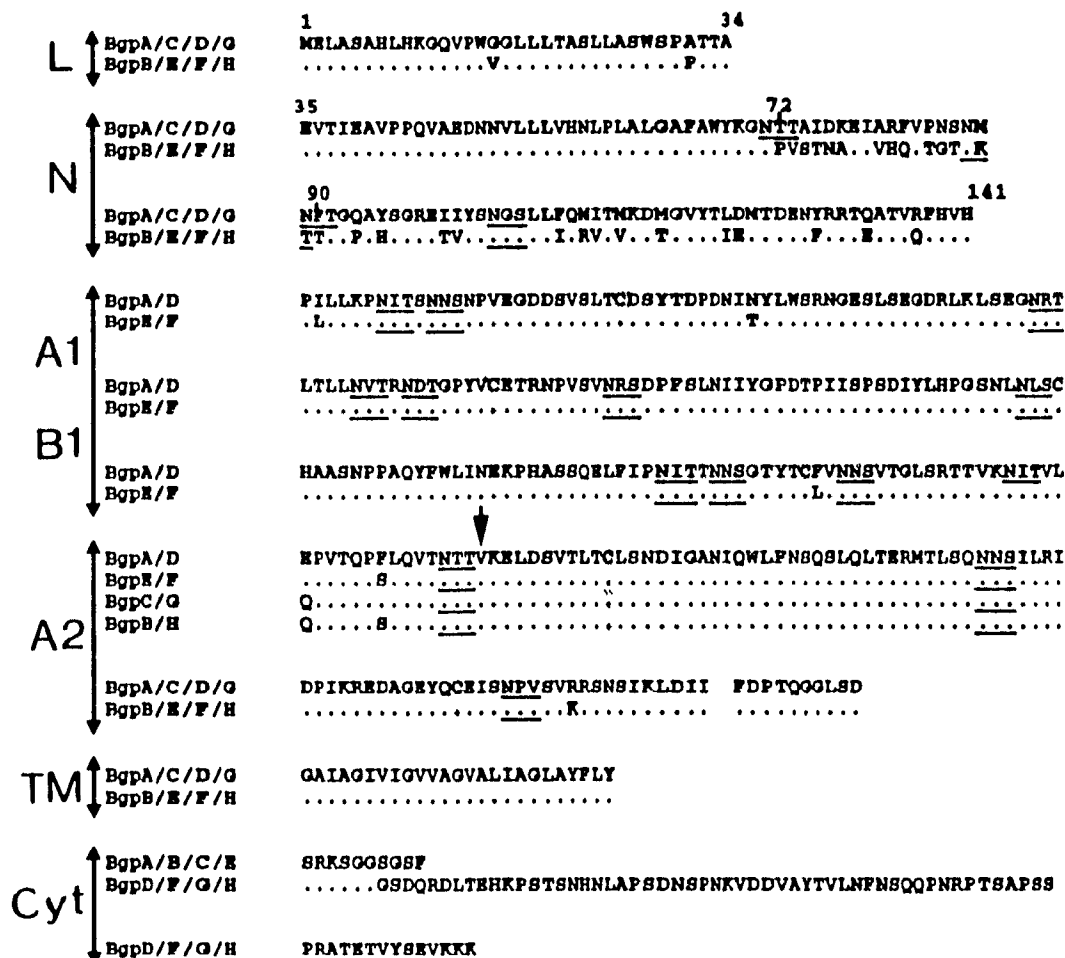
LEGEND

↓↑ 53 bp exon

* position of insertion of 3' 11112 1

↑ A $\text{poly}(A)$ ist

Figure 3c



C : conserved Cys thought to be involved in disulfide bonding

Figure 4. Northern Analyses and Quantification.

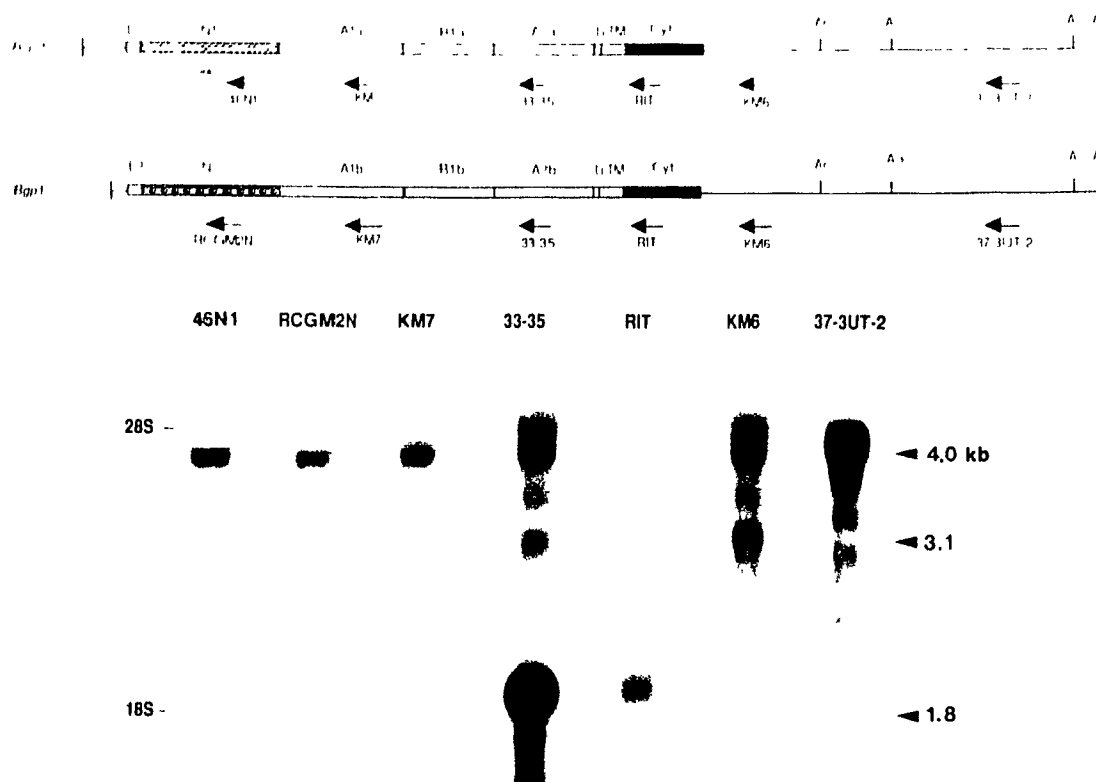
(a) Positions of oligos in the *BgpD* and *BgpF* isoforms: The positions of the oligos in the various regions of the cDNAs are indicated by arrows below the structures.

(b) Northern analyses: Positions of the 18S: 1.86 kb and 28S: 4.71 kb as well as the three major transcripts quantitated in c: are indicated.

(c) Quantification: Autoradiograms of the above Northern blots were scanned by laser densitometry using a Bio-Imager scanner (Millipore, Montreal, CANADA). The 4.0 kb band revealed by hybridization to oligo 46N1 was weighed as 1.0. All other readings were computed proportionally. n.d., not determined.

Methods: 20 μ g of total colon RNA, prepared by guanidium isothiocyanate extraction and centrifugation, was electrophoresed through a 2.2 M formaldehyde-1.5% agarose gel, transferred to GeneScreen Plus membranes. Hybridizations were carried out as previously described (McCuaig et al., 1992) with the [32 P]oligos (Sambrook et al., 1989) indicated above the autoradiograms. The filters were washed to a final stringency corresponding to 5°C below the predicted melting temperature of each oligo.

Figure 4



	46N1	RCGM2N	KM7	33-35	RIT	KM6	37-3UT-2
4.0 kb	1.0 n=7	0.40 ± 0.05 n=5	1.0 ± 1 n=6	2.5 ± 5 n=5	0.01 n=1	2.8 ± 5 n=5	2.4 n=1
3.1 kb	0.01 n=1	n.d.	0.03 n=1	0.44 ± 0.04 n=4	n.d.	1.5 ± 1 n=5	n.d.
1.8 kb	0.02 n=1	0.03 n=1	0.02 n=1	14 ± 5 n=5	0.6 ± 2 n=5	0.2 n=1	n.d.

CHAPTER THREE

THE EXPRESSION OF MOUSE BILIARY GLYCOPROTEIN, A CARCINOEMBRYONIC ANTIGEN-RELATED PROTEIN, IS DOWNREGULATED IN MALIGNANT TISSUES

A manuscript by **Madelaine Rosenberg**, Patrick Nédellec, Serge Jothy, David Fleischer, Claire Turbide and Nicole Beauchemin, accepted for publication in the journal *Cancer Research*.

ABSTRACT

Mouse biliary glycoprotein (Bgp) is a member of the carcinoembryonic antigen (CEA) gene family and is highly expressed in the epithelial cells of normal hepatic biliary ducts and intestine. Nine mouse Bgp isoforms have been identified through molecular cloning and shown to be splice and allelic variants of a single *Bgp* gene. These glycoproteins function *in vitro* as intercellular adhesion molecules and serve as the mouse hepatitis virus receptors. Since human CEA is overexpressed in gastrointestinal tumors, we have investigated the expression of mouse Bgp in primary tumors and carcinoma cell lines. Our results demonstrate that the expression of the major mouse Bgp isoforms is downregulated in tumors at the transcriptional and the posttranscriptional levels. This decrease in expression is corroborated by immunostaining of primary colonic tumors with anti-mouse Bgp antibodies. In addition, Bgp expression is influenced by transcriptional control mechanisms involving DNA methylation of the *Bgp* gene upstream regulatory region. Our results demonstrate that mouse Bgp protein expression is decreased upon malignant transformation and further suggest that Bgp proteins may be involved in the maintenance of the differentiated cellular phenotype.

INTRODUCTION

Carcinoembryonic antigen (CEA) is a heavily glycosylated protein used clinically as a human tumor marker to detect recurrences of numerous types of cancer (Gold and

Freedman, 1965; Shuster et al., 1980). CEA is the prototype of a large group of related proteins which belong to the immunoglobulin supergene family (Thompson et al., 1991). Following the cloning of the CEA cDNA in 1987, the genes of many related proteins were also cloned and are presently classified into two subgroups: the CEA-related subgroup (CEA, BGP, NCA, CGM1, CGM6, CGM7) and the pregnancy-specific subgroup (PSG1-11) (Thompson et al., 1991). These cDNAs code for proteins that display similar features where the amino-terminal domain resembles a variable immunoglobulin (Ig) domain and the internal repeating units are C2-like Ig domains (Thompson et al., 1991; Williams and Barclay, 1988). These proteins differ, however, in their cell membrane attachment mechanisms. Biliary glycoprotein (BGP) is unique in this family since it bears a transmembrane domain and an intracytoplasmic domain (Hinoda et al., 1990; Barnett et al., 1993) while CEA, NCA and CGM6 are attached to the cell membrane by glycosphospholipid anchors (Thompson et al., 1991). In contrast, PSG subgroup family members are generally secreted and usually associated with placenta, testis and choriocarcinomas (Thompson et al., 1991).

Most members of the human CEA-related subgroup are expressed at the apical membrane of epithelial cells of many normal tissues and tumor cells (Thompson et al., 1991). BGP, in particular, is expressed in normal biliary ducts of the liver (Hinoda et al., 1990), in colon (Barnett et al., 1989), in granulocytes (Drzeniek et al., 1991) and in leukocytes (Kuroki et al., 1991) as well as in a variety of tumors such as colon, pancreas, and liver carcinomas and in leukemia cell lines (Barnett et al., 1989; Hinoda et al., 1990; Drzeniek et al., 1991).

The functions of the CEA subgroup family members have been investigated. BGP, CEA and NCA are cell adhesion molecules as demonstrated by *in vitro* aggregation assays (Benchimol et al., 1989; Rojas et al., 1990; Thompson et al., 1991; Oikawa et al., 1992). BGP, CEA and NCA also bind *Escherichia coli* isolated from human gut and various strains of *Salmonella* (Leusch et al., 1991). Furthermore, CEA has been implicated in at least three of the four stages of colorectal carcinogenesis (Jessup and Thomas, 1989) and is postulated to play a role in intestinal tissue organization during development (Benchimol et al., 1989).

We have been studying a mouse model for the human CEA gene family; in previous works, we and others have characterized nine distinct cDNA isoforms of one mouse CEA-related gene (Beauchemin et al., 1989; Turbide et al., 1991; Dveksler et al., 1991; McCuaig et al., 1992; McCuaig et al., 1993; Dveksler et al., 1993). These CEA-related isoforms are the mouse homologs of human BGP as defined by their deduced amino acid sequences, their mode of membrane attachment and their expression patterns (Thompson et al., 1991; McCuaig et al., 1993). The distinguishing features of the predicted mouse Bgp protein isoforms are found in the two diverse N-terminal domains, the highly conserved C2-set Ig domains and an intracytoplasmic (IT) domain of either 10 or 73 amino acids (McCuaig et al., 1993). The nine distinct cDNA isoforms are generated through complex alternative splicing of one *Bgp* gene (McCuaig et al., 1992; McCuaig et al., 1993) and allelic variation (Dveksler et al., 1993).

We have previously demonstrated that several mouse *Bgp* cDNAs, when transfected into eukaryotic cells, express Bgp proteins at the cell surface and that these

Bgp proteins function as cell adhesion molecules *in vitro* (Turbide et al., 1991; McCuaig et al., 1992), much like their human counterpart (Rojas et al., 1990; Oikawa et al., 1992). Several of the mouse Bgp isoforms also function as mouse hepatitis viral receptors (Dveksler et al., 1991; Dveksler et al., 1993); this is the first example of multiple isoforms functioning as receptors for a single viral capsid protein (Dveksler et al., 1993). The mouse Bgp deduced amino acid sequences are also highly homologous to several well characterized rat hepatocyte proteins including the cell-CAM 105/ecto-ATPase/pp120/HA4 proteins (Aurivillius et al., 1990; Culic et al., 1992; Najjar et al., 1993). Moreover, the amino acid sequence of the long IT domains of the mouse, rat and human BGP proteins share high homology (Barnett et al., 1989; Culic et al., 1992; McCuaig et al., 1993) suggesting that this domain may be essential for certain function(s). In fact, the long IT domain has been implicated in the transport of bile salts through the canalicular domain of hepatocytes (Sippel et al., 1993), in phosphorylation events activated by the insulin receptor tyrosine kinase (Najjar et al., 1993) and protein kinase C (Afar et al., 1992), and in binding calmodulin (Blikstad et al., 1992).

The overexpression of CEA in certain human tumors has been shown to result from alterations in transcriptional and posttranscriptional mechanisms (Tran et al., 1988; Boucher et al., 1989; Hauck et al., 1991). Although rat BGP protein expression is diminished in rat hepatocellular carcinomas (Hixson et al., 1985; Hixson and McEntire, 1989), the regulatory mechanisms operative on the expression of human, rat and mouse BGP remain unknown. To determine if mouse Bgp expression is altered upon malignant transformation, we have investigated the expression of several mouse Bgp isoforms in

primary colonic tumors and in colon and liver tumor cell lines in comparison to their normal tissue counterparts.

Our results demonstrate a marked decrease in mouse *Bgp* gene transcriptional activity and a corresponding decrease in protein expression in primary colon tumors and colon and liver carcinoma cell lines. Furthermore, our results clearly indicate that *Bgp* expression is influenced by transcriptional control mechanisms that correlate with changes in the DNA methylation status of the *Bgp* gene upstream regulatory regions.

MATERIALS AND METHODS

Cell Culture and Transfections

Mouse embryonic fibroblast cells (NIH 3T3), kindly provided by Dr. André Veillette (McGill Cancer Center, McGill University, Montréal, Québec, Canada), mouse (C57BL ICRF a¹) methyl azoxymethanol acetate-induced rectal carcinoma cells CMT-93 (American Type Culture Collection (ATCC), Rockville, MD) (Franks and Hemmings, 1978), mouse (C57BL X Af F1 hybrid female) spontaneous mammary tumor cells MMT 060562 (ATCC), and the Balb/c colonic carcinoma cell lines CT 26, CT 36 and CT 51 (Brattain et al., 1980), generously provided by Dr. Michael Brattain (University of Alabama in Birmingham Medical Center, Birmingham, AL) were grown at 37°C in monolayer cultures in α -MEM medium (GIBCO, Burlington, Ontario, Canada) supplemented with 10% fetal bovine serum (GIBCO). Cells of the Balb/c embryonic liver cell line BNL CL.2 (ATCC), of the methylcholanthrene epoxide-transformed Balb/c

liver cell line BNL 1MEA.7R.1 (ATCC), of the nitrosoguanidine-transformed Balb/c liver cell line BNL 1NG A.2 (ATCC) and of the mouse (C3H/CrGl) spontaneous mammary tumor cell line Mm5MT (ATCC) were grown at 37°C in monolayer cultures in D-MEM with 4.5 g/L glucose (GIBCO) supplemented with 10% fetal bovine serum (GIBCO). All media were supplemented with 50 U/ml of penicillin and 50 µg/ml of streptomycin. NIH 3T3 transfectants, stably expressing BgpA (previously named mmCGM1a), an isoform bearing a 10 amino acid IT domain, were cultured as previously reported (McCuaig et al., 1992). Transfectant cells expressing BgpD, an isoform exhibiting a 73 amino acid IT domain, were generated by calcium phosphate-mediated coprecipitation (Parker and Stark, 1979), into 5×10^5 NIH 3T3 cells, of 5 µg of the *BgpD* cDNA (McCuaig et al., 1993) inserted in the sense or antisense orientations in the p91023B expression vector (Wong et al., 1985) along with 5 µg of carrier genomic DNA and 0.5 µg of the dominant selectable marker pSV2neo (Southern and Berg, 1982). Transfectant clones were manually picked after selection in medium containing 1.0 mg/ml of Geneticin^R powder (active form: 630 µg/mg) (G418, GIBCO) and evaluated for BgpD protein production by immunoblotting analyses. The transfectant cells were subsequently maintained in α -MEM medium containing 10% fetal bovine serum and 0.5 mg/ml of Geneticin^R (total powder).

Generation of Antibodies

A polyclonal rabbit anti-mouse Bgp antibody (serum 231) has previously been described (McCuaig et al., 1992). To generate a polyclonal antibody reacting

specifically with the long IT domain of mouse Bgp isoforms, a restriction fragment of the *BgpD* cDNA (McCuaig et al., 1993) corresponding to the long IT domain was subcloned into the pGEX2T plasmid vector (Pharmacia, Montréal, Québec, Canada). The plasmid was inserted into *E. coli* 1161 and transformed colonies were selected. Glutathione *S*-transferase-IT fusion proteins were prepared, adsorbed onto glutathione-agarose beads (Sigma, St. Louis, MO) and submitted to cleavage by 0.3 U/ml of thrombin (Boehringer Mannheim, Montréal, Québec, Canada) (Smith and Johnson, 1988). The eluted peptide was purified on 15% SDS-acrylamide gels and the appropriate ~*Mr* 8,000 band was retrieved, emulsified with either complete or incomplete Freund's adjuvant (GIBCO), and injected into pathogen-free New Zealand rabbits. The rabbits were boosted at two week intervals. This antibody, further referred to as rabbit anti-IT antibody (serum 837), reacts with mouse Bgp proteins bearing long IT domains upon immunoprecipitation but is unable to recognize these Bgp isoforms by immunoblotting.

Immunodetection Analyses

Cells were grown as monolayers to confluency, washed twice with PBS (130 mM NaCl, 7 mM Na₂HPO₄, 3 mM NaH₂PO₄, pH 7.0), harvested with a solution of PBS-citrate (PBS-15 mM sodium citrate, pH 7.0) and lysed by sonication in PBS with protease inhibitors (aprotinin 2 µg/ml, leupeptin 5 µg/ml, and pepstatin 0.4 µg /ml; Sigma and Boehringer Mannheim) at 4°C. Balb/c colon and liver membrane preparations were generated as previously described (McCuaig et al., 1992). After determination of total protein concentration (Peterson, 1977), protein aliquots were

precipitated and processed for SDS-PAGE analysis as previously described (McCuaig et al., 1992). The proteins were resolved by SDS-PAGE on 10% acrylamide gels (Laemmli, 1970), transferred to nitrocellulose membranes (Schleicher & Schuell, Keene, NH) according to Towbin (Towbin et al., 1979). Specific glycoproteins were revealed by incubation with anti-mouse Bgp antibody (serum 231) and ^{125}I -labeled protein A (Amersham Canada, Oakville, Ontario, Canada), followed by exposure to Kodak XAR-5 film (Eastman Kodak, Rochester, NY).

Immunoprecipitation

Proteins from cell lysates were immunoprecipitated for 2h at 20°C or 16h at 4°C with anti-mouse Bgp antibody (serum 231) or anti-IT antibody (serum 837). The antibody-antigen complexes were incubated with Protein A Sepharose CL-4B (Pharmacia,) for 4h at 4°C, washed five times with PBS-1% Lubrol PX (Sigma) and subsequently boiled in SDS-PAGE sample buffer for 10 min. After gel electrophoresis and immunoblotting, the immunoprecipitated proteins were detected using anti-mouse Bgp antiserum 231 as described above.

Immunocytochemistry of Tumors

Colonic tissue from CD-1 adult mice and from 1,2-dimethylhydrazine-induced colon adenocarcinomas (Glickman et al., 1987) were excised, fixed in phosphate buffered neutral 4% formaldehyde solution and embedded in paraffin. Immunohistochemical staining by a peroxidase anti-peroxidase technique was performed as described (Jothy et

al., 1986) on 4 μ m sections with a 1:100 dilution of anti-mouse Bgp antibody (serum 231), with 1:100 dilution of polyclonal anti-human CEA antibody (Dakopatts, Glostrup, Denmark) or with rabbit pre-immune serum. Antigen-antibody complexes were visualized using swine anti-rabbit IgG antibodies, followed by rabbit horseradish-conjugated peroxidase anti-peroxidase complexes. The sections were counterstained with hematoxylin.

[³²P]-Labeled Probes

A 1.0-kilobase *EcoR*I-*Sst*I restriction fragment of the *BgpA* cDNA corresponding to the signal sequence, N-terminal domain and two C2-set Ig domains (McCuaig et al., 1992), and a 1.5-kilobase *Pst*I restriction fragment of chicken β -actin cDNA (Cleveland et al., 1980) were excised from 0.8% low melting point agarose gels and labeled with [α -³²P]dATP (NEN Dupont, Mississauga, Ontario, Canada) by random priming (Feinberg and Vogelstein, 1983). A 393 base pair *Nco*I restriction fragment corresponding to a promoter fragment of the mouse *Bgp* gene and a *Bam*HI-*Hind*III restriction fragment corresponding to an upstream region of the mouse *Bgp* gene both obtained from a genomic λ phage clone containing 5-kilobases of upstream regulatory elements (Nédellec, unpublished results) were prepared and labeled as described above.

RNA Preparation and Northern Analyses

Total RNA was prepared by guanidium isothiocyanate extraction and centrifugation as previously described (McCuaig et al., 1992) from the following

samples: normal Balb/c mouse colon and liver, 1,2-dimethylhydrazine-induced primary colon tumors in CD-1 mice and their adjacent mucosae, and the cell lines CT 26, CT 36, CT 51, BNL CL.2, BNL 1MEA.7R.1, BNL 1NG A.2, MMT 060562, Mm5MT and CMT-93. RNA samples were separated by electrophoresis on 2.2 M formaldehyde-containing 1.5% agarose gels, transferred to Hybond-N membranes (Amersham Canada) and probed with radiolabeled restriction fragments as previously described (McCuaig et al., 1992). 18S rRNA (1.86-kilobases) and 28S rRNA (4.71-kilobases) were used as markers. The membranes were rehybridized with a radiolabeled β -actin cDNA fragment to confirm that equal amounts of RNA were loaded onto the gel.

Genomic DNA Analyses

Normal Balb/c spleen genomic DNA was purchased from Jackson Laboratory (Bar Harbour, MA). Genomic DNA was prepared from CT 26, CT 36, CT 51 and CMT-93 cells, and normal Balb/c colons according to established protocols (Sambrook et al., 1989). 5 μ g of genomic DNA were digested by the restriction enzymes *HindIII* (Pharmacia), *EcoRI* (Pharmacia), *BamHI* (Pharmacia), *SstI* (Pharmacia), *MspI* (New England Biolabs, Beverly, MA) and *HpaII* (New England Biolabs) according to the manufacturers' specifications. The digests were electrophoresed in 0.8% agarose gels (buffered in 40 mM Tris-HCl, 33 mM sodium acetate, 1 mM EDTA, pH 7.6) and transferred by capillarity to GeneScreen Plus membranes (NEN Dupont). The membranes were hybridized with radiolabeled restriction fragments as previously described (McCuaig et al., 1992).

RESULTS

Transcriptional Activity

To explore the control mechanisms involved in mouse *Bgp* gene expression, we first analyzed the mouse *Bgp* gene transcriptional activity in various tissues and cell lines. Northern analyses were performed with radiolabeled *BgpA EcoRI-SstI* cDNA restriction fragments (McCuaig et al., 1992) on total RNA prepared from normal mouse colon and liver, from colon, liver and breast carcinoma cell lines, and from primary colonic tumors. Multiple transcripts were detected in normal colon with predominant transcripts hybridizing at 4.0 and 1.8-kilobases (Fig. 1A, lane 2). According to our cloning data (McCuaig et al., 1992; McCuaig et al., 1993), the 4.0-kilobase transcripts correspond to mRNAs enclosing a 5'UTR of 391-ribonucleotide bases (Nédellec, unpublished results), a typical N-terminal domain, three C-2 set Ig domains (denoted as A1, B1, A2), a characteristic short or long IT domain and 1.4-kilobases of 3'UTR. The 1.8-kilobase transcripts represent splice variants in which the two C-2 set domains (A1 and B1) have been excised and in which a shorter 3'UTR is used (McCuaig et al., 1992; McCuaig et al., 1993). 4.0 and 1.8-kilobase transcripts of lesser intensity were observed in the rectal carcinoma cell line CMT-93 (Fig. 1A, lane 6). No such signals were detected in the colon carcinoma cell lines CT 26, CT 36 and CT 51 (Fig. 1A, lanes 3, 4 and 5 respectively).

The abundance of mouse *Bgp* transcripts in normal liver and liver tumor cell lines was also evaluated. The hybridization probe recognized 4.0-kilobase and 1.8-kilobase

transcripts in normal Balb/c liver (Fig. 1A, lane 1), as previously demonstrated (16-18); the amount of RNA loaded in this lane was low in comparison to other samples, as can be judged by hybridization of the β -actin probe (Fig. 1B). No similar *Bgp* mRNAs were observed in samples from the liver cell lines tested (Fig. 1A, lanes 7-9). However, upon five-fold longer exposure to film, faintly hybridizing *Bgp* transcripts of 4.0 and 1.8-kilobases were detected in samples from the normal embryonic liver cell line BNL CL.2 and from the transformed liver cell line BNL 1NG A.2 (data not shown).

Robbins *et al.* (Robbins et al., 1991) have reported that the transformation of mouse breast tissue by MMTV or chemical carcinogens leads to the overexpression of mouse *Bgp* transcripts which are not normally synthesized in this tissue. Therefore, the level of *Bgp* mRNA expression in two mouse breast tumor cell lines, MMT 060562 and Mm5MT, was evaluated (Fig. 1A, lanes 10 and 11 respectively). Although no *Bgp* transcripts were observed in these samples in the exposure shown in Fig. 1A, faintly hybridizing *Bgp* transcripts of 4.0 and 1.8-kilobases were detected in RNA from the cell line Mm5MT upon five fold longer exposure to film (data not shown). RNA loading was evaluated by rehybridizing with a β -actin probe (Fig. 1B).

To verify if the reduction in mRNA expression was due to clonal variability of the carcinoma cell lines, *Bgp* mRNA expression was assessed in primary tumors (Fig. 1C). Primary CD-1 mouse tumors of the descending colon (Fig. 1C, lanes 4 and 7), in comparison to their normal adjacent mucosa (Fig. 1C, lanes 3 and 6) or to normal mucosa of the ascending colon (Fig. 1C, lanes 2 and 5), also demonstrated reduced expression of the major mouse *Bgp* transcripts. RNA loading was assessed by

rehybridization with a radiolabeled β -actin probe (Fig. 1D).

These findings suggest that mouse *Bgp* mRNAs are not efficiently transcribed in colon and liver tumor cells in comparison to the significant expression found in their normal tissue counterparts. Diffuse ~ 0.5 -kilobase *Bgp* cross-hybridizing transcripts were detected in all samples tested. Although the exact nature of these RNAs is unknown, it is unlikely that they represent *bona fide* *Bgp* transcripts due to their small size.

Posttranscriptional Activity

To determine if the observed reduction in *Bgp* transcriptional activity could be correlated with posttranscriptional events, immunoblotting and immunoprecipitation experiments were carried out using two polyclonal anti-mouse *Bgp* antibodies. The first polyclonal rabbit anti-mouse *Bgp* antibody (serum 231) was raised against purified mouse colon *Bgp* proteins and does not recognize other cell adhesion molecules of similar molecular weight such as E-cadherin or N-CAM, or of different molecular weight such as purified CEA (McCuaig et al., 1992). This antiserum recognizes numerous *Bgp* epitopes as well as the different *Bgp* isoforms that are expressed in mouse colon and liver: $\sim Mr$ 110,000-120,000 isoforms containing N-terminal, A1, B1, A2, and short or long IT domains, and $\sim Mr$ 54,000-60,000 *Bgp* proteins possessing N-terminal, A2, and short or long IT domains (McCuaig et al., 1992).

Fig. 2A represents an immunoblot analysis of normal colon and liver proteins as well as proteins of colon, liver, rectal and breast carcinoma cell lines using anti-mouse

Bgp antiserum 231. A predominant protein band of $\sim Mr$ 120,000 was revealed in normal colon membrane proteins (Fig. 2A, lane 1) and in proteins from the rectal carcinoma cell line CMT-93 (Fig. 2A, lane 2), while low amounts of $\sim Mr$ 120,000 were detected in normal embryonic liver (BNL CL.2) (Fig. 2A, lane 8), in the liver carcinoma cell line (BNL ING A.2) (Fig. 2A, lane 10) and in the breast carcinoma cell line (Mm5MT) (Fig. 2A, lane 7). Bgp proteins of $\sim Mr$ 110,000 were detected in normal liver membrane proteins (Fig. 2A, lane 11). This polyclonal antiserum also recognized proteins of $\sim Mr$ 54,000-60,000 in proteins samples from normal colon (Fig. 2A, lane 1), from the colon carcinoma cell line CT 26 (Fig. 2A, lane 3) and from normal liver (Fig. 2A, lane 11). No Bgp proteins of $\sim Mr$ 110,000-120,000 were detected in protein samples from the breast carcinoma cell line MMT 060562 (Fig. 2A, lane 6) or the colon carcinoma cell lines CT 26, CT 36, and CT 51 (Fig. 2A, lanes 3, 4 and 5 respectively). A nonspecific band of $\sim Mr$ 64,000 was seen in all samples and is considered to be a contaminant since it is also present in immunoblots incubated with a pre-immune serum (Fig. 2C).

We have previously determined the existence of mouse *Bgp* mRNA isoforms bearing short or long IT domains (McCuaig et al., 1993). As can be seen in immunoblots using the polyclonal anti-mouse Bgp antiserum 231, both BgpA- and BgpD-transfectant cells synthesized Bgp proteins of $\sim Mr$ 120,000 abundantly (Fig. 2B, lanes 3 and 4) in comparison to CMT-93 (Fig. 2B, lane 1) or to normal colon membrane proteins (Fig. 2B, lane 2), while parental NIH 3T3 cells and cells transfected with an antisense *BgpD* cDNA-driven construct did not significantly express these proteins (Fig.

2B, lanes 5 and 6 respectively). Furthermore, BgpA or BgpD proteins from transfected cells were readily immunoprecipitated with the anti-mouse Bgp antibody (Fig. 2F, lanes 3 and 4 respectively).

Since the long IT domain of Bgp has been implicated in bile acid transport and possibly signal transduction, we assessed the relative expression of isoforms bearing long IT domains in comparison to total Bgp protein expression. A second antibody was raised against a fusion peptide containing the long IT domain to specifically detect Bgp isoforms bearing this domain. The specificity of this antiserum was assessed by immunoprecipitation analyses of NIH 3T3 transfectant cells expressing either BgpA or BgpD proteins: the anti-IT antibody formed specific complexes with BgpD proteins (long IT domain) but not with BgpA proteins (short IT domain) (Fig. 2D, lanes 1 and 2 respectively).

Immunoprecipitation analyses using a polyclonal anti-mouse Bgp antiserum 231 or anti-IT serum 837 were performed on several cell lines. Bgp proteins of either $\sim Mr$ 120,000 and 60,000 were immunoprecipitated by the polyclonal anti-mouse Bgp antibody (serum 231) from proteins samples of normal colon, normal liver, the rectal carcinoma cell line CMT-93, the normal embryonic cell line BNL CL.2, the liver carcinoma cell line BNL ING A.2 and the breast carcinoma cells Mm5MT (Fig. 2E, lanes 1-3 and 6-8 respectively). Bands of $\sim Mr$ 75,000 or 90,000 (Fig. 2E, lanes 1 and 2 respectively) have previously been identified as breakdown products of the major $\sim Mr$ 120,000 Bgp glycoproteins (Odin et al., 1986). Only a small fraction of Bgp proteins from normal colon, normal liver and Mm5MT breast carcinoma were, however, immunoprecipitated

by the anti-IT antiserum 837 (data not shown). These findings confirm that, in both normal and transformed cells, Bgp isoforms bearing a long IT domain represent only a fraction of the total population of Bgp proteins (Culic et al., 1992; McCuaig et al., 1993) and that these particular isoforms are only expressed at very low levels in transformed cells.

Cellular Expression

Cellular localization of the Bgp proteins was assessed by immunostaining of normal and primary colon tumor tissues. Sections were cut from paraffin-embedded normal mouse colon and colonic tumors and immunostained for expression of mouse Bgp (Fig. 3). The most intense staining was obtained with the polyclonal rabbit anti-mouse Bgp antibody (serum 231), although a polyclonal anti-human CEA antibody also cross-reacted with the same proteins (data not shown). In normal colon (Fig. 3A), Bgp proteins are abundantly expressed at the apical membranes of crypt and superficial epithelial cells and in the cytoplasm of the crypt epithelial cells. Immunostaining of well-differentiated mouse colon adenocarcinomas invading the submucosa, the equivalent of a human colorectal Dukes stage A adenocarcinoma, revealed no Bgp protein expression (Fig. 3B). Furthermore, in sections containing both primary colon tumors and normal adjacent mucosa, Bgp expression was only detected in normal tissue (data not shown). Immunostaining with rabbit pre-immune serum was negative in normal and tumor tissues (data not shown). These results confirm the protein expression data observed in the cell lines: colon tumors do not express the major Bgp protein isoforms.

Regulation of Expression

Although *CEA* gene transcription is known to be upregulated by hypomethylation (Tran et al., 1988; Boucher et al., 1989) and by posttranscriptional events (Hauck and Stanners, 1991), little is known about the control of human or mouse *Bgp* gene transcription. We have examined several regulatory mechanisms including chromosomal rearrangement, gene amplification and DNA methylation of the *Bgp* gene, which may be partially responsible for the observed decreases in *Bgp* gene transcription and Bgp protein expression.

The possibility of chromosomal rearrangement or gene amplifications or deletions occurring in the vicinity of the *Bgp* genomic locus was assessed by comparing the patterns and intensities of hybridization of specific probes to restriction enzyme-digested genomic DNA from various mouse tissues or cell lines (Boucher et al., 1989). Equal amounts of genomic DNA from normal Balb/c spleen, from Balb/c-derived colon carcinoma cell lines (CT 26, CT 36, CT 51) and from the C57BL ICRF a¹-derived rectal carcinoma cell line (CMT-93) were digested with the restriction enzymes *EcoRI*, *HindIII*, *BamHI*, or *SstI* and subjected to Southern analyses using either a radiolabeled *EcoRI-SstI* *BgpA* cDNA fragment (enclosing the N-terminal and two C-2 set Ig domains of the *BgpA* cDNA) or a radiolabeled *BamHI-HindIII* fragment (corresponding to a segment of the mouse *Bgp* gene upstream region) as hybridization probes. For each particular restriction enzyme and hybridization probe combination, all of the DNA samples tested showed the same pattern and the same relative intensity of hybridization (data not shown). These results indicate that neither chromosomal rearrangement nor gene amplifications or

deletions have taken place close to the *Bgp* gene in the transformed cell lines tested. Therefore these two transcriptional regulatory mechanisms do not account for the differential *Bgp* expression seen in tumors *versus* normal cells.

Decreased DNA methylation in the upstream regions of the human *CEA* gene has been correlated with increased *CEA* gene transcription (Tran et al., 1988; Boucher et al., 1989). Because the mouse *Bgp* gene is a CEA-related gene that displays a different expression pattern than human CEA, we investigated whether altered methylation of the *Bgp* gene upstream region could be responsible for the observed decreases in *Bgp* gene transcription and corresponding decreases in *Bgp* protein expression in primary colon tumors and carcinoma cell lines.

Mammalian DNA is methylated at cytosine residues in the dinucleotide sequence CpG (Doerfler, 1983). By using the isoschizomeric restriction enzymes *MspI* and *HpaII* that are specific for the (5'-C↓CGG-3') site, changes in DNA methylation patterns are detectable (Boucher et al., 1989). Both enzymes are able to recognize this sequence when the cytosine residues are unmethylated; however, methylation of the second cytosine (5'-C^mCGG-3') alters this recognition site such that only the *MspI* enzyme can cleave the DNA chain at this site. Methylation on the first cytosine (5'-^mCCGG-3') is sufficient to render this site resistant to cleavage by both enzymes.

Genomic DNA from normal Balb/c colon, from the colon carcinoma cell lines CT 26, CT 36, CT 51 and from the rectal carcinoma cell line CMT-93 was digested with the restriction enzymes *HindIII*, *MspI* and *HindIII*, or *HpaII* and *HindIII*, and then submitted to Southern analyses. A 393 base pair *NcoI* fragment from a genomic λ phage

clone, corresponding to a segment of the proximal mouse *Bgp* gene promoter was used as the hybridization probe (Fig. 4). All genomic DNA samples digested with *Hind*III displayed identical hybridization patterns consisting of three discrete fragments of 1.3, 3.5 and 5.8-kilobases (Fig. 4A, lanes 1-5). Extensive characterization of *Bgp* gene-containing Balb/c genomic clones has led us to establish that the 1.3-kilobase *Hind*III fragment corresponds to a portion of the upstream region of the mouse *Bgp* gene and encloses the *Nco*I probe in its entirety (Nédellec, unpublished data). The *Nco*I probe crosshybridized to 3.5 and 5.8-kilobase fragments which represent two distinct CEA-related genes that share homology with the mouse *Bgp* gene (Nédellec unpublished results). Attempts to hybridize other fragments from the *Bgp* upstream region specific only to the *Bgp* gene have been unsuccessful due to their repetitive nature. All five genomic DNA samples digested with *Hind*III and *Msp*I (which cleaves its recognition sequence irrespective of the state of methylation of the second cytosine) showed identical hybridization patterns consisting of 3 distinct fragments of 0.7, 1.0 and 1.9-kilobases (Fig. 4B, lanes 1-5). Through sequence analyses of a mouse *Bgp* genomic clone we have determined that the 1.0-kilobase *Hind*III-*Msp*I fragment corresponds specifically to the known mouse *Bgp* gene (Nédellec, unpublished results).

Hybridization of the *Nco*I probe to *Hpa*II-*Hind*III digested DNA samples from normal colon and the cell line CMT-93 (Fig. 4C, lanes 1 and 5 respectively), which both express *Bgp* RNA and proteins, detected a 1.3-kilobase *Bgp* fragment similar to the 1.3-kilobase fragment obtained by digestion with *Hind*III alone (Fig. 4A, lanes 1 and 5), indicating that the *Hpa*II recognition sites found within the 1.3-kilobase *Hind*III-*Hind*III

fragment are methylated. Southern analyses of *HpaII-HindIII* digests of genomic DNA from the colon carcinoma cell lines CT 26, CT 36 and CT 51 detected two *Bgp* fragments of 1.0 and 1.3-kilobases at various hybridization intensities (Fig. 4C, lanes 2, 3 and 4 respectively). The presence of 1.0-kilobase *Bgp* fragments in these *HpaII-HindIII* digests, similar to the 1.0-kilobase fragments observed in the *MspI-HindIII* digests demonstrates that only a population of the cells in the colon carcinoma cell lines bears a *Bgp* gene that is unmethylated at the recognition sequence. As can be seen in Fig. 4C, lanes 1-5, the other two CEA-related genes recognized by the probe are also differentially methylated in normal and carcinoma cells. Overall, the transformed colon cell lines which do not express *Bgp* mRNA or proteins are less methylated than DNA from normal colon or CMT-93 cells, which both express *Bgp* proteins abundantly. This suggests that the differences in *Bgp* gene expression in tumor and normal cells are associated with changes in the structure of the 5' region of the *Bgp* gene. This correlation with expression is, however, different from other gene paradigms where demethylation is associated with increased gene expression.

DISCUSSION

Because CEA expression is upregulated in numerous human cancers, it was important in the development of a mouse model to evaluate the expression of the mouse *Bgp* gene, a CEA-related gene, in normal and malignant tissues. The results presented in this paper demonstrate that, in distinction to the human *CEA* gene, the expression of

mouse *Bgp* transcripts and protein isoforms in primary colon tumors and in colon and liver carcinoma cell lines is markedly decreased in comparison with their normal tissue counterparts. Furthermore our data shows that repression of the mouse *Bgp* gene in carcinoma cell lines correlates with DNA methylation of its upstream regulatory regions.

Northern analyses on total RNA revealed that normal colon and liver tissues express major *Bgp* transcripts of 4.0 and 1.8-kilobases. In contrast, several mouse colon and liver carcinoma cell lines expressed *Bgp* mRNAs at very low levels or not at all. The observed reduction in *Bgp* expression is not due to clonal variability of the cell lines since primary mouse tumors of the descending colon, in comparison to their normal adjacent mucosa or normal mucosa of the ascending colon also demonstrate lower expression of the major mouse *Bgp* transcripts.

Immunoblotting analyses with anti-mouse *Bgp* antibodies revealed a marked decrease in the expression of mouse *Bgp* isoforms in colon and liver carcinoma cell lines. *Bgp* proteins of ~*Mr* 110,000-120,000 that are abundantly expressed in normal colon, in normal liver and to a lesser extent in the rectal carcinoma cell line CMT-93 are expressed at very low levels, if at all, in the colon, liver and breast tumor cell lines tested.

Similar decreases in rat BGP (cell-CAM 105) expression have been demonstrated in regenerating liver (Odin and Öbrink, 1986) and in numerous rat hepatocellular carcinomas (Hixson et al., 1985, Hixson and McEntire, 1989). Moreover, the small fraction of rat liver tumors that do express BGP proteins, express an altered form with a more basic pI (Hixson and McEntire, 1989). Another report confirms that a rat

hepatocyte plasma membrane protein of $\sim Mr$ 110,000, shown by antibody crossreactivity to be identical to cell-CAM 105 and therefore rat BGP (Hixson and McEntire, 1989), is reduced in expression after malignant transformation (Becker et al., 1985).

Since identical *BgpA* cDNAs have been cloned from both adult liver (Dveksler et al., 1991) and colon (McCuaig et al., 1992), the differences in the apparent molecular weights of the major Bgp isoforms expressed in normal colon, in a normal embryonic liver cell line and in a transformed liver cell line ($\sim Mr$ 120,000) and normal liver ($\sim Mr$ 110,000) are likely due to variations in glycosylation; there are sixteen N-linked glycosylation consensus sites in the predicted amino acid sequence of the BgpA proteins (Dveksler et al., 1991; McCuaig et al., 1992) suggesting that these posttranslational modifications may be different in these two tissues or that these modifications may be associated with transformation. Interestingly, altered CEA N-glycosylation has been reported in human cancer cells and is postulated to be associated with early events in carcinogenesis (Garcia et al., 1991).

Contrary to previous findings by Robbins *et al.*, our data on *Bgp* mRNA expression in breast carcinoma cell lines does not entirely concur with the finding that *Bgp* transcripts are highly overexpressed when breast tissue undergoes carcinogenic transformation (Robbins et al., 1991). The discrepancy may, however, arise from the different amounts of RNA used in the both analyses. Our data reveal that Bgp proteins are expressed at low levels in the breast carcinoma cell line Mm5MT. However, Bgp expression is slightly increased in these MMTV-transformed Mm5MT cells in

comparison to normal mammary glands where no Bgp expression is detected. A more rigorous analysis of a greater number of tumors and cell lines will be required to confirm whether Bgp plays a significant role in breast tumor development.

Results from immunohistochemical analyses of primary tumors further demonstrate that Bgp protein expression is markedly decreased upon malignant transformation. Bgp proteins were also detected in the cytoplasm of the normal crypt epithelial cells. This cytoplasmic immunodetection may indicate that Bgp is synthesized in all crypt cells and that its expression may become restricted to the luminal membrane once the cells have migrated to the top of the crypt. Alternatively, these proteins may be Bgp degradation products found within lysosomes (Darnell et al., 1986).

The abundance of isoforms possessing long IT domains is significantly lower than those bearing short cytoplasmic domains, as shown by immunoprecipitation experiments, potentially reflecting a difference in the function(s) of these isoforms (McCuaig et al., 1993). In fact, isoforms bearing the long IT domain, which is highly conserved between mouse, rat and human, may be involved in signal transduction. A pp120 protein, shown by antibody cross-reactivity to be equivalent to rat BGP (Hixson and McEntire, 1989; Margolis et al., 1990), has been characterized as a substrate for tyrosine phosphorylation by the insulin receptor (Najjar et al., 1993) while a cAMP-dependent serine phosphorylation consensus sequence is found within the long IT of the rat ecto-ATPase (Lin and Guidotti, 1989). Furthermore, protein kinase C-dependent serine and threonine phosphorylation of human BGP has been demonstrated (Afar et al., 1992). This potential role of Bgp in signal transduction is worthy of further investigation since phosphorylation

by tyrosine kinase receptors and oncogenes are essential steps in cascade activation during embryonic development and differentiation (Ullrich and Schlessinger, 1990). A rat liver ecto-ATPase isoform bearing a long IT domain has been identified as being a canalicular bile acid transport protein whose function is dependent upon phosphorylation of its long IT domain (Sippel et al., 1993). In addition, the long IT domain is reported to associate with calmodulin in a Ca^{2+} -dependent manner and has been postulated to interact with cytoskeletal elements (Blikstad et al., 1992).

As yet, no rodent CEA-related cDNA homologs encoding glycopospholipid-anchored proteins have been isolated from the mouse or rat animal models. Human BGP is highly expressed in normal biliary ducts of the liver (Hinoda et al., 1990) and, to a lesser extent, in colon (Barnett et al., 1989), granulocytes (Drzenick et al., 1991) and leukocytes (Kuroki et al., 1991). In contrast, mouse Bgp transcripts and proteins are more abundant in colon tissues than in liver tissues (Beauchemin et al., 1989; Turbide et al., 1991; McCuaig et al., 1992; McCuaig et al., 1993). This difference in tissue expression may reflect differences in the function(s) of Bgp proteins in humans and mice.

To understand the diminished mouse *Bgp* gene transcription and Bgp protein expression, we studied several regulatory mechanisms that can alter gene transcription. Chromosomal rearrangements were not detected in the vicinity of the *Bgp* gene and therefore do not account for the observed decrease in *Bgp* gene transcription in colon tumor cell lines. Our results demonstrate, however, that the upstream regulatory region of the *Bgp* gene is less methylated in non-expressing colonic carcinoma cell lines than in normal expressing tissues. DNA methylation has been implicated in a number of

important biological processes: changes in DNA methylation states can perturb DNA-protein interactions, protect DNA against restriction endonucleases, affect overall DNA structure and regulate gene expression (Doerfler, 1983). Modification of the methylation pattern with *de novo* methylation or demethylation occurs throughout development, but is not well understood (Doerfler, 1983). Although the colon tumor cell lines tested show different states of DNA methylation in the *Bgp* gene upstream region, they are all less methylated in comparison to normal colon DNA. While human *CEA* gene transcription is upregulated by hypomethylation of its upstream region, our results demonstrate that hypomethylation of the *Bgp* gene is associated with the decreased expression of Bgp proteins upon tumor formation. The sites detected in our study that are differentially methylated may play a direct role in regulating *Bgp* gene expression. Alternatively, these differentially methylated sites may give rise to secondary changes in the chromatin structure or lead to methylation events at other sites. Additional experiments will be required to clarify these questions.

Although the human *CEA* and *BGP* genes are very homologous in their upstream regulatory regions (in 1.1-kilobases, they are 72% identical), their transcription is governed by different *trans*-acting factor binding (Hauck et al., submitted; Hauck and Stanners, submitted). In contrast to the human *CEA* gene (Hauck and Stanners, submitted), no silencer elements exist in the human *BGP* gene upstream region (Hauck et al., submitted). Therefore, transcriptional activation of the mouse *Bgp* gene, as reported during embryonic development (Huang et al., 1990) or transcriptional repression, as seen in malignant processes (this paper) are likely the consequence of the

several complex mechanisms involving altered methylation of the *Bgp* promoter, uniqueness of *cis*-acting elements and differential binding of *trans*-acting factors.

The results presented in this paper are consistent with a model in which the numerous functions associated with Bgp proteins such as adhesion, bile salt transporter, possible receptor and signal transduction, are essential for the maintenance of normal cellular architecture and of a differentiated phenotype. To determine if Bgp plays a direct role in the progression and the maintenance of cancerous states, future work will focus on understanding if disruption in the expression patterns of these normally abundant glycoproteins can be correlated with tumor progression and metastases, as has been elegantly shown for another adhesion molecule, E-cadherin (Frixen et al., 1991; Chen and Öbrink, 1991; Vleminckx et al., 1991).

Acknowledgements

We thank Dr. Moshe Szyf for discussions and critical reading of the manuscript, Dr. Michael Brattain for the generous gift of the colon carcinoma cell lines CT 26, CT 36, and CT 51, and Dr. André Veillette for the kind gift of NIH 3T3 cells.

Figure 1. Northern Analyses of Mouse *Bgp* mRNA. Twenty μ g of total RNA were electrophoresed on formaldehyde agarose-gels, transferred to nylon membranes and hybridized with the 32 P-labeled *Eco*R1-*Sst*I fragment of *BgpA* cDNA.

(A) Transcriptional activity in normal tissues and carcinoma cell lines: lane 1, normal Balb/c liver; lane 2, normal Balb/c colon; lane 3, CT 26; lane 4, CT 36; lane 5, CT 51; lane 6, CMT-93; lane 7, BNL CL.2; lane 8, BNL 1MEA.7R.1; lane 9, BNL 1NG A.2; lane 10, MMT 060562; lane 11, Mm5MT.

(B) Hybridization of the same membrane with a labeled *Pst*I β -actin fragment.

(C) Transcriptional activity in primary colon tissues and tumors: lane 1, normal CD-1 colon RNA. Lanes 2-7: RNA prepared from carcinogen treated mice: lane 2, normal ascending colon; lane 3, normal adjacent mucosa of tumor 1; lane 4, descending colon tumor 1; lane 5, normal ascending colon; lane 6, normal adjacent mucosa of tumor 2; lane 7, descending colon tumor 2.

(D) Hybridization of the same membrane with a labeled *Pst*I β -actin fragment. 18S rRNA (1.86-kilobases) and 28S rRNA (4.71-kilobases) are molecular size markers.

Figure 1

A 1 2 3 4 5 6 7 8 9 10 11

28S-

18S-

B

C

1 2 3 4 5 6 7

28S-

18S-

D

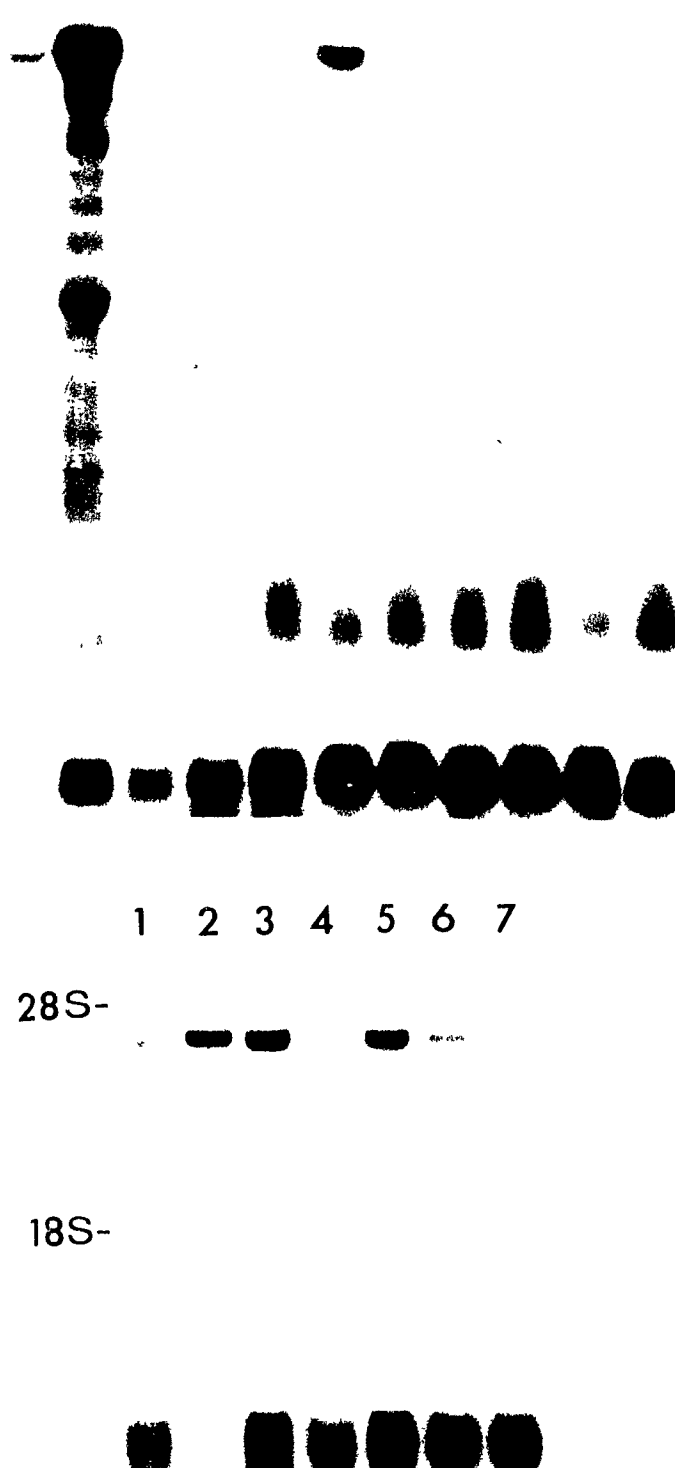


Figure 2. Bgp Protein Isoform Expression. Immunoblot and immunoprecipitation analyses using polyclonal anti-mouse antibodies.

(A) Immunoblot analyses using anti-mouse Bgp antibody (serum 231 at 1:1000 dilution) and 300 μ g of total cell lysate proteins from established cell lines, 5 μ g of normal colon membrane proteins, and 500 μ g of normal liver membrane proteins: lane 1, normal colon; lane 2, CMT-93; lane 3, CT 26; lane 4, CT 36; lane 5, CT 51; lane 6, MMT 060562; lane 7, Mm5MT; lane 8, BNL CL.2; lane 9, BNL 1MEA.7R.1; lane 10, BNL 1NG A.2; lane 11, normal liver.

(B) Immunoblot analysis of 150 μ g of transfectant cell proteins and 5 μ g of normal colon membrane proteins using anti-mouse Bgp antibody (serum 231 at 1:2000 dilution): lane 1, CMT-93; lane 2, normal colon; lane 3, BgpA transfectant cell proteins; lane 4, BgpD transfectant cell proteins; lane 5, antisense-driven BgpD transfectant cell proteins; lane 6, parental NIH 3T3 cell proteins.

(C) Immunoblot analysis of 150 μ g transfectant cell proteins with rabbit pre-immune serum (at 1:2000 dilution): lane 1, BgpA transfectant cell proteins; lane 2, CMT-93.

(D) Immunoprecipitation analysis of 200 μ g of transfectant cell proteins with the anti-IT antibody (serum 837), immunoblotted with anti-mouse Bgp antibody (serum 231 at 1:2000 dilution): lane 1, BgpD transfectant cell proteins; lane 2, BgpA transfectant cell proteins; lane 3, antisense-driven BgpD transfectant cell proteins; lane 4, parental NIH 3T3 cell proteins.

(E) Immunoprecipitation analysis of 500 μ g total cell proteins, 8 μ g colon membrane proteins, and 500 μ g liver membrane proteins with anti-mouse Bgp antibody (serum 231)

immunoblotted with anti-mouse Bgp antibody (serum 231 at 1:1000 dilution): lane 1, normal colon; lane 2, normal liver; lane 3, CMT-93; lane 4, CT 36; lane 5, CT 51; lane 6, BNL CL.2; lane 7, BNL 1NG A.2; lane 8, Mm5MT.

(F) Immunoprecipitation analysis of 200 μ g transfectant cell proteins and 8 μ g colon membrane proteins using the anti-mouse Bgp antibody (serum 231) and immunoblotted with the same antibody (at 1:2000 dilution): lane 1, normal colon; lane 2, CMT-93; lane 3, BgpA transfectant cell proteins; lane 4, BgpD transfectant cell proteins; lane 5, antisense-driven BgpD transfectant cell proteins; lane 6, parental NIH 3T3 cell proteins. Molecular size markers in kilodaltons. Large arrow indicates heavy chain IgG migrating at $\sim Mr$ 50,000.

Figure 2

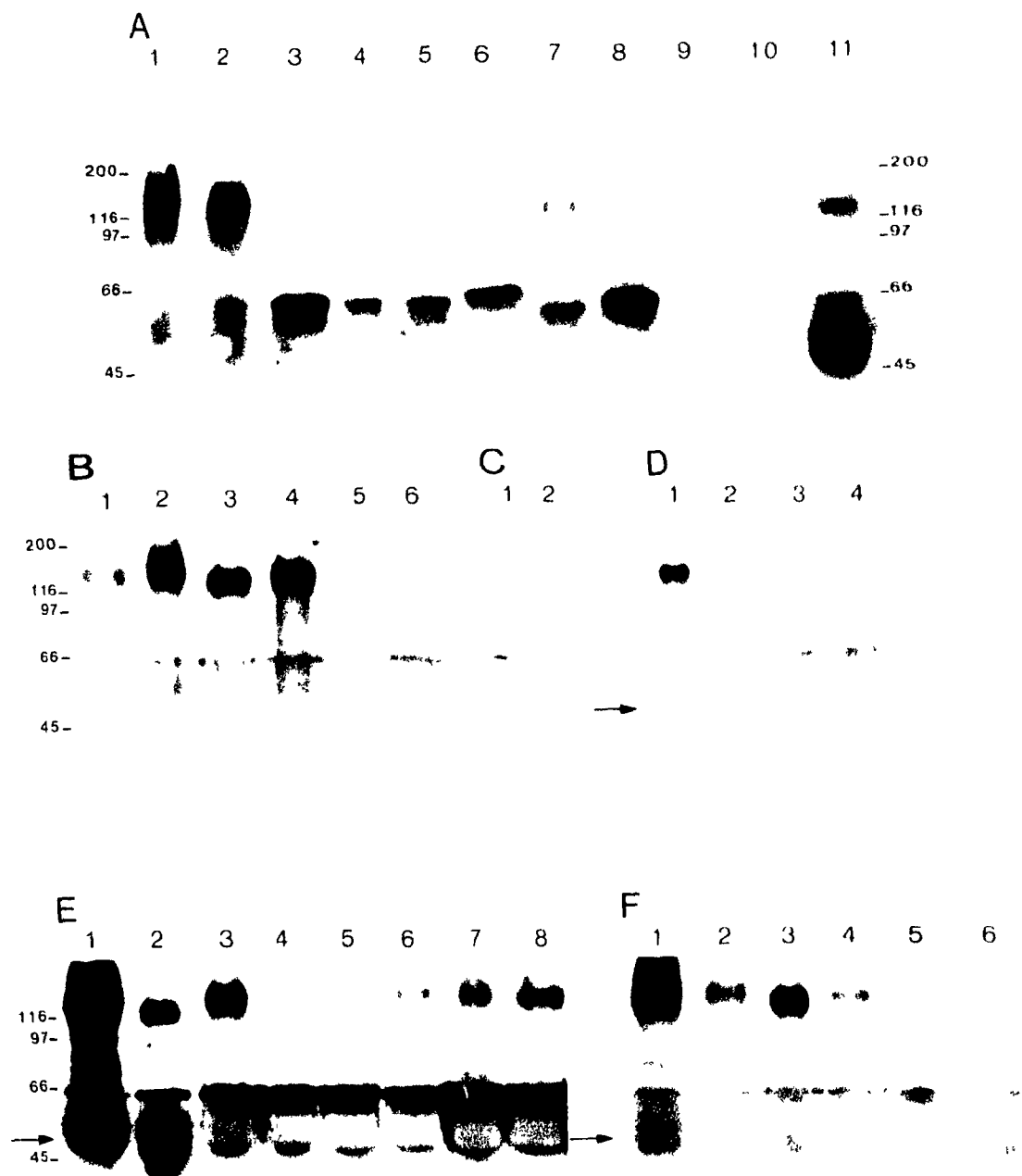


Figure 3. Immunostaining.

(A) Histological section of normal mouse colon immunostained with polyclonal rabbit anti-mouse Bgp antibody. The apical membrane of the epithelial cells in the crypt and on the surface as well as the cytoplasm of the crypt cells express Bgp proteins (X 100 magnification, paraffin section, immunoperoxidase with polyclonal anti-mouse Bgp antiserum 231 and counterstained with hematoxylin). Large arrows indicate luminal staining, small arrows indicate apical staining, while arrowheads indicate cytoplasmic staining.

(B) Histological section of mouse colon tumor showing no evidence of anti-mouse Bgp antibody reactivity (X 160 magnification).

Figure 3

A

B



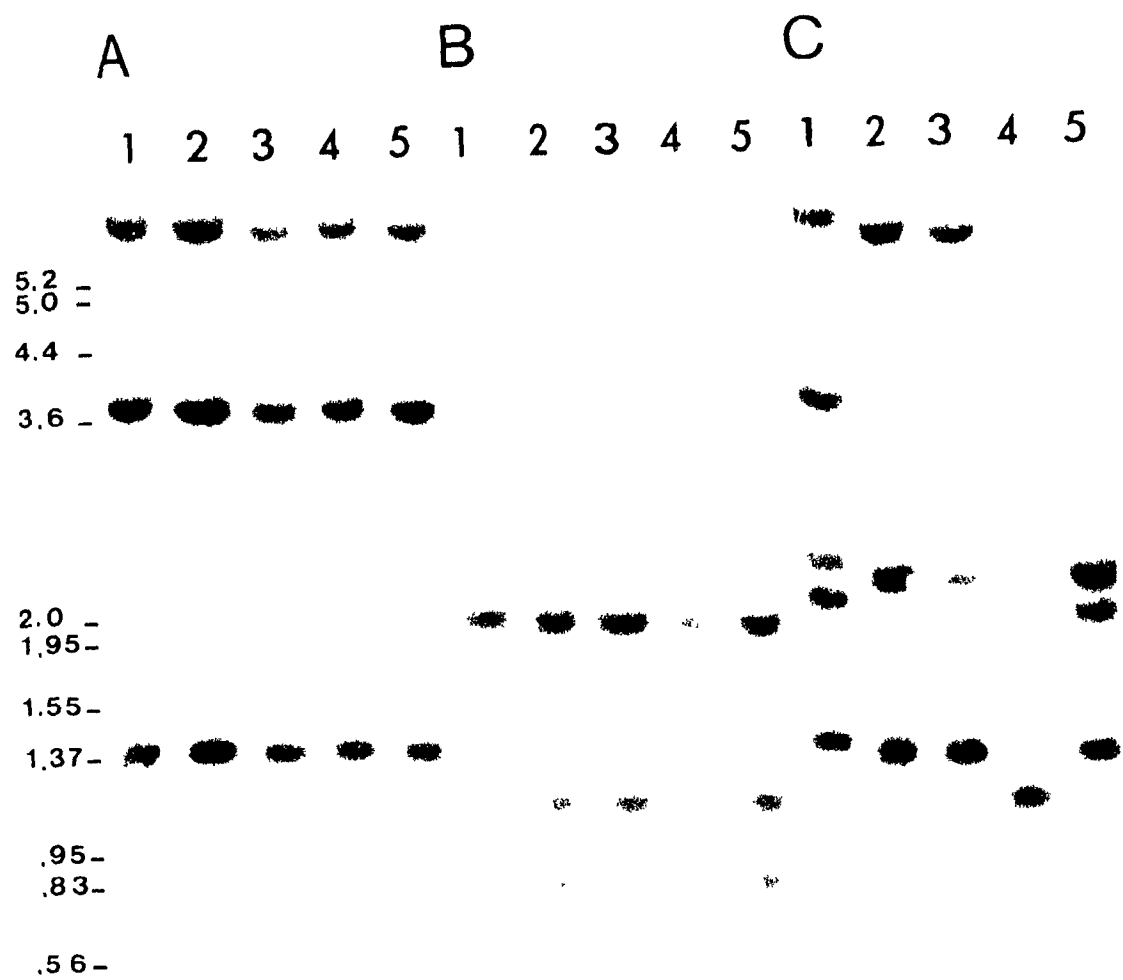
Figure 4. Genomic Southern Analyses. Five μg of genomic DNA were digested with various restriction enzymes and migrated on 0.8% agarose gels, transferred to a GeneScreen Plus membrane and probed with ^{32}P -labeled *NcoI* restriction fragment corresponding to the proximal upstream region of the mouse *Bgp* gene.

(A) Genomic DNA digested with *HindIII*.

(B) Genomic DNA digested with *MspI* and *HindIII*.

(C) Genomic DNA digested with *HpaII* and *HindIII*: lane 1, normal Balb/c colon; lane 2, CT 26; lane 3, CT 36; lane 4, CT 51; lane 5, CMT-93. Molecular size markers in kilobases.

Figure 4



CHAPTER FOUR

DISCUSSION AND CONCLUSION

The Characterization and Expression of Mouse Biliary Glycoproteins in Normal and Malignant Tissues

A mouse model is being developed as an *in vivo* approach to further investigate the functions of CEA-related antigens. Two full length mouse CEA-related cDNAs were previously identified and characterized (Turbide et al., 1991; McCuaig et al. 1992). These CEA-related cDNA isoforms are the mouse homologs of human BGP as defined by their deduced amino acid sequences, their mode of membrane anchorage and their expression patterns (Rojas et al., 1990; McCuaig et al., 1992; Barnett et al., 1993). Both mouse Bgp isoforms have been shown to function *in vitro* as cell adhesion molecules (Turbide et al., 1990; McCuaig et al., 1992) and to exhibit spatial and temporal expression patterns during development (Huang et al., 1990). (A) This thesis describes the cloning of several novel *Bgp* cDNAs as well as the structural characterization and the expression of nine similar but not identical mouse *Bgp* cDNA isoforms (Chapter 2, McCuaig et al., 1993). The data strongly suggest that the nine distinct *Bgp* isoforms represent splice and allelic variants of a single *Bgp* gene. (B) This dissertation also describes the study of mouse Bgp expression in normal and malignant tissues and examines several transcriptional control mechanisms which may regulate *Bgp* gene expression (Chapter 3, Rosenberg et al., in press). These findings reveal that mouse Bgp expression is downregulated in primary colon tumors and in colon and liver carcinoma cell lines. Our results show that Bgp expression is influenced by transcriptional control mechanisms involving DNA methylation of the *Bgp* gene upstream regulatory regions. A disruption in the expression of Bgp isoforms, such as that

observed in malignancy, may affect cell adhesion and disrupt normal tissue architecture.

The Role of Alternative Splicing and Allelic Variation in the Generation of Multiple Bgp cDNA Isoforms

Mouse *Bgp* isoforms are presently thought to be the products of complex alternative splicing of a single genomic transcriptional unit as well as allelic variants. Posttranscriptional processing of primary transcripts includes the removal of intervening sequences (introns) and the cleavage and polyadenylation of the 3'ends (Leff et al., 1986). For certain genes, the splicing of the primary transcripts always occurs in a constitutive fashion and gives rise a single protein product (Leff et al., 1986). For other genes, specific *bona fide* encoded exons are excluded from the mature RNA in a regulated process known as alternative splicing. This differential splicing of exons results in the production of multiple protein products from a single genomic transcriptional unit (Leff et al., 1986). Alternative usage of polyadenylation sites which generates 3'UTRs of varying lengths has also been described for a variety of genes and may influence the stability of mature transcripts (Leff et al., 1986).

We have shown that alternative splicing of mouse *Bgp* precursor mRNAs generates multiple transcripts which lack or exhibit the C2-set A1 and B1 domain exons and which include or exclude the 53 base pair exon producing two different IT forms (Chapter 2, McCuaig et al., 1993). The *Bgp* transcripts also display 3'UTRs of different lengths depending on the polyadenylation site employed during processing. Similar rat *Bgp* splice variants bearing either a long and short IT domain have been identified by

several laboratories. The cDNAs coding for the rat liver proteins ecto-ATPase, cell-CAM 105 and pp120 have been cloned and found to code for the same rat *Bgp* protein comprised of an N-terminal domain, three C-2 set Ig domains, a transmembrane domain and a long IT domain (71 amino acids) (Lin and Guidotti, 1989; Aurivillius et al., 1990; Najjar et al., 1993). Similar rat *Bgp* cDNA clones bearing short IT domains (10 amino acids) have also been characterized and are produced by the specific exclusion of a 53 base-pair exon from the mature transcript (Culic et al., 1992; Najjar et al., 1993). Rat *Bgp* splice variants lacking the A' and B1 domain exons, similar to several mouse *Bgp* splice variants, have not been identified through cloning or PCR amplification of reverse transcribed RNA (Najjar et al., 1993).

Extensive alternative splicing has also been reported for the human *BGP* gene (Hinoda et al., 1988; Barnett et al., 1989; Barnett et al., 1993) but differs somewhat from that of the murine *Bgp* genes (McCuaig et al., 1992; Chapter 2, McCuaig et al., 1993; Najjar et al., 1993). The most common human *BGP* splice variants are produced through the differential splicing of the C2-set A2 domain exon and of a 53 base pair exon which generates either a short or long IT domain (Barnett et al., 1989). However, other unusual *BGP* splice variants also exist. In the isoforms *BGPg*, *BGP_h* and *BGP_i*, differential splicing occurs such that these isoforms lack transmembrane and IT domains (Kuroki et al., 1991). In the recently characterized isoforms *BGPx* and *BGPx'*, the N-terminal domain exon is spliced to the exon encoding the transmembrane domain (Barnett et al., 1993) while in the isoforms *BGP_y* and *BGP_z*, the B1 domain exon is spliced to one of two small *Alu*-like exons found within the intervening sequence separating the A2

domain exon from the transmembrane domain exon (Barnett et al., 1993). The existence of similar unusual mouse or rat *Bgp* splice variants is plausible but remains to be demonstrated.

Alternative splicing has also been demonstrated for other Ig superfamily cell adhesion molecules which bear IT domains. Differential splicing of N-CAM primary transcripts generates at least five different RNAs whose protein products are either bound to the membrane by GPI-linkages similar to CEA and NCA or possess transmembrane and short or long IT domains (Cunningham et al., 1987; Edelman and Crossin, 1991). In both muscle and brain, the individual N-CAM isoforms are expressed in a spatial and temporal fashion during development, suggesting that specific isoforms are required for particular functions (Rutishauser and Jessel, 1988). Two isoforms of myelin-associated glycoprotein (MAG) which differ only in the lengths of their cytoplasmic domains are produced through developmentally-regulated alternative splicing of precursor transcripts (Salzer et al., 1987). The fact that human, rat and mouse BGP isoforms as well as N-CAM and MAG isoforms exist as splice variant which display similar long or short IT domains, suggests that both these domains may mediate important functions.

In contrast to human *BGP* (Barnett et al., 1993), two different N-terminal domains have been identified in mouse *Bgp* cDNA isoforms (McCuaig et al., 1992). The *Bgp* isoforms are grouped into two classes depending on which N-terminal domain and C2-set domains they contain. In each of the cDNA clones characterized, the N-terminal domain is associated with specific C2-set domains such that the N1 domain is found only associated with the A1a, B1a or A2a domains while the N2 domain is only joined with

the A1b, B1b or A2b domains (Chapter 2, McCuaig et al., 1993). However, close inspection reveals that the sequences of the transmembrane domain and the two IT forms found in N1- and N2-bearing *Bgp* isoforms are identical. This suggests that either an extremely complex alternative splicing mechanism exists or that two highly homologous but distinct *Bgp* genes are present in the mouse genome.

Evidence of allelic variation emanates from analyses of various mouse strains (Dveksler et al., 1993; Chapter 2, McCuaig et al., 1993). While both N1- and N2-containing *Bgp* cDNA isoforms have been isolated from the outbred CD-1 mouse strain (Turbide et al., 1991; McCuaig et al., 1992; Chapter 2, McCuaig et al., 1993), only N1-bearing *Bgp* isoforms have been identified through cloning or PCR amplification in the inbred Balb/c, CH3 and C57BL/6 mouse strains (Dveksler et al., 1993). It has recently been shown that N1-bearing *Bgp* isoforms serve as the mouse hepatitis virus (MHV) receptors in the MHV-susceptible mouse strains Balb/c, CH3 and C57BL/6 (discussed below) (Dveksler et al., 1993). In contrast, no N1-containing *Bgp* isoforms have been found in the inbred SJL/J mice which are resistant to MHV infection; Southern blot analysis of PCR amplified reverse transcribed RNA indicates that this mouse strain possesses only N2-containing *Bgp* isoforms (Dveksler et al., 1993). Viral binding assays have shown that intestinal and hepatocyte membranes of the F1 progeny of Balb/c X SJL/J matings bind MHV with only 40% of the Balb/c parental MHV-binding efficiency (Dveksler et al., 1993) while the membrane proteins of the SJL/J parent show no virus binding activity (Boyle et al., 1987). This value is similar to the predicted 50% viral-binding efficiency if N1- and N2-bearing *Bgp* proteins are encoded by two co-dominant

alleles of a single gene (Dveksler et al., 1993). These data suggest that the N1 and N2-bearing Bgp isoforms may be allelic variants of a single *Bgp* gene.

To elucidate the mechanisms involved in the generation of the multitude of *Bgp* splice variants, a thorough analysis of the Balb/c *Bgp* gene has been initiated (P. Nédellec and N. Beauchemin, unpublished results). Extensive sequencing of several genomic cosmid clones as well as Southern analyses of Balb/c genomic DNA reveal that the Balb/c *Bgp* gene contains only the N1 domain and has a genomic organization that is similar to that of the human and rat *BGP* genes (P. Nédellec, unpublished results; Barnett et al., 1993; Najjar et al., 1993). While the lengths of the intervening sequences vary, all three genes possess the following exons in the following order: 5'UTR/Leader, Leader/N, A1, B1, A2, TM, CytS, CytL/3'UTR (P. Nédellec, unpublished results; Barnett et al., 1993; Najjar et al., 1993); in the human *BGP* gene two small *bona fide* exons, Ily and Ilz, are found between the A2 and TM domain exons (Barnett et al., 1993). Analyses of overlapping genomic cosmid clones does not indicate the presence of N2-containing genomic segments in the vicinity of Balb/c *Bgp* gene, suggesting that N2 and N1 domains are not present in the same genomic transcription unit (P. Nédellec, unpublished results). Taken together, these results demonstrate that N1- and N2-bearing isoforms are encoded by two distinct genes.

Recent evidence indicates that strain-specific splice variants of rat *Bgp* may also exist. Culic et al. reported the sequence of a novel rat *Bgp* cDNA isoform which contains numerous amino acid substitutions in the N-terminal domain as well as scattered differences throughout the remainder of its sequence in comparison to the previously

identified rat *Bgp* cDNA isoform ecto-ATPase (Culic et al., 1992). Similar sequence variations were found in several rat *Bgp* isoforms isolated from an outbred Sprague-Dawley rat liver cDNA library (Stratagene, La Jolla, CA) (B. Öbrink, personal communication). These sequence differences are reminiscent of the variations observed in mouse *Bgp* isoforms and suggest that two *Bgp* genes may exist in the rat. Further investigations are required to determine if two distinct allelic *Bgp* genes are present in various rat strains. Because of the close evolutionary relationship between rodents, the possible existence of allelic *Bgp* genes in rats is not surprising. No similar evidence of allelic variation in human *BGP* has been reported but may, in fact, exist.

The existence of multiple mouse *Bgp* splice variants with similar structures suggests that these isoforms may have distinct yet overlapping functions. While exact of role of mouse *Bgp* proteins in normal cellular physiology is unclear, a number of different functions are associated with various mouse, rat and human *BGP* isoforms and are discussed below.

A Role in Cell Adhesion

While Ig superfamily members display an array of different functions, most members are involved in recognition roles at the cell surface (Williams and Barclay, 1988). Molecules containing C1-set Ig domains, such as the classical immunoglobulin molecules and the β and γ chains of T-cell receptors, are most often directly involved in immunorecognition (Williams and Barclay, 1988; Thompson et al., 1991). Antigens bearing C2-set Ig domains are generally associated with adhesion or receptor functions

(Thompson et al., 1991). Members of this group include the adhesion molecules ICAM, N-CAM, MAG, and CD2, and the cell surface receptors PDGFR and Fc-receptor (Williams and Barclay, 1988). The cell surface localization and sequence homology of CEA-related antigens to other C2-set Ig molecules strongly suggest that CEA family members may function *in vivo* as adhesion molecules and/or cell surface receptors (Thompson et al., 1991).

CEA, NCA, BGP and mouse Bgp all function as cell adhesion molecules when expressed at the surface of transfected cells (Benchimol et al., 1989; Oikawa et al., 1989; Rojas et al., 1990; Turbide et al., 1991; McCuaig et al., 1992; Oikawa et al., 1992). *In vitro* aggregation assays have demonstrated that mouse BgpA mediates cell adhesion in a Ca^{2+} - and temperature-independent manner (McCuaig et al., 1992) while mouse BgpB requires Ca^{2+} and physiological temperatures for cell adhesion (Turbide et al., 1991). The differences in the adhesion properties may arise from the different extracellular structures of the two isoforms: BgpA possesses an N1 domain and three C2-set domains while BgpB contains an N2 domain and one C2-set domain; both isoforms possess identical transmembrane and short IT domains. While the two N-terminal domains are 87% homologous, many of the amino acid sequence differences are non-conservative and involve changes in amino acid charge; the N1 domain also contains an additional potential N-linked glycosylation site (McCuaig et al., 1992). As the N-terminal domain appears to be a critical denominator for CEA, NCA and CGM6 adhesion (Oikawa et al., 1991), the differences between the two mouse Bgp N-terminal domains may contribute to the different adhesion properties. Alternatively, it has been

suggested that the levels of protein expression at the cell surface may influence the requirement for Ca^{2+} (M. Rojas et al., personal communication; Oikawa et al., 1992). While it remains to be demonstrated, the Bgp isoforms C-H are expected to be capable of mediating *in vitro* cellular aggregation since they contain similar if not identical extracellular domains as BgpA and BgpB; the presence of the long IT domain in certain isoforms is not expected to inhibit adhesion since human BGP proteins bearing long IT domains mediate cell adhesion (Rojas et al., 1990).

The rat Bgp homolog cell-CAM 105 was first identified as a cell surface protein capable of mediating rat hepatocyte aggregation *in vitro* (Ocklind and Öbrink, 1982) and was subsequently found to mediate homotypic Ca^{2+} -independent adhesion, similar as mouse BgpA (Tingström et al., 1990). Cell-CAM 105 is localized primarily to epithelial structures but is also found in granulocytes and platelets (Odin et al., 1988). The localization of cell-CAM 105 to hepatocyte cell-cell contacts, to the apical brush border of the intestine, and to the surface of activated platelets strongly suggests that rat Bgp participates in several different cell surface interactions involving membrane-membrane binding (Ocklind et al., 1983; Odin et al., 1988; Hansson et al., 1989). The tissue distribution of mouse *Bgp* transcripts and proteins during embryonic development and in normal adult tissues is consistent with an *in vivo* cellular adhesion function (Huang et al., 1990; Rosenberg et al., in press). However, detailed immunostaining experiments will be necessary to determine the specific subcellular localization of mouse Bgp proteins in various tissues. Furthermore, the determination of the specific expression patterns of individual Bgp protein isoforms may help determine the role of each isoform in cell

adhesion.

Ecto-ATPase Activity

Rat Bgp proteins have also been shown to exhibit ecto-ATPase activity (Lin and Guidotti, 1989). Ecto-ATPases are found on the surface of a large variety of cells and are characterized by being activated by Ca^{2+} and Mg^{2+} ions and by their ability to hydrolyse nonspecifically several different nucleotides (Lin and Guidotti, 1989). Cell lysates of transfectant cells expressing rat Bgp isoform have been shown to hydrolyse ATP in a Ca^{2+} -dependent fashion (Lin and Guidotti, 1989; Sippel et al., 1993). The structure and the cell surface localization of rat Bgp/ecto-ATPase are compatible with such an ectonucleosidase activity *in vivo* (Lin and Guidotti, 1989). It has been suggested that ecto-ATPase activity may influence Bgp-mediated adhesion since the addition of extracellular ATP stimulates rat hepatocyte aggregation *in vitro* (Öbrink, 1991).

A Potential Role in Signal Transduction

Biliary glycoproteins are unique members of the CEA family due to the fact that they contain transmembrane and IT domains of either 10 or 71-73 amino acids. In human, mouse and rat BGP, both the short and the long IT domains are rich in serine and threonine residues which can be phosphorylated (Barnett et al., 1989; Lin and Guidotti et al., 1989; McCuaig et al., 1993). Furthermore, two consensus tyrosine phosphorylation sequences are conserved in the long IT domains of all three species. Sequences similar to the consensus sequences for PKC dependent phosphorylation and

to a cAMP-dependent serine phosphorylation site have also been identified within the rat Bgp long IT domain (Lin and Guidotti, 1989; Sippel et al., 1993). The phosphorylation of proteins induces conformational changes and may alter protein function (Hunter and Cooper, 1985). Since protein phosphorylation is implicated in a number of different signal transduction pathways, it is tempting to speculate that BGP proteins may be involved in signal transduction events (Ullrich and Schlessinger, 1982). This potential role of BGP proteins in signal transduction is intriguing since phosphorylation by tyrosine kinase receptors and oncogenes are essential steps in cascade activation during embryonic development and differentiation (Ullrich and Schlessinger, 1982; Hunter and Cooper, 1985).

The rat Bgp protein pp120 has been characterized as a substrate of the tyrosine kinase activity of the insulin receptor (Rees-Jones and Taylor, 1985), the epidermal growth factor (EGF) receptor (Phillips et al., 1987) and the insulin-like growth factor I (IGF-I) receptor (Fanciulli et al., 1989). This stimulated rat Bgp tyrosine phosphorylation has been demonstrated in cell free systems, in intact cultured hepatoma cells, and in perfused liver where normal polarity of the cell membranes is maintained (Rees-Jones and Taylor, 1985; Accili et al., 1986; Phillips et al., 1987; Margolis et al., 1988; Fanciulli et al., 1989). *In vivo* serine phosphorylation is also likely since the rat Bgp protein cell-CAM 105 is phosphorylated on serine residues when isolated rat hepatocytes are cultured in [32 P]P_i (Odin et al., 1986).

Phosphorylation of human BGP has also been demonstrated. Human BGP isoforms expressed in transfected cells, in a colon carcinoma cell line and in an acute

myelogenous leukemia cell line, are phosphorylated on serine and threonine residues upon stimulation with TPA, an activator of PKC (Afar et al., 1992). BGP isoforms bearing long IT domains are also phosphorylated on tyrosine residues (Afar et al., 1992). However the endogenous protein kinase(s) involved in the phosphorylation of human BGP remain to be identified.

As yet, no phosphorylation assays on mouse Bgp isoforms have been performed. The phosphorylation status of mouse Bgp isoforms can be assessed by incubating transfectant cells that express Bgp isoforms bearing the long IT domain (with short IT expressing-transfectants used as controls) with ^{32}P -orthophosphate and with or without stimulation by insulin, EGF or IGF-1. The phosphate content of Bgp proteins can be assessed by lysing the labeled cells and immunoprecipitating the Bgp proteins with specific anti-mouse Bgp antibodies or anti-(long)-IT antibodies and subjecting the immunoprecipitated proteins to Western analyses or phospho-amino acid analysis. Similar experiments can be performed on the rectal carcinoma cell line CMT93 which expresses Bgp isoforms which display long IT domains. Alternatively, the phosphotyrosine content of Bgp proteins can be determined by immunoprecipitation and detection with anti-phosphotyrosine antibodies in conjunction with anti-mouse Bgp antibodies. Because of the high homology between the mouse, human and rat long IT domains, mouse Bgp isoforms are likely to be phosphorylated on serine, threonine and tyrosine residues *in vitro* and possibly *in vivo*. It has been suggested that the phosphorylation of human BGP proteins and by extension mouse Bgp proteins may be involved in modulating cell adhesion or transmembrane signalling events, although no

data supports these hypotheses (Afar et al., 1992).

Potential Interactions with Cytosolic Proteins

Cell adhesion molecules (CAMs) are implicated in a variety of important functions including morphogenesis, immunological events and cell-cell signalling (Öbrink, 1991). It is generally accepted that CAM mediates such processes by transmitting the binding signal across the membrane either to their own cytoplasmic domains or to membrane-associated proteins (Blikstad et al., 1992). E-cadherin, an important epithelial cell adhesion molecule, has been shown to be associated with three intracellular proteins called α , β and γ catenins through its IT domain (Takeichi, 1991). Integrins, important cell-matrix adhesion proteins, also interact with several cytoskeletal proteins including talin and α -actinin and are implicated in signal transduction events (Hynes, 1992).

In order to elucidate the function(s) of the long IT domain of Bgp proteins, a search for proteins associating with this domain was initiated (Öbrink et al., 1988). The rat Bgp protein cell-CAM 105 bearing a long IT domain was recently shown to bind calmodulin, a ubiquitous cytosolic regulatory protein involved in the modulation of intracellular Ca^{2+} concentrations (Öbrink et al., 1988; Blikstad et al., 1992). Calmodulin bound specifically to the long IT domain in a Ca^{2+} -dependent and reduction-sensitive manner (Blikstad et al., 1992). These analyses also revealed that rat Bgp/cell-CAM 105 is a major calmodulin-binding protein in liver plasma membranes. Calmodulin is known to interact with various components of the cytoskeleton as well as with other

proteins (Cheung, 1980; Howe et al., 1980; Glenney and Weber, 1980). The binding of Bgp molecules on different cellular membranes may alter the conformation of the IT domain and affect the binding of the long IT with calmodulin and other proteins (Blikstad et al., 1992). In such a model, cell-cell adhesion could influence cytoskeletal associations or affect other cellular components (Blikstad et al., 1992).

To determine if mouse Bgp proteins bind to calmodulin, we will conduct crosslinking and immunoprecipitation analyses under reducing and non-reducing conditions using anti-mouse Bgp antibodies, antibodies specific for the long IT domain, and commercial anti-calmodulin antibodies. The crosslinking and immunoprecipitation experiments will also help to determine if other proteins bind to the Bgp short and long IT domains. These experiments should shed light into the role of mouse Bgp proteins in a cascade of events implicating intercellular adhesion and signal transmission.

A Role in Hepatic Bile Acid Transport

Hepatocytes are polarized cells which possess distinct sinusoidal (basolateral) and canalicular (apical) plasma membrane domains; these domains are morphologically and functionally distinct with each domain containing specific proteins (Hong and Doyle, 1987). The continuous trans-hepatocellular secretion of bile acids from the blood to the bile canalicular lumen involves the uptake of bile acids from the blood at the sinusoidal membrane, their translocation across the cell to the bile canalicular domain where they are finally transported across the plasma membrane (Meier, 1989; Sippel et al., 1993). Recently a ~110kD glycoprotein involved in the transport of bile salts across the

canalicular domain was identified through amino acid sequencing as the rat Bgp isoform ecto-ATPase which displays a 71 amino acid IT domain (Sippel et al., 1993). COS cells transfected with the rat Bgp ecto-ATPase cDNA exhibited bile acid transport activity *in vitro*. In contrast, COS cell expressing a truncated rat Bgp lacking the IT domain were unable to pump out bile acids, demonstrating that the long IT domain is necessary for bile acid efflux. It was also shown that rat Bgp/ecto-ATPase is phosphorylated on its IT domain by PKC activators (Sippel et al., 1993). Furthermore, inhibition of phosphorylation by vanadate (a phosphorylation inhibitor) and staurosporine (a PKC inhibitor) abolishes bile acid transport activity indicating that phosphorylation is essential for this function. Ecto-ATPase activity was not affected by phosphorylation inhibitors demonstrating that phosphorylation is not essential for ecto-ATPase activity. It is unclear if ecto-ATPase activity is necessary for bile acid efflux (Sippel et al., 1993).

The mechanism of rat Bgp-mediated bile acid transport across the canalicular domain is unknown. The data suggest that the intracytoplasmic domain contains a bile acid binding domain (Sippel et al., 1993). It has been previously demonstrated that canalicular membranes proteins are first translocated to the sinusoidal membrane before being targeted to the canalicular membrane (Bartles et al., 1987). It is inviting to speculate that rat Bgp proteins may in some way mediate the uptake and the transport of bile acids across the hepatocyte in addition to mediating their efflux from the hepatocyte (Sippel et al., 1993).

As the majority of membrane transport proteins identified possess multiple membrane spanning domains (Jennings, 1989), the single transmembrane domain

structure of rat Bgp is atypical for a transport protein (Sippel et al., 1993). However, several transport proteins containing only one transmembrane domain have been identified suggesting that Bgp monomers may be functionally active (Sippel et al., 1993). Alternatively, rat Bgps bearing long IT domains may be functionally active as dimers as has been shown for the Na,K-ATPase (DeTomaso et al., 1993) or as multimers. Since the transport of bile acids across the canalicular domain is the rate-limiting step in the process of hepatic biliary secretion (Sippel et al., 1993), it is tempting to speculate that the differential expression of the alternatively spliced isoforms may regulate the secretion of hepatic bile acids.

Mouse Hepatitis Virus Receptors

Several mouse Bgp proteins also function as receptors for the mouse hepatitis virus (MHV) A-59 (Dveksler et al., 1993), a murine coronavirus which causes respiratory infections, hepatitis, and demyelinating diseases in susceptible mice strains (Wege et al., 1982). The Bgp isoforms expressed in the MHV-susceptible inbred Balb/c, CH3 and C57BL/6 mouse strains, have been characterized and found to contain only N1-bearing isoforms (Dveksler et al., 1993). In collaboration with Dr. K. Holmes' laboratory, we have recently shown that the Bgp N1-containing isoforms MHVR1, BgpC and BgpD are able to confer MHV susceptibility when transfected into MHV resistant non-Bgp expressing cells, confirming that these N1-bearing Bgp isoforms are functional MHV receptors (Dveksler et al., 1993).

In contrast, adult SJL/J mice which express only N2-bearing Bgp isoforms are

resistant to MHV infection. This finding initially suggested that the MHV virus could bind to N1-bearing isoforms but not to N2-containing isoforms. However, *BgpB*, an N2-containing isoforms isolated from CD-1 colon, was surprisingly able to confer MHV susceptibility when transfected into resistant non-Bgp expressing cells (Dveksler et al., 1993). The reasons for this discrepancy are unclear and warrant further investigation. Preliminary genomic Southern analyses indicate that an alteration may exist in the SJL/J *Bgp* gene (P. Nédellec, unpublished data) while Western analyses have shown that these mice express Bgp-like proteins which have a slightly inferior molecular weight in comparison to Bgp proteins from Balb/c or CD-1 (Williams et al., 1990; McCuaig et al., 1992; Rosenberg et al., in press).

MHV is the first virus to be identified which can use multiple protein isoforms as receptors (Dveksler et al., 1993). Several other Ig family members have also been shown to serve as virus receptors: CD4, which is expressed in helper T lymphocytes, is the viral receptor of HIV (Maddon et al., 1986), the cell adhesion molecule ICAM-1 functions as a receptor for the common cold rhinoviruses (Staunton et al., 1989; Greve et al., 1989), and an unnamed Ig family member serves as the poliovirus receptor (Mendelsohn et al., 1989). The usurped viral receptor functions are not startling since Ig superfamily members are large carbohydrate-rich molecules expressed at the cell surface, the location where the virion begins its assault on a host cell (White and Littman, 1989).

Functional Implications of the Existence of Multiple Bgp Isoforms

Together, these results suggest that mouse Bgp proteins may be involved in a cascade of events implicating intercellular adhesion, bile acid transport, response to growth factors and possibly signal transduction to other cellular compartments or proteins. The existence of isoforms with two alternate extracellular structures combined with two distinct intracellular forms suggests that multiple functional forms may exist. Certain Bgp function(s) may be mediated by Bgp monomers while others may require the dimerization of Bgp proteins or the formation of multimeric Bgp complexes. Functionally important heterotypic interactions have been demonstrated for several other Ig family members such as CD8-CD1 (Snow et al. 1985), CD3-TCR (Brenner et al., 1985) and CD2-LFA-3 (Mentzer et al., 1987). The ectopic expression of *BgpA* or *BgpB* cDNA isoforms in non-adhesive cells is sufficient to confer adhesion, suggesting that the association between structurally distinct Bgp isoforms is not necessary for cell adhesion *in vitro* (McCuaig et al., 1992; Turbide et al., 1991). However, homotypic or heterotypic associations between Bgp isoforms may be involved in the modulation of Bgp adhesion *in vivo*. Cross-linking and immunoprecipitation of Bgp proteins from different tissues may help determine their functional form(s).

We have investigated the expression of mouse *Bgp* splice variants in normal adult colon. Using domain-specific oligonucleotide probes, we have shown that transcripts bearing long IT domains are ~10 fold less abundant than other isoforms (Chapter 2; McCuaig et al., 1993). This finding has been corroborated at the protein level by immunoprecipitation analyses which demonstrated that Bgp proteins which display a long

IT domain are expressed at a 10-20 fold lower abundance than other isoforms (Chapter 3, Rosenberg et al., in press). Furthermore, Northern analyses demonstrated that isoforms bearing only one C2-set domain (A2) are found only in small amounts in comparison to those containing three C2-set domains (Chapter 2, McCuaig et al., 1993). However, the functional significance of this differential expression remains unknown. It is possible that the regulated expression of a discrete set of Bgp isoforms may give rise to tissue-specific functions. For example, the differential expression of Bgp isoforms bearing different extracellular domains may be important in the modulation of cell adhesion. It can also be envisaged that variations in the ratios of short and long IT-bearing Bgp proteins may modulate intracellular responses to adhesion or other Bgp functions. RNase protection and PCR amplification analyses are presently being carried out to define the discrete expression patterns of the various mouse *Bgp* isoforms in normal and transformed tissues as well as in embryonic development.

The relevance of allelic variation of the mouse *Bgp* gene is unclear. Balb/c and CH3 mice are not adversely affected by being homozygous for the N1 *Bgp* allele and are as healthy as outbred CD-1 mice. SJL/J mice, in which only N2 *Bgp* isoforms have been identified (Dveksler et al., 1993), are phenotypically normal until they reach adulthood, at which point they begin to develop tumors (Wanebo et al., 1966). Preliminary genomic DNA analyses indicate that a deletion or aberration may exist in the SJL/J *Bgp* gene (P. Nédellec, unpublished data). While it is tempting to speculate that an alteration of the SJL/J *Bgp* gene may play a role in tumor formation, further experiments are required to determine the structure of the SJL/J *Bgp* gene.

The Role of Mouse Bgp Proteins in Malignant and Regenerating Tissues

Cell adhesion plays a central role in embryonic development, in the maintenance of normal tissue architecture and in the pathology of multicellular organisms (Edelman, 1988). In tumorigenesis, profound alterations in tissue organization occur as a result of fundamental changes in cell adhesion. This is most pronounced in tumor invasion and metastasis, a complex phenomena involving the detachment of malignant cells from adjacent tissues and their migration to another site where invasion and resumed growth may occur (Sommers et al., 1991). The altered expression of CAMs is postulated to play a key role in the detachment of malignant cells from the primary site and their invasion and adhesion at a secondary site. E-cadherin, a Ca^{2+} -dependent homophilic cell adhesion molecule, is postulated to act as an invasion suppressor molecule (Behrens et al., 1989; Vleminckx et al., 1991; Frixen et al., 1991). Detailed studies have shown that E-cadherin expression is downregulated in numerous human malignancies (Behrens et al., 1989; Frixen et al., 1991). Furthermore, the loss of E-cadherin expression has been correlated with increased invasiveness of human carcinoma cells (Behrens et al., 1989; Frixen et al., 1991). These results are corroborated by invasion assays which demonstrate that the transfection and high expression of E-cadherin into highly invasive epithelial tumor cell lines abrogates invasion, presumably by the formation of cell-cell contacts and the inhibition of cellular motility (Vleminckx et al., 1991; Chen and Öbrink, 1991). In contrast, CEA and NCA expression is upregulated in intestinal cancers (Thompson et al., 1991). It is speculated CEA and NCA overexpression may contribute to the carcinogenic process by disrupting normal tissue architecture or through adhesion

in the process of invasion and metastasis (Benchimol et al., 1989; Jessup and Thomas, 1989). However, the role of CEA and NCA in tumor formation and progression remains to be determined experimentally.

Biliary glycoprotein expression is decreased in carcinomas, similar as E-cadherin expression. Our results demonstrate that mouse Bgp mRNA and protein expression is markedly decreased in colon and liver carcinomas cell lines and in primary colon tumors in comparison to their normal tissue counterparts (Chapter 3, Rosenberg et al., in press). Decreased rat Egp protein expression has also been demonstrated in liver cell lines and in numerous transplantable hepatocellular carcinomas (Hixson et al., 1985; Becker et al., 1989; Hixson and McEntire, 1989). The majority of these hepatocellular carcinomas do not express detectable levels of rat Bgp (Hixson and McEntire, 1989). The small fraction of the hepatocellular carcinomas that do express Bgp, express an altered form which has a more basic pI (Hixson and McEntire, 1989). The alteration in the pI has been attributed to glycosylation differences (McEntire et al., 1989). In the case of human BGP, no conclusive evidence exists on the expression of BGP proteins in carcinomas. Barnett et al. (1993) have recently shown that the transcripts of several unusual BGP isoforms are found in larger quantities in RNA from tumor cell lines than in RNA from normal tissues. The loss of expression of mouse Bgp adhesion molecules in tumors may lead to loss of intercellular adhesion and facilitate the emigration of carcinoma cells from the primary site and may represent an important step in the acquisition of an invasive and malignant phenotype. However, it is clear that numerous other alterations are also involved in the carcinogenic process (Nowell, 1986).

Regenerating and fetal liver express low amounts of Bgp proteins in comparison to the levels found in normal liver. Rat liver Bgp proteins are expressed late in gestation, appearing at 16-18 days p.c. and increasing until reaching maximum levels shortly after birth (Odin and Öbrink, 1986; Feracci et al., 1987). Reduced expression in regenerating liver after partial hepatectomy has been documented for the rat Bgp homologs cell-CAM 105 and HA4 (Odin and Öbrink, 1986; Bartles and Hubbard, 1986). Thus, regenerating and neoplastic rat liver cells share the common feature of expressing little or no Bgp. Diminished Bgp expression is correlated with cell proliferative processes which do not require tight intercellular-binding, suggesting that Bgp expression may be involved in the maintenance of normal tissue architecture or in normal cellular growth processes.

Transcriptional Regulation of the Mouse *Bgp* Gene

Despite their extensive homology, the members of the CEA family display diverse expression patterns in both normal and malignant tissues (Thompson et al., 1991). One approach to understand the differential expression of CEA family members has been to investigate the mechanisms which regulate their expression. Changes in the methylation status of genomic DNA can alter chromatin structure, perturb DNA-protein interactions, and regulate gene activity (Doefler, 1983; Cedar, 1988). Modification of the methylation pattern with *de novo* methylation or demethylation occurs throughout development, but is poorly understood (Doefler, 1983; Cedar, 1988). The overexpression of the *CEA* gene in colon carcinomas has been correlated with hypomethylation of the upstream regulatory

regions (Tran et al., 1988; Boucher et al., 1989). In contrast, our results demonstrate that the upstream regulatory regions of the mouse *Bgp* gene are less methylated in non-*Bgp*-expressing colonic carcinoma cell lines than in normal expressing tissues (Chapter 3, Rosenberg et al., in press). Since the control of gene expression is regulated by the DNA methylation of highly specific sites (Doefer, 1983), it is possible that additional sites which were not detected in our study may play a direct role in regulating *Bgp* gene expression. Alternatively, a change in the methylation status of these specific sites may give rise to secondary changes in the chromatin structure which may indirectly repress *Bgp* gene transcription. Additional studies on the overall methylation status of the mouse *Bgp* gene are required to resolve these questions. Since demethylated genes often adopt a DNAase I sensitive structure (Doefer, 1983), treatment with DNAase I may reveal the extent of DNA methylation of the *Bgp* gene in various cell lines and tissues. Because human and rodent BGPs possess structural features and expression patterns which differ from CEA and NCA, it is likely that BGP expression is controlled by different transcriptional activation mechanisms. Analyses of the putative promoter regions of *CEA*, *NCA* and human *BGP* reveal that these genes do not possess the typical promoter TATA or CAAT boxes (Schrewe et al., 1990; Hauck and Stanners, submitted; Hauck et al., submitted), a feature found in all CEA family members studied thus far (Thompson et al., 1991). While *CEA* transcription is controlled by an upstream silencer region, no similar silencing elements were identified in the *BGP* gene (Hauck and Stanners, submitted; Hauck et al., submitted). Although the human *CEA* and *BGP* genes are over 72% homologous over a 1.1-kilobase region, footprinting and mobility shift

assays reveal that their transcription is governed by different transcription factors (Hauck and Stanners, submitted; Hauck et al., submitted). A HNF-4/LF-A1-like factor was found to bind specifically to the upstream region of the *BGP* gene (Hauck et al., submitted) while SP1 and SP1-like factors were only recognized by elements found in the upstream region of the *CEA* gene (Hauck and Stanners, submitted). The transcription factor USF and an AP2-like factor were found to bind to upstream regulatory regions of both *BGP* and *CEA* (Hauck and Stanners, submitted; Hauck et al., submitted).

The upstream regulatory regions of the mouse *Bgp* gene are presently being characterized by members of Dr N. Beauchemin's laboratory. The aims of these investigations are to identify the *cis*-regulatory elements that are responsible for the transcriptional activation in different cell types. Preliminary data reveals that the mouse *Bgp* gene also lacks the classical promoter elements such as the TATA and CAAT boxes (P. Nédellec, unpublished data). Deletion construct analyses of the *Bgp* upstream regulatory regions using the luciferase reporter gene will help determine if certain sequence elements are involved in the modulation of the transcriptional activity of the mouse *Bgp* gene. Since rat Bgp protein expression in the uterus is controlled by ovarian steroid hormones (Svalander et al., 1990), it is tempting to speculate that mouse Bgp expression may also be hormonally regulated and that estrogen and progesterone-responsive elements may be present in the upstream regulatory regions of the mouse *Bgp* gene. The transcriptional activation of the mouse *Bgp* gene, as reported during embryonic development (Huang et al., 1990) or repression as seen in malignant processes (Rosenberg et al., in press) are likely to be the consequence of the several complex

mechanisms involving altered methylation of the regulatory regions of the *Bgp* gene, unique transcription factor binding and the activation of *cis*-acting regulatory elements.

CONCLUSIONS

We have identified nine distinct mouse *Bgp* cDNA isoforms and have shown that these isoforms are splice and allelic variants of a single gene. Mouse *Bgp* proteins are implicated in a variety of cellular functions including cell adhesion, hepatic bile acid transport and signal transduction. However, the significance of the existence of multiple mouse *Bgp* protein isoforms is unclear. Since the differential expression of mouse *Bgp* proteins is postulated to play an important role in embryonic tissue differentiation (Huang et al., 1990), it is possible that the expression of a particular set of *Bgp* isoforms may give rise to tissue-specific functions. It is, therefore, important to determine exactly which *Bgp* isoforms are expressed at various periods of development. This can be accomplished by Northern analyses, RNase protection assays, and PCR amplification of reverse transcribed RNA obtained from pre-implantation blastocysts and micro-dissected embryonic tissues obtained at different stages of development. Likewise, the characterization of the *Bgp* transcript and protein isoforms expressed in tumors is particularly important since we have shown that *Bgp* expression is decreased or altered in malignant mouse tissues and may represent an important step in the acquisition of an invasive and malignant phenotype. The determination of the tissue expression patterns of specific *Bgp* isoforms will help to elucidate the role(s) of *Bgp* proteins in embryonic

development, in the maintenance of normal tissue architecture, and in tumor formation and progression.

The study of the mouse *Bgp* gene promoter and upstream regulatory elements will shed light into the transcriptional control mechanisms which regulate *Bgp* expression. Deletion construct analyses using the luciferase reporter gene are presently being carried out and will yield information on *Bgp* gene promoter activity. Experiments will be conducted to elucidate the transcription factors and the *cis*-acting regulatory elements that are involved in mouse *Bgp* gene expression. The determination of the transcription factors and the *cis*-acting regulatory elements may yield clues on the mechanisms that cause the loss of *Bgp* expression in tumors and may help determine the role of *Bgp* proteins in development.

Two novel mouse *CEA*-related genes which share high homology with the *Bgp* gene described above have recently been identified by Patrick Nédellec of Dr. Nicole Beauchemin's laboratory. The structures of these new identified *CEA*-related genes are presently being characterized through sequence analyses of genomic cosmid clones (P. Nédellec, unpublished results). The Southern analyses presented in Chapter 3 reveal that the upstream regulatory regions of these two newly identified mouse *CEA*-related genes also exhibit altered methylation states in carcinoma cell lines (Chapter 3, Rosenberg et al., submitted). However, it is not possible at the present time to draw any conclusions from their expression and their DNA methylation status since the expression patterns of the two novel mouse *CEA*-related genes have not yet been assessed.

The ongoing investigations into the functions of mouse *Bgp* proteins and into the

regulation of mouse *Bgp* gene expression as well as the characterization of the two novel *CEA*-related genes are instrumental in the establishment of the mouse model for the *CEA* family of related genes. The ideas presented in this thesis are based on the results available at the present moment. Therefore, it is likely that the new results generated by the genomic analyses on the mouse *Bgp* gene and the two novel mouse *CEA*-related genes will engender new hypotheses on the structure and organization of the mouse *CEA* gene family and may necessitate the modification or the reformulation of our current ideas.

REFERENCES

- Abeyounis, C.J., and Milgrom, F. A thermostable antigen characteristic for carcinogen-induced rat intestinal tumors. *J. Immunol.* 116: 30-34, 1976.
- Accili, D., Perrotti, N., Rees-Jones, R., and Taylor, S.I. Tissue distribution and subcellular localization of an endogenous substrate (pp120) for the insulin receptor-associated tyrosine kinase. *Endocrinol.*, 119: 1274-1280, 1986.
- Afar, D.E.H., Stanners, C.P., and Bell, J.C. Tyrosine phosphorylation of biliary glycoprotein, a cell adhesion molecule related to carcinoembryonic antigen. *Biochim. Biophys. Acta*, 1134: 46-52, 1992.
- Aurivillius, M., Hansen, O.C., Lazrek, M.B.S., Bock, E., and Öbrink, B. The cell adhesion molecule cell-CAM 105 is an ecto-ATPase and a member of the immunoglobulin superfamily. *FEBS Letters*, 264: 267-269, 1990.
- Barnett, T., Drake, L., and Pickle, W. Human biliary glycoprotein gene: characterization of a family of novel alternatively spliced RNAs and their expressed proteins. *Mol. Cell. Biol.* 13: 1273-1282, 1993.
- Barnett, T., Kretschmer, A., Austen, D.A., Goebel, S.J., Hart, J.T., Elting, J.J., and Kamarck, M.E. Carcinoembryonic antigens: alternative splicing accounts for the multiple mRNAs that code for novel members of the carcinoembryonic antigen family. *J. Cell Biol.*, 108: 267-276, 1989.

Bartles, J.R., and Hubbard, A.L. Preservation of hepatocyte plasma membrane domains during cell division *in situ* in regenerating liver. *Dev. Biol.*, 118: 286-295, 1986.

Bartles, J.R., Feracci, H.M., Steiger, B., and Hubbard, A.L. Biogenesis of the rat hepatocyte plasma membrane in vivo: comparison of the pathways taken by apical and basolateral proteins using subcellular fractionation. *J. Cell Biol.*, 105: 1241-1251, 1987.

Bates, P.A., Luo, J., and Sternberg, M.J.E. A predicted three-dimensional structure for the carcinoembryonic antigen (CEA). *FEBS Letters*, 301: 207-214, 1992.

Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanners, C.P. Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. *Mol. Cell. Biol.*, 7: 3221-3230, 1987.

Beauchemin, N., Turbide, C., Afar, D., Bell, J., Raymond, M., Stanners, C.P., and Fuks, A. A mouse analogue of the human carcinoembryonic antigen. *Cancer Res.*, 49: 2017-2021, 1989a.

Beauchemin, N., Turbide, C., Huang, J.Q., Benchimol, S., Jothy, S., Shirota, K., Fuks, A., and Stanners, C.P. Studies on the function of carcinoembryonic antigen. In *The Carcinoembryonic Antigen Gene Family*, (A. Yachi and J.E. Shively, eds.), pp. 49-64. Amsterdam: Elsevier, 1989b.

Becker, A., Neumeier, R., Park, C.S., Gossrau, R., and Reutter, W. Identification of a transformation-sensitive 110-kDa plasma membrane glycoprotein of rat hepatocytes.

Eur. J. Cell Biol., 39: 417-423, 1985.

Behrens, J., Mareel, M., Van Roy, F., and Birchmeier, W. Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. J. Cell Biol., 108: 2435-2447, 1989.

Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shiota, K., and Stanners, C.P. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. Cell, 57: 327-334, 1989.

Berling, B., Kolbinger, F., Grunert, F., Thompson, J.A., Brombacher, F., Buchegger, F., von Kleist, S., and Zimmermann, W. Cloning of a carcinoembryonic antigen gene family member expressed in leukocytes of chronic myeloid leukemia patients and bone marrow. Cancer Res., 50: 6534-6539, 1990.

Blikstad, I., Wikström, T., Aurivillius, M., and Öbrink, B. C-CAM (cell-CAM 105) is a calmodulin binding protein. FEBS Letters, 302: 26-30, 1992.

Boucher, D., Cournoyer, D., Stanners, C.P., and Fuks, A. Studies on the control of gene expression of the carcinoembryonic antigen family in human tissue. Cancer Res., 49: 847-852, 1989.

Boyles, J.F., Weismiller, D.G., and Holmes K.V. Genetic resistance to mouse hepatitis virus correlates with absence of virus binding activity on target tissues. J. Virol., 61: 185-189, 1987.

Brattain, M.G., Strobel-Stevens, J., Fine, D., Webb, M., and Sarraf, A.M. Establishment of mouse colonic carcinoma cell lines with different metastatic properties. *Cancer Res.*, 40: 2142-2146, 1980.

Brenner, M., Trowbridge, I., and Strominger, J. Cross-linking of human T cell receptor proteins: association between the T cell idiotype beta subunit and the T3 glycoprotein of the heavy subunit. *Cell*, 40: 183-190, 1985.

Cedar, H. DNA-methylation and gene activity. *Cell*, 53: 3-4, 1988.

Cerni, C., Tatra, G., and Bohm, H. Immunosuppression by human placenta lactogen (HPL) and the pregnancy-specific β 1-glycoprotein (SP1): inhibition of mitogen-induced lymphocyte transformation. *Arch. Gynäkol.*, 223: 1-7, 1977.

Chan, W.Y. The pregnancy specific β 1-glycoprotein family. *Adv. Contra. Del. Sys.*, 7: 21-52, 1991

Chen, W., and Öbrink, B. Cell-cell contacts mediated by E-cadherin (uvomorulin) restrict invasive behavior of L-cells. *J. Cell Biol.*, 114: 319-327, 1991.

Cheung, W.Y. Calmodulin plays a pivotal role in cellular regulation. *Science (Wash. DC)*, 207: 19-27, 1980.

Chu, T.M., Reynoso, G., and Hansen, H.J. Demonstration of carcinoembryonic antigen in normal human plasma. *Nature*, 238:152-153, 1972.

Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, N.J., Rutter, W.J., and Kirschner, M.W. Number and evolutionary conservation of α - and β -tubulin and cytoplasmic β - and γ -actin genes using specific cloned cDNA probes. *Cell*, 20: 95-105, 1980.

Cross, G.A.M. Glycolipid anchoring of plasma membrane proteins. *Ann. Rev. Cell Biol.*, 6: 1-39, 1990.

Culic, O., Huang, Q.H., Flanagan, D., Hixson, D., and Lin, S.H. Molecular cloning and expression of a new rat liver cell-CAM 105 isoform. *Biochem. J.*, 285: 47-53, 1992.

Cunningham, B., Hemperley, J., Murray, B., Prediger, E., Brackenbury, R., and Edelman, G. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation and alternative splicing. *Science (Wash. DC)*, 236: 799-806, 1987.

Darnell, J., Lodish H., and Baltimore D. Principles of cellular organization and function. In *Molecular Cell Biology*, pp.163-165. New York: Scientific American Books Inc., 1986.

DeTomaso, A.W., Xie, Z.J., Liu, G., and Mercer, R.W. Expression, targetting, and assembly of functional Na,K-ATPase polypeptides in baculovirus-infected insect cells. *J. Biol. Chem.*, 268: 1470-1478, 1993.

Devereux, J., Haeberli, P., and Smithies, O. A comprehensive set of sequence analysis

programs for the VAX. Nucl. Acids Res., 12: 387-395, 1984.

Doerfler, W. DNA methylation and gene activity. Ann. Rev. Biochem., 52: 93-124, 1983.

Drzeniek, Z., Lamerz, R., Fenger, U., Wagener, C., and Haubeck, H.D. Identification of membrane antigens in granulocytes and colonic carcinoma cells by a monoclonal antibody specific for biliary glycoprotein, a member of the carcinoembryonic antigen family. Cancer Lett., 56: 173-179, 1991.

Dveksler, G.S., Dieffenbach, C.W., Cardellichio, C.B., McCuaig, K., Pensiero, M.N., Jiang, G.S., Beauchemin, N., and Holmes, K.V. Several members of the mouse carcinoembryonic antigen-related glycoprotein family are functional receptors for the coronavirus mouse hepatitis virus-A59. J. Virol., 67: 1-8, 1993.

Dveksler, G.S., Pensiero, M.N., Cardellichio, C.B., Williams, R.K., Jiang, G.S., Holmes, K.V., and Dieffenbach, C.W. Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. J. Virol., 65: 6881-6891, 1991.

Edelman, G.M. Morphoregulatory Molecules. Biochem., 27: 3533-3543, 1988.

Edelman, G.M., and Crossin, M. Cell molecules: implications for a molecular histology. Ann. Rev. Biochem., 60: 155-190, 1991.

Egans, M.L., Pritchard, D.G., Todd, C.W., and Go, V.L.W. Isolation and immunochemical and chemical characterization of carcinoembryonic antigen-like substances in colon lavages of healthy individuals. *Cancer Res.*, 37: 2638-2643, 1977.

Engvall, E., Vuento, M., and Ruoslahti, E. A monkey antigen crossreacting with carcinoembryonic antigen, CEA. *Br. J. Cancer*, 34: 341-344, 1976.

Fanciulli, M., Paggi, M.G., Mancini, A., Del Carlo, C., Floridi, A., Taylor, S.I., and Perrotti, N. pp120: a common endogenous substrate for insulin and IGF-1 receptor associated tyrosine kinase activity in the highly malignant AS-30D rat hepatoma cells. *Biochem. Biophys. Res. Comm.*, 160: 168-173, 1989.

Feinberg, A.P., and Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, 132: 6-13, 1983.

Feracci, H., Connolly, T.P., Margolis, R.N., and Hubbard, A.L. The establishment of hepatocyte cell surface polarity during fetal liver development. *Dev. Biol.*, 123: 73-84, 1987.

Franks, L.M., and Hemmings, V.J. A cell line from an induced carcinoma of mouse rectum. *J. Pathol.*, 124: 35-38, 1978.

Fritsche, R., and Mach, J.P. Isolation and characterization of carcinoembryonic antigen (CEA) extracted from normal human colonic mucosa. *Immunochemistry*, 14: 119-127, 1977.

Frixen, U.H., Behrens, J., Sachs, M., Eberle, G., Voss, B., Warda, A., Löchner, D., and Birchmeier, W. E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J. Cell Biol.*, 113: 173-185, 1991.

Frohman, M. A., Dush, M.K., and Martin G.R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U.S.A.*, 85: 8998-9001, 1988.

Garcia, M., Seigner, C. Bastid, C., Choux, R., Payan, M.J., and Reggio, H. Carcinoembryonic antigen has a different molecular weight in normal colon and cancer cells due to *N*-glycosylation differences. *Cancer Res.*, 51: 5679-5686, 1991.

Glenney, J.R., and Weber, K. Calmodulin-binding proteins of the microfilaments present in isolated brush borders and microvilli of intestinal epithelial cells. *J. Biol. Chem.*, 255: 10551-10554, 1980.

Glickman, L.T., Suissa, S., and Fleiszer, D.M. Proliferative characteristics of colonic crypt cells in C57BL/6J and A/J mice as predictors of subsequent tumor formation. *Cancer Res.*, 47: 4766-4770, 1987.

Gold, P., and Freedman, S.O. Specific carcinoembryonic antigens of the human digestive system. *J. Exp. Med.*, 122: 467-481, 1965.

Greve, J.M., Davis, G., Meyer, A.M., Forte, C.P., Connolly, S.Y., Marlbor, C.W., Kamarck, M.E., and McClelland, A. The major human rhinovirus receptor is ICAM-1.

Cell, 56: 839-847, 1989.

Haagensen, D.E.Jr., Metzgar, R.S., Swenson, B., Dilley, W.G., Cox, C.E., Davis, S., Murdoch, J., Zamchek, N., and Wells, S.A. Carcinoembryonic antigens in non-human primates. J. Natl. Cancer Inst., 69: 1073-1076, 1982.

Hansson, M., Blikstad, I., and Öbrink, B. Cell-surface location and molecular properties of cell-CAM 105 in intestinal epithelial cells. Exp. Cell Res., 181: 63-74, 1989.

Hauck, W., Nédellec, P., Turbide, C., Stanners, C.P., Barnett, T., and Beauchemin, N. Distinct transcriptional regulation of the human biliary glycoprotein gene, a CEA family member. Submitted, 1993.

Hauck, W., and Stanners, C.P. Identification of transcription factors and multiple interacting sites regulating the tissue-specific and differentiation-dependent expression of the carcinoembryonic antigen gene a human colon tumor marker. Submitted, 1993.

Hauck, W., and Stanners, C.P. Control of carcinoembryonic antigen gene family expression in a differentiating colon carcinoma cell line, Caco-2. Cancer Res., 51: 3526-3533, 1991.

Hefta, S.A., Hefta, L.J., Lee, T.D., Paxton, R., and Shively, J.E. Carcinoembryonic antigen is anchored to membranes by covalent attachment to a glycosylphosphatidylinositol moiety: Identification of the ethanolamine linkage site. Proc.

Natl. Acad. Sci. USA, 85: 4648-4652, 1988.

Hefta, S.A., Paxton, R.J., and Shively, J.E. Sequence and glycosylation site identity of two distinct glycoforms of nonspecific cross-reacting antigen as demonstrated by sequence analysis and fast atom bombardment mass spectrometry. *J. Biol. Chem.*, 265: 8618-8626, 1990.

Hemperly, J.J., Edelman, G.M., and Cunningham, B.A. cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a membrane-spanning region consistent with evidence for membrane attachment via a phosphatidylinositol intermediate. *Proc. Natl. Acad. Sci. U.S.A.*, 83: 9822-9826, 1986.

Hinoda, Y., Imai, K., Nakagawa, N., Ibayashi, Y., Nakano, T., Paxton, R.J., Shively, J.E., and Yachi, A. Transcription of biliary glycoprotein I gene in malignant and non-malignant human liver tissues. *Int. J. Cancer*, 45: 875-878, 1990.

Hinoda, Y., Neumaier, M., Hefta, S.A., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L.J.F., Shively, J.E., and Paxton, R.J. Molecular cloning of a cDNA coding biliary glycoprotein I: primary structure of a glycoprotein immunologically crossreactive with carcinoembryonic antigen. *Proc. Natl. Acad. Sci. U.S.A.* 85: 6959-6963, 1988.

Hixson, D.C., and McEntire, K.D. Detection of an altered form of cell-CAM 105 on rat transplantable and primary hepatocellular carcinomas. *Cancer Res.*, 49: 6788-6794, 1989.

Hixson, D.C., McEntire, K.D., and Öbrink, B. Alterations in the expression of a hepatocyte cell adhesion molecule by transplantable rat hepatocellular carcinomas. *Cancer Res.*, 45: 3742-3749, 1985.

Howe, C.L., Mooseker, M.S., and Graves, T.A. Brush-border calmodulin. *J. Cell Biol.*, 85: 916-923, 1980.

Huang, J.Q., Turbide, C., Daniels, E., Jothy, S., and Beauchemin, N. Spatiotemporal expression of murine carcinoembryonic antigen (CEA) gene family members during mouse embryogenesis. *Development*, 110: 573-588, 1990.

Hunter, T., and Cooper J.A. Protein-tyrosine kinases. *Ann. Rev. Biochem.*, 54: 897-930, 1985.

Hynes, R.O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell*, 69: 11-25, 1992.

Jennings, M.L. Topography of membrane proteins. *Ann. Rev. Biochem.* 58: 999-1027, 1989.

Jessup, J.M., and Thomas, P. Carcinoembryonic antigen: function in metastasis by human colorectal carcinoma. *Cancer Metastasis Rev.*, 8: 263-280, 1989.

Jothy, S., Brazinsky, S.A., Chin-A-Loy, M., Haggarty, A., Krantz, M.J., Cheung, M., and Fuks, A. Characterization of monoclonal antibodies to carcinoembryonic antigen

with increased tumor specificity. *Lab. Invest.*, 54: 108-117, 1986.

Kamarck, M.E., Elting, J.J., Hart, J.T., Goebel, S.J. Rae, P.M.M., Nothdurft, M.A., Nedwin, J.J., and Barnett, T.R. Carcinoembryonic antigen family: expression of a mouse L-cell transfectant and characterization of a partial cDNA in bacteriophage λ gt11. *Proc. Natl. Acad. Sci. U.S.A.*, 84: 5350-5354, 1987.

Khan, W. N., Teglund, S., Bremer, K., and Hammarström, S. The pregnancy-specific glycoprotein family of the immunoglobulin superfamily: identification of new members and estimation of family size. *Genomics*, 12: 780-787, 1992.

Kodelja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J.A., and Zimmermann, W. Identification of a carcinoembryonic antigen gene family in the rat. *J. Biol. Chem.*, 264: 6906-6912, 1989.

Koike, S., Horie, H., Ise, I., Okitsu, A., Yoshida, M., Iisuka, N., Takeuchi, K., Takegami, T., and Nomoto, A. The poliovirus receptor protein is produced both as membrane-bound and secreted forms. *EMBO J.*, 9: 3217-3224, 1990.

Koike, S., Ise, I., and Nomoto, A. Functional domains of the poliovirus receptor. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 4104-4108, 1991.

Kolbinger, F., Schwarz, K., Brombacher, F., von Kleist, S., and Grunert, F. Expression of an NCA cDNA in NIH/3T3 cells yields a 110K glycoprotein, which is anchored into the membrane via glycosyl-phosphatidylinositol. *Biochem. Biophys. Res.*

Commun., 161: 1126-1134, 1989.

Kupchik, H.Z., and Zamcheck, N. Carcinoembryonic antigen(s) in liver disease: isolation from human cirrhotic liver and serum and from normal liver. *Gastroenterology*, 63: 95-101, 1972.

Kuroki, M., Arakawa F., Matsuo Y., Oikawa, S. Nakazato H., and Matsuoka Y. Three novel molecular forms of biliary glycoprotein deduced from cDNA clones from a human leukocyte library. *Biochem. Biophys. Res. Comm.*, 176: 578-585, 1991.

Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)*, 227: 680-685, 1970.

Leff, S.E., Rosenfeld, M.G., and Evans, R.E. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Ann. Rev. Biochem.*, 55: 1091-1117, 1986.

Leusch, H.G., Drzeniek, Z., Markos-Pusztai, Z., and Wagener, C. Binding of *Escherichia coli* and *Salmonella* strains to members of the carcinoembryonic antigen family: differential binding inhibition by aromatic α -glycosides of mannose. *Infect. Immun.*, 59: 2051-2057, 1991.

Lin, S.H., and Guidotti, G. Cloning and expression of a cDNA coding for a rat liver plasma membrane ecto-ATPase. *J. Biol. Chem.*, 264: 14408-14414, 1989.

Lin, S.H., Culic, O., Flanagan, D., and Hixson, D. Immunochemical characterization of two isoforms of rat liver ecto-ATPase that show an immunological and structural identity with a glycoprotein cell-adhesion molecule with M_r 105 000. *Biochem. J.*, 278: 155-161, 1991.

Lin, T.-M., Halpert, S.P., and Spellacy, W.N. Measurement of pregnancy-associated plasma proteins during human gestation. *J. Clin. Invest.*, 54: 576-582, 1974.

Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A., and Axel, R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*, 47: 333-348, 1986.

Margolis, R.N., Schell, M.J., Taylor, S.I., and Hubbard, A.L. Hepatocyte plasma membrane ecto-ATPase (pp120/HA4) is a substrate for tyrosine kinase activity of the insulin receptor. *Biochem. Biophys. Res. Comm.*, 166: 562-566, 1990.

Margolis, R.N., Taylor, S.I., Seminara, D., and Hubbard, A.L. Identification of pp120, an endogenous substrate for the hepatocyte insulin receptor tyrosine kinase, as an integral membrane glycoprotein of the bile canalicular domain. *Proc. Natl. Acad. Sci. U.S.A.*, 85: 7256-7259, 1988.

Mattila, P., Korpela, J., Tenkänen, T., and Pitkänen, K. Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase. An extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.*, 19: 4967-4973, 1991.

McCabe, P.C. Production of single-stranded DNA by asymmetric PCR. In: Innis, M.A., Gelfand, D.H., Sninsky, J.I. and White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications pp. 76-83. Academic Press, San Francisco, CA, 1990.

McCuaig, K., Rosenberg, M., Nédellec, P., Turbide, C., and Beauchemin N. Expression of the Bgp gene and characterization of mouse colon biliary glycoprotein isoforms. *Gene*, in press, 1993.

McCuaig, K., Turbide, C., and Beauchemin, N. mmCGM1a: a mouse carcinoembryonic antigen gene family member, generated by alternative splicing, functions as an adhesion molecule. *Cell Growth and Diff.*, 3: 165-174, 1992.

McEntire, K.D., Mowery, J., and Hixson, D.C. Comparison of the structural characteristics of cell-CAM 105 from hepatocytes with those of an altered form expressed in rat transplantable hepatocellular carcinomas. *Cancer Res.*, 49: 6795-6802, 1989.

Meier, P.J. The bile salt secretory polarity of hepatocytes. *J. Hepatol.*, 9: 124-129, 1989.

Mendelsohn, C.L., Wimmer, E., and Racaniello, V.R. Cellular receptor for poliovirus: molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily. *Cell*, 56:855-865, 1989.

Mentzer, S., Smith, B., Barbosa, J., Crimmins, M.A.V., Herrmann, S.H., and Burakoff, S.J. CTL adhesion and antigen recognition are discrete steps in the human

CTL-target cell interaction. *J. Immunol.* 138: 1325-1330, 1987.

Mowery, J., and Hixson, D.C. Detection of cell-CAM 105 in the pericanalicular domain of the rat hepatocyte plasma membrane. *Hepatology*, 13: 47-56, 1991.

Najjar, S.M., Accili, D., Philippe, N., Jernberg, J., Margolis, R., and Taylor, S.I. pp120/ecto-ATPase, an endogenous substrate of the insulin receptor tyrosine kinase, is expressed as two variably spliced isoforms. *J. Biol. Chem.*, 268: 1201-1206, 1993.

Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A., and Shively, J.E. Characterization of a cDNA clone for the nonspecific cross-reacting antigen (NCA) and a comparison of NCA and carcinoembryonic antigen (CEA). *J. Biol. Chem.*, 263: 3202-3207, 1988.

Nowell, P.C. Mechanisms of tumor progression. *Cancer Res.*, 46: 2203-2207, 1986.

Öbrink, B. C-CAM (cell-CAM 105) - a member of the growing immunoglobulin superfamily of cell adhesion proteins. *BioEssays*, 13: 227-234, 1991.

Öbrink, B., Blikstad, I., Hansson, M., Odin, P., Svalander, P., and Tingström, A. Cell-CAM 105 - an adhesive cell surface glycoprotein. *Protoplasma*, 145: 182-187, 1988.

Ocklind, C., Forsum, U., and Öbrink, B. Cell surface localization and tissue distribution of a hepatocyte cell-cell adhesion glycoprotein (cell CAM 105). *J. Cell Biol.*, 96: 1168-1171, 1983.

Ocklind, C., and Öbrink, B. Intercellular adhesion of rat hepatocytes: identification of a cell surface glycoprotein involved in the initial adhesion process. *J. Biol. Chem.*, 257: 6788-6795, 1982.

Odin, P., Asplund, M. Busch, C. and Öbrink, B. Immunohistochemical localization of cell CAM 105 in rat tissues: appearance in epithelia, platelets, and granulocytes. *J. Histochem. Cytochem.*, 36: 729-739, 1988.

Odin, P., and Öbrink, B. Dynamic expression of the cell adhesion molecule cell-CAM 105 in fetal and regenerating rat liver. *Exp. Cell. Res.*, 164: 103-114, 1986.

Odin, P., Tingström, A., and Öbrink, B. Chemical characterization of cell-CAM 105, a cell-adhesion molecule isolated from rat liver membranes. *Biochem. J.*, 236: 559-568, 1986.

Ogilvie, S., Kvello-Stenstrom, A.G., Hammond, G., Buhi, W.C., Larkin, L.H., and Shiverick, K.T. Identification of proteins immunochemically related to human pregnancy-specific β_1 -glycoprotein in the rat placenta. *Endocrinology*, 125: 287-294, 1989.

Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G., and Nakazato, H. The carcinoembryonic antigen (CEA) contains multiple immunoglobulin-like Ig domains. *Biochem. Biophys. Res. Comm.*, 144: 634-642, 1987a.

Oikawa, S., Inusuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G. and Nakazato, H. Cell adhesion of non-specific cross-reacting antigen (NCA) and carcinoembryonic antigen

(CEA) expressed on CHO cell surface: homophillic and heterophillic adhesion. *Biochem. Biophys. Res. Comm.*, 164: 39-45, 1989.

Oikawa, S., Inuzuka, C., Kuroki, M., Arakawa, F., Matsuoka, Y., Kosaki, G. and Nakazato, H. A specific heterotypic cell adhesion activity between members of carcinoembryonic antigen family, W272 and NCA, is mediated by N-domains. *J. Biol. Chem.* 266: 7995-8001, 1991.

Oikawa, S., Kuroki, M., Matsuoka, Y., Kosaki, G., and Nakazato, H. Homotypic and heterotypic Ca^{++} -independent cell adhesion activities of biliary glycoprotein, a member of carcinoembryonic antigen family, expressed on CHO cell surface. *Biochem. Biophys. Res. Comm.*, 186: 881-887, 1992.

Oikawa, S., Nakazota, H., and Kosaki, G.. Primary Structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequences. *Biochem. Biophys. Res. Commun.* 142: 511-518, 1987b.

Parker, B.A., and Stark, G.R. Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present in early infections by virus or viral DNA. *J. Virol.*, 31: 360-369, 1979.

Peterson, G.L. A simplification of the protein assay method of Lowry *et al.* which is more generally applicable. *Anal. Biochem.*, 83: 346-356, 1977.

Phillips, S.A., and Taylor, S.I. Rat liver membranes contain a 120 kDa glycoprotein

which serves a substrate for the tyrosine kinases of the receptors for insulin and epidermal growth factor. *FEBS Let.*, 212: 141-144, 1987.

Pollard, S.R., Meier, W., Chow, P., Rosa, J.J., and Wiley, D.C. CD4-binding regions of human immunodeficiency virus envelope glycoprotein gp120 defined by proteolytic digestion. *Proc. Natl. Acad. Sci. U.S.A.*, 88: 11320-11324, 1991.

Proudfoot, N.J., and Brownlee, G.G.: 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature*, 263: 211-214, 1976.

Rebstock, S., Lucas, K., Thompson, J.A., and Zimmermann, W. cDNA and gene analyses imply a novel structure for a rat carcinoembryonic antigen-related protein. *J. Biol. Chem.*, 265: 7872-7879, 1990.

Rees-Jones, R.W. and Taylor, S.I. An endogenous substrate for the insulin receptor-associated tyrosine kinase. *J. Biol. Chem.*, 260: 4461-4467, 1985.

Register, R.B., Uncapher, C.R., Naylor, A.M., Linerberger, D.W., and Colonno, R.J. Human-murine chimeras of ICAM-1 identify amino acid residues critical for rhinovirus and antibody binding. *J. Virol.*, 65: 6589-6596, 1991.

Robbins, J., Robbins, P.F., Kozak, C.A., and Callahan, R. The mouse biliary glycoprotein gene (*Bgp*): partial nucleotide sequence, expression, and chromosomal assignment. *Genomics*, 10: 583-587, 1991.

Rojas, M., Fuks, A., and Stanners, C.P. Biliary glycoprotein, a member of the immunoglobulin supergene family, functions *in vitro* as a Ca^{2+} -dependent intercellular adhesion molecule. *Cell Growth and Diff.*, 1: 527-533, 1990.

Rosenberg, M., Nédellec, P., Jothy, S., Fleiszer, D., Turbide, C., and Beauchemin, N. The expression of mouse biliary glycoprotein, a carcinoembryonic antigen-related protein, is downregulated in malignant tissues. Submitted for publication to the journal *Cancer Research*.

Rudert, F., Saunders, A.M., Rebstock, S., Thompson, J.A., and Zimmermann, W. Characterization of murine carcinoembryonic antigen gene family members. *Mammalian Genome*, 3: 262-273, 1992.

Ruthishouser, U., and Jessel, T.M. Cell adhesion molecules in vertebrate neural development. *Physiological Rev.*, 68: 819-857, 1988.

Salzer, J.L., Holmes, P.W., and Colman, D.R. The amino acid sequence of the myelin-associated glycoproteins: homology to the immunoglobulin gene superfamily. *J. Cell Biol.*, 104: 957-965, 1987.

Sambrook, J., Fritsch, E.F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd Ed., New York: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.

Sanger, F., Nicklen, S., and Coulson, A.R. DNA sequencing with chain-terminating

inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74: 5463-5467, 1977.

Schrewe, H., Thompson, J., Bona, M., Hefta, L.J.F., Maruya, A., Hassauer, M., Shively, J.E., von Kleist, S., and Zimmermann, W. Cloning of the complete gene for carcinoembryonic antigen: analysis of its promoter indicates a region conveying cell type-specific expression. *Mol. Cell. Biol.*, 10: 2738-2748, 1990.

Shively, J.E., and Beatty, J.D. CEA-related antigens: molecular biology and clinical significance. *CRC Crit. Rev. Oncol. Hematol.*, 2: 355-399, 1985.

Shuster, J.D., Thompson, M.P., Fuks, A., and Gold, P. Immunologic approaches to diagnostic of malignancies. *Prog. Exp. Tumor Res.*, 25: 89-139, 1980.

Sippel, C.J., Suchy, F.J., Ananthanarayanan, M., and Perlmutter, D.H. The rat liver ecto-ATPase is also a canalicular bile acid transport protein. *J. Biol. Chem.*, 268: 2083-2091, 1993.

Smith, D.B., and Johnson, K.S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene*, 67: 31-40, 1988.

Snow, P., Van de Rijn, M., and Terhorst, C. Association between the human thymic differentiation antigens T6 and T8. *Eur. J. Immunol.*, 15: 529-532, 1985.

Sommers, C.L., Thompson, E.W., Torri, J.A., Kemler, R., Gelmann, E.P., and Byers, S.W. Cell adhesion molecule uvomorulin expression in breast cancer cell lines:

relationship to morphology and invasive capacities. *Cell Growth and Diff.*, 2: 365-372, 1991.

Southern, P.J., and Berg, P. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under the control of the SV40 early region promoter. *J. Mol. Appl. Genet.*, 1: 327-314, 1982.

Stanners, C.P., Rojas, M., Zhou, H., Fuks, A. and Beauchemin, N. The CEA family: a system in transitional evolution ? *The Int. J. Biol. Markers*, 7: 137-142, 1992.

Staunton, D.E., Merluzzi, V.J., Rothlein, R., Barton, R., Marlin, S.D., and Springer, T.A. A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell*, 56: 849-853, 1989.

Staunton, D.E., Dustin, M.L., Erickson, H.P., and Springer, T.A. The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell*, 61: 243-254, 1990.

Svenberg, T. Carcinoembryonic antigen-like substances of human bile: isolation and partial characterization. *Int. J. Cancer*, 17: 588-596, 1976.

Svenberg, T., Hammarström, S., and Hedin, A. Purification and properties of biliary glycoprotein 1 (BGP 1). Immunochemical relationship to carcinoembryonic antigen. *Mol. Immunol.*, 16: 245-252, 1979a.

Svenberg, T., Hammarström, S., Zeromski, J. Immunofluorescence studies on the occurrence and localization of the CEA-related biliary glycoprotein I (BGP I) in normal human gastrointestinal tissues. *Clin. Exp. Immunol.*, 36: 436-441, 1979b.

Svenberg, T., Wahren, B., and Hammarström, S. Elevated serum levels of a biliary glycoprotein (BGPI) in patients with liver or biliary tract disease. *Clin. Exp. Immunol.*, 36: 317-325, 1979c.

Takami, N., Misumi, Y., Kuroki, M., Matsuoka, Y., and Ikehara, Y. Evidence of carboxyl-terminal processing and glycolipid-anchoring of human carcinoembryonic antigen. *J. Biol. Chem.*, 263: 12716-12720, 1988.

Takeichi, M. Cadherin cell adhesion receptors as a morphogenic regulator. *Science (Wash. DC)*, 251: 1451-1455, 1991.

Thompson, J.A., Grunert, F., and Zimmermann, W. Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. *J. Clin. Lab. Analys.*, 5: 344-366, 1991.

Thompson, J., Koumari, R., Wagner, K., Barnert, S., Schleussner, C., Schrewe, H., Zimmermann, W., Muller, G., Zanietta, D., Ammaturo, D., and Hardman, N. The human pregnancy-specific glycoprotein genes are tightly linked on the lon arm of chromosome 19 and are coordinately expressed. *Biochem. Biophys. Res. Commun.*, 167: 848-859, 1990.

Thompson, J., and Zimmermann, W. The carcinoembryonic antigen gene family: structure, expression and evolution. *Tumor Biol.*, 9: 63-83, 1988.

Thompson, J., Zimmermann, W., Osthus-Bugat, P., Schleussner, C., Eades-Perner, A.M., Barnert, S., von Kleist, S., Willcocks, T., Craig, I., Tynan, K., Olsen, A., and Mohrenweiser, H. Long-range chromosomal mapping of the carcinoembryonic antigen (CEA) gene family cluster. *Genomics*, 12: 761-772, 1992.

Tingström, A., Blikstad, I., Aurivillius, M., and Öbrink, B. C-CAM (Cell-CAM 105) is an adhesive cell surface glycoprotein with homophilic binding properties. *J. Cell Sci.*, 96: 17-25, 1990.

Towbin, H., Staehelin, T., and Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A.*, 76: 4350-4354, 1979.

Tran, R., Kashmiri, S.V.S., Kantor, J., Greiner, J.W., Pestka, S., Shively, J.E., and Schlom, J. Correlation of DNA hypomethylation with expression of carcinoembryonic antigen in human colon carcinoma cells. *Cancer Res.*, 48: 5674-5679, 1988.

Tse, A.G.D., Barclay, A.N., Watts, A., and Williams, A.F. A glycopospholipid tail at the carboxyl-terminus of the Thy-1 glycoprotein of neurons and thymocytes. *Science (Wash. DC)*, 230: 1003-1008, 1985.

Turbide, C., Rojas, M., Stanners, C.P., and Beauchemin, N. A mouse

carcinoembryonic antigen gene family member is a calcium-dependent cell adhesion molecule. *J. Biol. Chem.*, 266: 309-315, 1991.

Ullrich, A., and Schlessinger, J. Signal transduction by receptors with tyrosine kinase activity. *Cell*, 61: 203-212, 1990.

Vleminckx, K., Vakaet, L., Mareel, M., Fiers, W., and Van Roy, F. Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role. *Cell*, 66: 107-119, 1991.

Wanebo, H.J., Gallmeier, W.M., Boyse, E.A., and Old, L.J. Paraproteinemia and reticulum cell sarcoma in an inbred mouse strain. *Science*, 154: 901-903, 1966.

Watanabe, S., and Chou, J.Y. Isolation and characterization of complementary DNAs encoding human pregnancy-specific β 1-glycoprotein. *J. Biol. Chem.*, 263: 2049-2054, 1988.

Wege, H., Siddell, St., and ter Meulen, V. The biology and pathogenesis of coronaviruses. *Curr. Top. Microbiol. Immunol.*, 99: 165-200, 1982.

White, J.M., and Littman, D.R. Viral receptors of the immunoglobulin superfamily. *Cell*, 56: 725-728, 1989.

Williams, A.F. A year in the life of the immunoglobulin superfamily. *Immunol. Today*, 8:298-303, 1987.

Williams, A.F., and Barclay, A.N. The immunoglobulin superfamily: domains for cell surface recognition. *Annu. Rev. Immunol.*, 6: 381-405, 1988.

Williams, R.K., Jiang, G.S., Snyder, S.W., Frana, M.F., and Holmes, K.V. Purification of the 110-kilodalton glycoprotein receptor for mouse hepatitis virus (MHV)-A59 from mouse liver and identification of a nonfunctional homologous protein in MHV-resistant SJL/J mice. *J. Virol.*, 64: 3817-3823, 1990.

Wong, G.G., Witek, J.S., Temple, P.A., Wilkins, K.M., Leary, A.C., Luxenberg, D.P., Jones, S.S., Brown, E.L., Kay, R.M., Orr, E.C., Schoemaker, C., Golde, D.W., Kaufman, R.J., Huwick, R.M., Wang, E.A., and Clark, S.C. Human M-CSF: molecular cloning of the complementary DNA and purification of the natural and recombinant proteins. *Science (Wash. DC)*, 228: 810-814, 1985.

Zhou, H., Fuks, A. and Stanners, C.P. Specificity of intercellular adhesion mediated by various members of the immunoglobulin supergene family. *Cell Growth and Diff.*, 1: 209-215, 1990.

Zimmermann, W., Ortlieb, R., Friedrich, R., and von Kleist, S. Isolation and characterization of cDNA clones encoding the human carcinoembryonic antigen reveal a highly conserved repeating structure. *Proc Natl. Acad. Sci. U.S.A.*, 84:2960-2964, 1987.