This is to certify that the

dissertation entitled

IDENTIFICATION AND CHARACTERIZATION OF CARBOHYDRATE BINDING PROTEIN 35 GENE STRUCTURE

presented by

SHIZHE JIA

has been accepted towards fulfillment
of the requirements for

DOCTOR OF PHILOSOPHY degree in BIOCHEMISTRY

Major professor

Date 3/13/91

IDENTIFICATION AND CHARACTERIZATION OF CARBOHYDRATE BINDING PROTEIN 35 GENE STRUCTURE

BY

SHIZHE JIA

A DISSERATATION

SUBMITTED TO

MICHIGAN STATE UNIVERSITY

IN PARTIAL FULFILLMENT OF REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

1990

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF CARBOHYDRATE BINDING PROTEIN 35 GENE STRUCTURE

By

Shizhe Jia

Carbohydrate Binding Protein 35 (CBP35) is a galactose-specific lectin identified in the cytoplasm and nucleus of mouse 3T3 fibroblasts, as well as a number of other tissues and cell types. Affinity purified antibodies directed against CBP35 were used to screen a lambda gt11 expression library derived from mRNA of mouse 3T3 fibroblasts. One positive clone containing cDNA for CBP35 was characterized by expression of a fusion protein containing beta-galactosidase and CBP35 sequences. Limited proteolysis of bacterial lysates containing the fusion protein, followed by SDS polyacrylamide gel electrophoresis and immunoblotting with anti-CBP35, yielded a peptide mapping pattern comparable to that obtained from parallel treatment of authentic CBP35. Such a limited proteolysis followed by affinity chromatography on a column of Sepharose derivatized with galactose also yielded a 30-kDa polypeptide that exhibited carbohydrate-binding activity. This polypeptide can be immunoblotted with anti-CBP35, but not with antibodies directed against beta-galactosidase. These results indicate that the positive clone is be an authentic CBP35 cDNA clone.

The complete nucleotide sequence of the cDNA clone for CBP35 has been determined. The deduced amino acid sequence showed that the

protein consists of two domains: (a) an amino terminal portion that contains an internal sequence homology featured by 8 repeats of a 9 amino acid motif which is rich in proline and glycine residues and which shows structure similarity to certain regions of proteins of the heterogeneous nuclear ribonucleoprotein complex (hnRNP); and (b) a carboxyl terminal portion that is homologous to beta-D-galactoside specific lectins isolated from a number of animal tissues.

A cloned genomic DNA segment for CBP35 gene was isolated . Nucleotide sequence analysis of the cloned gene, as well as the results of genomic Southern blot hybridization revealed that the gene is unique and spans 9 kilobases of genomic DNA. The mRNA for the lectin is encoded by five exons separated by four introns. Intron I must necessarily be in the 5′ transcribed but untranslated region of the primary transcript. Examination of the nucleotide sequence 5′ to the transcription initiation site revealed characteristic TATA and CCAAT boxes. A putative serum-responsive element has also been identified about 200 nucleotides upstream from the site for initiation of transcription. This may serve as a binding site for a class of transcription factors responsible for the serum-stimulated expression of the CBP35 gene.

DEDICATED TO MY PARENTS

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Dr. John Wang for all his encouragement and support. He has provided me with opportunities such that I can become more mature during my graduate career.

I would like to thank my committee, Drs. Lee McIntosh, Arnold Revzin, Leonard Robbins and John Wilson for their assistance during my education.

I thank Drs. Richard Anderson, Melvin Schindler, John Ho, Ioannis Moutsatsos for their advice and discussion. I am grateful to my laboratory mates for their cooperation and criticisms: Patty Voss, Kristen Yang, Sung-Yuan Wang, Neera Agrwal, Kim Hamann, Jamie Laing and Dr. Liz Cowles.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASGP: asialoglycoprotein

beta-galase: beta-galactosidase

bp: base pair(s)

CBP: carbohydrate binding protein

CLL-I: Chicken lactose lectin I

CLL-II: Chicken lactose lectin II

CRD: carbohydrate recognition domain

eBP: IgE binding protein

EDTA: (ethylenedinitrilo)-tetraacetic acid

Gal: galactose

hnRNP: heterogenous nuclear ribonucleoprotein complex

HRP: horseradish peroxidase

IPTG: isopropyl-beta-D-thio-galactopyranoside

IVS: Intervening sequence

kb: kilobase(s)

kD: kilodalton(s)

LBP35: laminin binding protein 35

L-34: mouse lectin 34

M-6-P: Mannose 6-phosphate

Mac-2: mouse macrophage surface antigen

PAGE: polyacrylamide gel electrophoresis

PMSF: phenylmethyl sulfonyl fluoride

SDS: sodium dodecyl sulfate

SRE: serum response element

TE: 125mM Tris-HCl, 1mM EDTA, pH 6.8

Tris: tris (hydroxymethyl) aminoethane

CHAPTER I

LITERATURE REVIEW

I INTRODUCTION

Carbohydrate binding proteins bind to saccharide-containing regions of glycoproteins and glycolipids. Lectins are carbohydrate binding proteins that were originally identified in plant extracts as agglutinins of erythrocytes (1). In the more recent literature, lectin and carbohydrate binding protein have become interchangeable terms, and these will be used similarly in this thesis. Although lectins were originally identified in plants, they have also been found in many other organisms including bacteria, slime molds and vertebrates (2). In this thesis, the gene structure of an animal lectin, designated Carbohydrate Binding Protein 35 (CBP35), will be described. The literature review will therefore focus on certain structural themes of animal lectins.

II ANIMAL LECTINS : STRUCTURAL MOTIFS

A: CLASSIFICATION

There are two main classification systems for animal lectins. The first system classifies the proteins according to solubility (Table I): a) integral membrane lectins which require detergents for their solubilization; and b) lectins soluble in aqueous buffer (17).

Key examples of the membrane bound lectin group include the asialoglycoprotein (ASGP) receptors (3, 4) that may function to clear serum glycoproteins and the mannose 6-phosphate (M6P) receptors that facilitate the transport of lysosomal enzymes from the golgi to their target organelles (5, 6). The soluble lectins in turn fall into two major families, one with molecular weights of 14,000-16,000 (14 kD -16 kD) and a second with molecular weights of 29,000 -35,000 (29 kD-35 kD).

Another classification of animal lectins is based on the dependence of the carbohydrate binding activity on cations (C-type lectins) or on thiols (S-type lectins) (18). The carbohydrate binding activity of C-type lectins is Ca^{2+}-dependent, while the S-type lectins are thiol-dependent for the activity. Table II summarizes the distinguishing features of the two types of lectins (18). From numerous structural studies, it is now clear that both C-type and S-type lectins consist of two different domains. The first domain, if one exists, is termed the special effector domain; this confers unique properties on the particular lectin. The second domain is the carbohydrate recognition domain (CRD).

B: C-TYPE LECTINS

Figure 1 is a summary of the structural features of C-type lectins (18). An important feature of the structure is that all these lectins have a CRD of 130 amino acid residues, in which 18 amino acid residues are conserved (20). Of particular interest are the four Cys residues that are involved in disulfide bonds. It should be noted that the carbohydrate binding specificities of different C-type lectins differ; some are specific for galactose (e.g. ASGP receptor)

Table I Examples of Membrane Bound and Soluble Lectins

CLASSIFICATION	LECTIN	SUBUNIT MW	LIGAND	SOURCE	REFERENCE
Membrane bound	ASGP receptor	26 K	GlcNAc	Chicken	3
		52 K	D-Gal	Rabbit	4
	M-6-P receptor	46 K	Man-6-P	Bovine	5
		225K	Man-6-P	Human	6
Soluble	CLL-I/CLL-II	14-16K	D-Gal	Chicken	7-9
	L-34 /CBP35 Mac-2/LBP35	29-35 K	D-Gal	Mouse	14-16
	eBP	31K	D-Gal	Rat	36

Table II Comparison of C-type and S-type Animal Lectins

Property	C-type lectins	S-type lectins
Ca^{2+} requirement	Yes	No
State of cysteines	Disulfides	Free thiols
Solubility	Variable	Aqueous solutions
Location	Extracellular	Intracellular and extracellular
Carbohydrate specificity	Various	Mostly B-galactosides

Figure 1. **Summary of the structural features of C-type animal lectins.** The invariant residues found in the common carbohydrate-recognition domain of the C-type lectins are shown, flanked by schematic diagrams of the special effector domains (if any) found in individual members of the family . EGF, epidermal growth factor; GAG, glycosaminoglycan. (adapted from ref. 18)

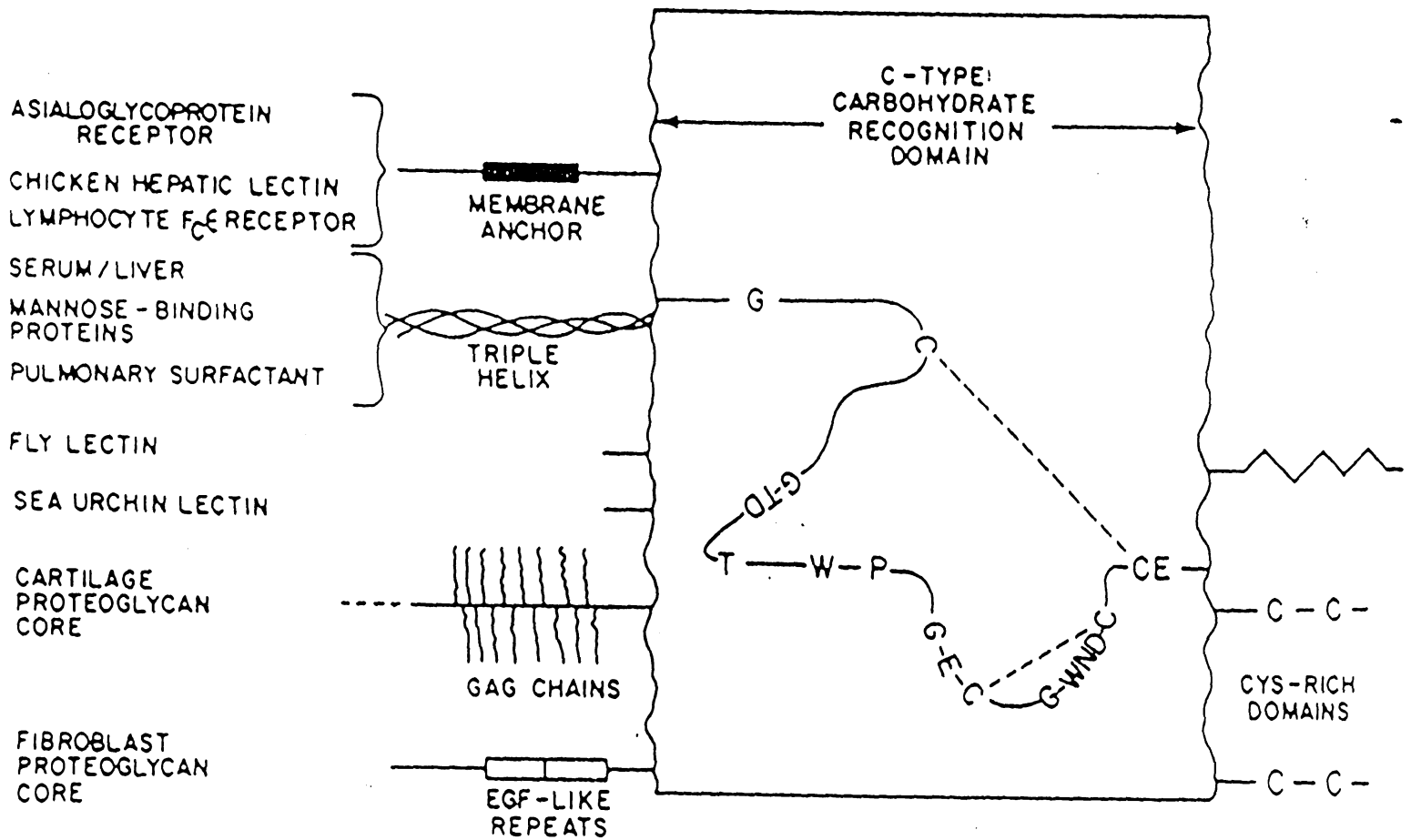

ASIALOGLYCOPROTEIN RECEPTOR
CHICKEN HEPATIC LECTIN
LYMPHOCYTE FcεRECEPTOR
SERUM/LIVER
MANNOSE-BINDING PROTEINS
PULMONARY SURFACTANT
FLY LECTIN
SEA URCHIN LECTIN
CARTILAGE PROTEOGLYCAN CORE
FIBROBLAST PROTEOGLYCAN CORE
MEMBRANE ANCHOR
TRIPLE HELIX
GAG CHAINS
EGF-LIKE REPEATS
C-TYPE CARBOHYDRATE RECOGNITION DOMAIN
G
C
G-I-D
T
W
P
CE
G-E-C
G-WND-C
C - C -
CYS-RICH DOMAINS
C - C -

Figure 1

while others are specific for mannose. Thus, the conservation of amino acids in the CRD is for carbohydrate-binding, not necessarily for a specific saccharide.

A second feature of the structure is that different lectins in the C-type category have distinct effector domains. For example, the ASGP receptor described earlier is a membrane bound lectin; it therefore has a membrane anchor as its effector domain (Fig. 1). Similarly, the cartilage proteoglycan core protein has a CRD that is linked to an effector domain to which many glycosaminoglycans are attached. Finally, certain C-type lectins, such as the fly lectin and the sea urchin lectin, contain only the CRD; they lack the effector domain.

Thus, although the exact functions of all of the C-type lectins are still not entirely understood, the structure information of this class of lectins is relatively clear. In fact, some of the structural information on the effector domain provides clear cut notions of their overall function. This theme is also shared by the domain delineation in the S-type lectins.

C: S-TYPE LECTINS

In S-type lectins, there are two groups of proteins: 14 -16 kD and 29-35 kD lectins. The 14-16 kD lectin group exhibits only one domain, the CRD. In addition to the CRD, the 29-35 kD lectins group exhibits a second domain. These features are schematically illustrated in Figure 2.

Like the C-type lectins, the CRD of S-type lectin exhibits

Figure 2. **Summary of the structural features of S-type animal lectins.** Conserved residues found in all of the family members so far sequenced are shown. In addition , the extra domain found in CBP35 is shown schematically.

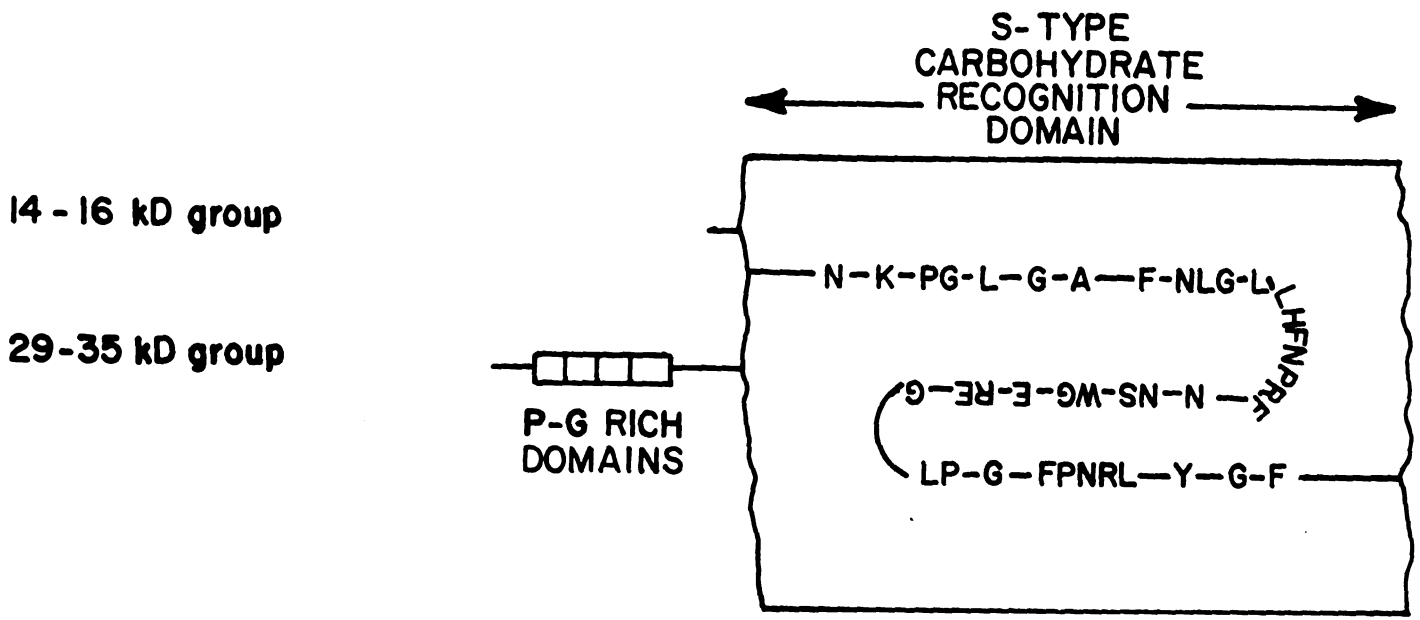

14 - 16 kD group
29-35 kD group
S-TYPE
CARBOHYDRATE
RECOGNITION
DOMAIN
N-K-PG-L-G-A—F-NLG-L
LHFNPRF
N-NS-WG-E-RE-G
LP-G-FPNRL—Y-G-F
P-G RICH
DOMAINS

Figure 2

conservation of certain residues, implying that these residues are necessary for carbohydrate binding. The 39 conserved amino acid pattern contains several charged residues. The CRD of S-type lectins is completely different from the CRD of C-type lectins. Unlike the C-type CRDs, the Cys residues are not conserved in the S-type CRDs. As indicated earlier, different CRDs of C-type lectins bind to different saccharides (Fig. 1). In contrast, all of the CRDs of S-type lectins thus far identified are specific for galactose/lactose-containing glycoconjugates.

At present, there has been no definitive identification of the function of any of the S-type lectins. The structural results, however, do provide ideas concerning the function of these proteins that can be tested. This will be the subject of more detailed discussion in the context of the structure of CBP35.

III CARBOHYDRATE BINDING PROTEIN 35

A: ISOLATION AND CHARACTERIZATION

CBP35 was initially isolated from the extracts of cultured 3T3 fibroblasts by fractionation on Sepharose columns covalently derivatized with asialofetuin (16). Three CBPs, with molecular weights of 35 kD, 16 kD and 13.5 kD, respectively, were purified from the columns. CBP35 was further purified from the other two CBPs by Sephadex G-150 chromatography. When purified CBP35 was subjected to sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) analysis, only one band was observed at 35 kD. CBP35 exhibits the following features: (a) it has agglutination activity when assayed with rabbit erythrocytes; (b) it specifically recognizes

galactose-containing glycoconjugates; and (c) it requires b-mercaptoethanol to retain carbohydrate binding activity, which is a characteristic of the S-type lectins.

B: DISTRIBUTION

Purified CBP35 was injected into rabbits to produce anti-CBP35 antibodies. The antibodies directed against CBP35 did not cross-react with the 13.5 kD and 16 kD lectins from 3T3 cells. When the extracts of 3T3 cells were subjected to SDS PAGE and immunoblotting analysis, only one band in the 35 kD position was observed (23). Using the antibodies directed against CBP35, the distribution of CBP35 was studied both in various tissues and in 3T3 cells (21, 23).

CBP35 was found in human, chicken and mouse fibroblasts and in a macrophage-like cell line, P388D1. In a study of the distribution of CBP35 in various tissues, both embryonic and adult tissues were searched for cross-reactivity with anti-CBP35. A cross-reactive polypeptide of about 35 kD was found in liver, lung, thymus, skin and muscle of embryonic tissues and in the lung, artery, thymus, and spleen of the adult mouse.

In 3T3 cells, the majority (>95%) of the CBP35 is found inside the cell. Although CBP35 could be detected at the cell surface (e. g. by anti-CBP35 antibodies), the interpretation of such results has not been clear cut. Much of these extracellular CBP35 could be due to cell lysis and leaking of intracellular lectins, which become bound to the cell surface. Therefore, most of the attention has been focused on intracellular CBP35.

The results from immunofluorescence studies with anti-CBP35 antibodies on formaldehyde-fixed and detergent permeabilized 3T3 cells have been particularly interesting. First, the lectin was predominantly localized in the nucleus of the proliferating 3T3 cells, while in quiescent 3T3 cultures the majority of the CBP35 was found in the cytoplasm. Second, CBP35 had a distinct punctate staining in the 3T3 cell nucleus (21). This indicated that CBP35 may be associated with some subnuclear structure.

C: NUCLEAR LOCALIZATION OF CBP35

Immunochemical studies have shown that CBP35 may not be present in a free form in 3T3 cells but rather may be associated with the heterogeneous nuclear ribonucleoprotein complex (hnRNP) in the nucleus (27) and ribonucleoprotein complexes (RNPs) in cytoplasm (Laing, J., and Wang, J., L. unpublished data).

When unfixed and detergent-permeabilized 3T3 cells were digested with ribonuclease A (RNase A) and deoxyribonuclease I (DNase I), only RNase A could release CBP35 from the nucleus. This indicated that the nuclear CBP35 may be associated with the ribonucleoprotein fraction in the nuclei. When subnuclear fractions were isolated by sucrose or cesium sulfate gradient centrifugation, CBP35 was found in the same fractions as hnRNP. Moreover, when nucleoplasm was subjected to affinity chromatography on a galactose-Sepharose column, the bound fraction eluted with lactose contained CBP35 as well as polypeptides corresponding to those identified in hnRNP (27). This was interpreted to indicate that CBP35 in the complex bound to the column which in turn co-purified the hnRNP components as a complex.

Since CBP35 was identified in the hnRNP complex, it is helpful to summarize some literature on hnRNP, particularly in the structural study area. In eukaryotic cells, nascent RNA polymerase II transcripts are rapidly assembled into ribonucleoprotein structures called heterogeneous nuclear ribonucleoprotein (hnRNP) complexes (41). The mammalian cell has a 40S hnRNP complex and is most likely a key structure in hnRNA processing : splicing, packaging and transporting out of the nucleus (42). The complex contains at least 20 polypeptides. The protein moiety of hnRNP particle has been characterized and shown to contain six major polypeptides, called "core proteins": A1/A2, B1/B2, C1/C2 (41).

The cDNA sequences of an hnRNP protein in Artemia salina (shrimp) (GRP-33) (44), the rat A1 protein (45) and human hnRNP proteins A1, A2, B1, C1 and C2 (46), have been determined. The hnRNP protein sequences deduced from the cDNA sequences have the following features: (a) the distribution of Gly and Pro residues in the polypeptide chains is nonuniform: for both GRP-33 and rat hnRNP A1 protein, approximately 77% of the Gly residues are located in the carboxyl-terminal portion of the protein while in the human C1 hnRNP protein, all the Pro residues are found in the amino-terminal half of the polypeptide chain (47); (b) there are domains containing consensus sequences characteristic of the RNA binding protein family (47); (c) some hnRNP proteins are almost identical to each other. For example, A2 and B1 cDNA are identical except for a 36 nucleotide in frame insert in B1, while C2 contains an extra 39 in frame nucleotides compared to C1 cDNA. The last feature may indicate that some of the diversity apparent among the hnRNP proteins is due to alternative hnRNP pre-mRNA splicing (46, 48).

D: REGULATION OF EXPRESSION

From quantitative immunoblotting, it was documented that the level of CBP35 was 3.5-fold higher in sparse proliferating 3T3 cells than in quiescent confluent monolayers of the same cells. When quiescent 3T3 cells were stimulated by the addition of serum, the level of CBP35 increased during the early part of G1, well before the onset of the S-phase of the cell cycle (21). Consistent with these results at the protein level, analysis of the expression of the CBP35 gene showed that there is elevation of both transcription and accumulation of mRNA for CBP35 in proliferating cells.

Using Northern blotting and nuclear transcription run-off assays, it was demonstrated that: (a) the mRNA level was increased as early as 30 minutes, and reached a peak at about 3 hours, after the addition of serum to stimulate the cells; (b) the rate of transcription of the CBP35 gene was elevated within 3 hours and reached its maximal level at about 10 hours after the serum stimulation; and (c) the elevation of the transcription rate of the CBP35 gene is not affeced by the presence of 10 ug/ml cycloheximide, suggesting that the increased transcription is a primary event of serum stimulation (24). Thus , the regulation of the expression of the CBP35 gene appears to be similar to that of other growth regulated genes, including c-fos, c-myc and actin genes (25,26). One might predict that there may be some structural similarity in the regulatory regions of CBP35 and those genes.

c-Fos and actin genes are serum stimulated genes. One mechanism for regulation of the expression of these genes involves the serum response element (SRE), which is present in the 5′ flanking regions

of both genes (39, 40). Using the serum stimulation assay to test the effects of different deletions of 5′ flanking regions of the human c-fos gene, Treisman defined the SRE to be between nucleotides -332 to -276 relative to the mRNA cap site (39). A similar SRE was also found in the actin promoter region, -94 to -75 nucleotides from the transcriptional start site (40). The SREs identified show the same consensus sequence: C- C- A- A/T- A- T- A/T- A/T- G- G. Both SRE sequences appear to bind the same serum response factors (40).

IV IDENTITY BETWEEN CBP35 AND PROTEINS STUDIED UNDER OTHER NAMES

From independent structural determinations, it is clear now that a protein corresponding to CBP35 has been identified and purified in several different laboratories. In many of these cases, the basis for the interest in the protein and for the purification protocol were totally independent of carbohydrate-binding activity. Thus, CBP35 exists in the literature under several names. These include: (a) lectin 34 (L-34) from a mouse fibrosarcoma (29); (b) mouse macrophage cell surface protein, Mac-2 (15); (c) a non-integrin type laminin-binding protein of mouse macrophage (LBP35) (49); and (d) an IgE-binding protein (eBP) from rat basophilic leukemia cells (36).

Once the amino acid sequence of each of these proteins became available, from direct sequencing at the protein level or deduced from the nucleotide sequence, searches in the sequence data bank showed them to be almost identical or homologous (e.g. rat versus mouse) to CBP35. This in turn led to tests of carbohydrate-binding activity and the results indicated that these proteins also exhibited galactose-specific binding activity. Thus, it was found that LBP35

bound to laminin and eBP bound to IgE via carbohydrate-specific recognition and that each of these interactions can be blocked by galactose/lactose.

A: L-34

L-34 was originally isolated from a mouse fibrosarcoma cell line by galactose-specific affinity chromatography. The lectin has a single polypeptide chain with a molecular weight of 34,000. It can recognize galactose-containing glycoconjugates (14).

A monoclonal antibody directed against L-34 was produced . L-34 was localized at the fibrosarcoma cell surface by the antibody (50). It was found that the monoclonal antibody was capable of inhibiting colony formation of some tumor cells under both anchorage-independent and anchorage-dependent growth conditions, suggesting that L-34 might be involved in regulation of tumor cell growth (50).

A cDNA clone from a mouse fibrosarcoma library was identified by the anti-L-34 antibodies and the DNA sequence of the clone was determined. The clone consisted of 144 5′-flanking , 792 coding, and 124 3′-flanking nucleotides (29). A comparison of CBP35 and L-34 cDNA coding regions shows that the CBP35 and L-34 sequences are identical. However, the 5′-untranslated regions of the CBP35 and L-34 clones were different; this will be discussed in Chapter IV.

B: Mac-2

Mac-2 is a 32 kD murine macrophage cell surface antigen. It was originally identified by Ho and Springer (54). Although found largely

in the cytosol, the Mac-2 antigen also appears in the extracellular medium and at the cell surface. The expression of Mac-2 is heterogeneous among different macrophage populations and in different macrophage cell lines (51).

Several cDNA clones of Mac-2 were identified from a library derived from mouse macrophage mRNA (15). One clone was expressed to produce Mac-2 antigen. It was found that this recombinant protein was able to bind to an asialofetuin-sepharose column, suggesting that Mac-2 has the same galactose-specific binding activity as CBP35. The cDNA sequences of Mac-2 clones were determined. For the coding region, the Mac-2 sequences are identical to CBP35. It is interesting that different cDNA clones have the same coding sequences; however, there are some difference in the 5′-flanking regions (15). In order to prove that the difference is not a cloning artifact, polymerase chain reaction (PCR) was used to synthesize mouse macrophage cDNA. A 20-nucleotide oligomer primer designed to hybridize to the mRNA about 130 nucleotides downstream of the ATG starting codon was used to extend the cDNA towards the 5′end, and the reverse direction was achieved with a 28-nucleotide primer which was complementary to the poly-G tail added to the 5′ end of the mRNA. The cDNAs derived from the experiment confirmed that there were two species of mRNA in mouse macrophages. One PCR-derived clone is highly homologous with the original L-34 clone in the 5′-flanking region. The second cDNA is identical to the first Mac-2 cDNA except for a 27-bp inframe insertion. One cDNA sequence may encode for a putative signal peptide for entry into the endoplasmic recticulm, ultimately transporting Mac-2 to the cell surface while the other mRNA, without the signal peptide, may be intracellular Mac-2.

C: Laminin binding protein (LBP35)

A laminin-affinity column was used to fractionate extracts of mouse macrophages. One protein with molecular weight of 35,000 was bound to the column and eluted only in high salt. This was designated as a laminin binding protein (LBP35). When the macrophage was surface labeled , the 35 kD LBP was the predominant surface protein that bound to the laminin affinity column (49).

Two peptide fragments of LBP35 were subjected to amino acid sequence analysis. The sequences obtained from the two fragments were 100% identical with CBP35 (49): Ile Val Leu Asp Phe Xxx Xxx Gly Asn Asp Val Ala Phe Xxx Asn Pro in LBP35 and Ile Val Leu Asp Phe Arg Arg Gly Asn Asp Val Ala Phe His Asn Pro in CBP35; Ile Gln Val Leu Val Glu Ala Asp Xxx Phe Lys Val in LBP35 and Ile Gln Val Leu Val Glu ALa Asp His Phe Lys Val in CBP35. On this basis, the binding of LBP35 to laminin was tested for carbohydrate specificity. It was found that galactose and lactose inhibited the binding of LBP35 to laminin.

D: IgE binding protein

IgE binding protein (eBP) was isolated from rat basophilic leukemia cells (RBL) on the basis of its binding to an IgE-containing affinity column (52). The protein has a molecular weight of 31,000.
By immunoblotting of subcellular fractions of rat basophilic leukemia cells with anti-eBP, the majority of the antigenic determinant was found within the cell, both in the cytoplasm and in the nucleus (53).

The eBP cDNA structure was determined and the amino acid sequence

of eBP showed 85% identity with CBP35, including the CRD. eBP can specifically bind to a galactose affinity column; conversely, CBP35 also can bind to an IgE affinity column (37). The binding of CBP35 and eBP to IgE can be inhibited by galactose. Antibodies directed against CBP35 and eBP were cross reactive with both proteins. Thus, eBP may be the rat homolog of CBP35.

V STRUCTURE OF 14 kD S-TYPE LECTINS

As stated earlier, the second subfamily of the S-type lectins consists of 14-16 kD soluble lectins (Fig. 2). A number of cDNA clones from different species, encoding all or parts of 14 kD lectins, have been sequenced. These include: chicken skin lectin (28), mouse fibrosarcoma lectin L-14 (55), bovine heart lectin (30), electric eel lectin (31) and human lung lectin (32). Some of the above lectins have been also sequenced by amino acid sequence analysis of peptide fragments (32-33).

A comparison of the sequences of 14-16 kD lectins shows a strong homology with each other (35). There are two regions of the sequences that are highly conserved in all proteins; featured by consensus sequences "His-Phe-Asn-Pro-Arg-Phe" and "Trp-Gly-Xxx-Glu-Xxx-Arg-Glu" (Fig. 2). The second consensus sequence, containing a tryptophan and two glutamic acid residues, corresponds to the proposed galactoside-binding site of the eel lectin (31).

The 14 kD chicken skin lectin genomic clone was isolated and the DNA sequence was determined (38). The lectin gene, spanning 4 kb, consists of four exons and three introns. A TATA box, located at -24 to -31 nucleotides, and a possible CCAAT box, located at -80 to -86

nucleotides 5′ to the transcription starting site, were identified. The polyadenylation signal sequence (AATAAA) was identified in the 3′ flanking region. It is interesting that the first exon of the gene is unusually small, only nine nucleotides in length. There is no consensus sequence for SRE in the 5′ flanking region.

VI GOAL OF THIS THESIS

In this thesis work, one cDNA clone and two genomic DNA clones for CBP35 have been isolated, identified and characterized. The goal is to provide the structural basis for: a) the carbohydrate binding activity of the lectin; b) the nuclear localization property; c) the serum stimulated regulation of expression; and d) the dual localization of CBP35 intracellularly and external to the cell.

REFERENCES

1. Barondes, S.H. (1981) Annu. Rev. Biochem. **50**, 207-231.

2. Goldstein, I. J., and Hayes, C. E. (1978) Adv. Carb. Chem. Biochem. **35**, 127-340.

3. Tanabe, T., Pricer, W. E., and Ashwell, G. (1978) J. Bio. Chem. **254**, 1038-1043.

4. Kawasaki, T., and Ashwell, G. (1977) J. Biol. Chem. **251**, 1296-1302.

5. Hoflack, B., and Kornfeld, S. (1985) J. Biol. Chem. **260**, 12008-12014.

6. Sahagian, G. G., Distler, J., and Jourdian, G. W. (1981) Proc. Natl. Acad. Sci. U. S. A. **78**, 4289-4293.

7. Beyer, E.C., and Barondes, S. H. (1982) J. Cell Biol. **92**, 27-37.

8. Barondes, S. H., and Haywood-Reid, P. (1981) J. Cell Biol. **91**, 568-572.

9. Beyer, E., and Barondes, S. H. (1982) J. Cell Biol. **92**, 28-33.

10. Oda, Y., and Kasai, K. (1983) Biochem. Biophys. Acta. **761**, 237-245.

11. Childs, R. A., and Feizi, T. (1980) Cell Biol. Int. Rep. **4**, 755-761.

12. Teichberg, V. I., Silman, I., Beitsch, D. D., and Rescheff, G. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 1383-1387.

13. Powell, J. T. (1980) Biochem. J. **187**, 123-129.

14. Raz, A., Meromsky, L., and Lotan, R. (1986) Cancer Res. **46**, 3667-3672.

15. Cherayil, B. J., Weiner, S. J., and Piliai, S. (1989) J. Exp. Med. **170**, 1959-1972.

16. Roff, C. F., and Wang, J. L. (1983) J. Biochem. Chem. **258**, 10657-10663.

17. Barondes, S. H. (1984) Science **233**, 1259-1264.

18. Drickamer, K. (1988) J. Biol. Chem. **263**, 9557-9560.

19. Briles, E. B., Gregory, W., Flettcher, P., and Kornfeld, S. (1979) J. Cell Biol. **81**, 528-537.

20. Drickamer, K. (1987) Kidney Int. **32**, S167-S180.

21. Moutsatsos, I, K., Wade, M., Schindler, M., and Wang, J. L. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 6452-6456.

22. Raz, A., and Lotan, R. (1987) Cancer Metastasis Rev. **6**, 433-452.

23. Crittenden, S. L., Roff, A. F., and Wang, J. L. (1984) Mol. Cell. Biol. **4**, 1252-1259.

24. Agrwal, N., Wang, J. L., and Voss, P. G. (1989) J. Biol. Chem. **264**, 17236-17242.

25. Lau, L. F., and Nathans, D. (1985) EMBO J. **4**, 3145-3151.

26. Mohun, T., Garrett, N., and Trieisman, R. (1987) EMBO J. **6**, 667-673.

27. Laing, J. G., and Wang, J. L. (1988) Biochem. **27**, 5329-5334.

28. Ohyama, Y., Hirabayashi, J., Oda, Y., Ohno, S., Kawasak, H., Suzuki, K., and Kasai, K. (1986) Biophys. Biochem. Res. Commun. **134**, 51-56.

29. Raz, A., Pazerini, G., and Carmi, P. (1989) Cancer Res. **49**, 3489-3493.

30. Southan, C., Aitken, A., Childs, R. A., Abbott, W. M, and FeiZi, T. FEBS Lett. **214**, 301-304.

31. Paroutoud, P., Levi, G., Teichberg, V. I., and Strosberg, A. D. (1987) Proc. Natl. Acad. Sci. U. S. A. **84**, 6345-6348.

32. Gitt, M. A., and Barondes, S. H. (1986) Proc. Natl. Acad. Sci. U. S. A. **83**, 7603-7607.

33. Hirabayashi, J., Kawasaki, H., Suzuki, K., and Kasai, K. (1987) J. Biochem. **101**, 775-787.

34. Abbott, W. M., and Feizi, T. (1989) Biochem. J. **259**, 291-294.

35. Jia. S., and Wang, J. L. (1988) J. Biol. Chem. **263**, 6009-6011.

36. Albrandt, K., Orida, N. K., and Liu, F.-T. (1987) Proc. Natl. Acad. Sci. U. S. A. **84**, 6858-6863.

37. Laing, J., Robertson, M. W., Gritzmacher, C. A., Wang, J. L., and Liu, F.-T. (19890 J. Biol. Chem. **264**, 1907-1910.

38. Ohyama, Y., and Kasai, K. (1988) J. Biochem. **104**, 173-177.

39. Treisman, R. C. (1986) Cell **46**, 567-574.

40. Gustafson, T. A., Miwa, T., Boxer, L. M., and Kedes, L. (1988) Mol. Cell. Biol. **8**, 4110-4119.

41. Dreyfuss, G. (1986) Ann. Rev. Cell Biol. **2**, 459-498.

42 Choi, Y. D., Grabowski, P. J., Sharp, P. A., And Dreyfuss, G.

(1986) Science **231**, 1534-1539.

44. Cruz-Alvarez, M., and Pellicer, A. (1987) J. Biol. Chem. **262**, 1377-1380.

45. Cobianchi, F., Sengupta, D. W., Zmudzka, B. Z., and Wilson, S. H. (1986) J. Biol. Chem. **261**, 3536-3543.

46. Burd, C. G., Swanson, M. S., Gorlach, M., and Dreyfuss, G. (1989) Proc. Natl. Aced. Sci. U. S. A. **86**, 9788-9792.

47. Swanson, S., Nakagawa, T. Y., LeVan, K., and Dreyfuss, G. (1987) Mol. Cell. Biol. **7**, 1731-1739.

48. Biamonti, G., Buvoli, M., Bussi, M. T., Morandi, C., Cobianchi, F., and Riva, S. (1989) J. Mol. Biol. **207**, 491-503.

49. Woo, H.-J., Shaw, L. M., Messies, J. M., and Mercurio, A. M. (1990) J. Biol. Chem. **265**, 7097-7099.

50. Lotan, R., Lotan, D., and Raz, A. (1985) Cancer Res. **45**, 4349-4353.

51. Nibbering, P. H., Leijh, P. C. J., and Furth, R. van. (1987) Cell Immunol. **105**, 374-380.

52. Liu, F.-T., and Orida, N. (1984) J. Biol. Chem. **259**, 10649-10652.

53. Gritzmacher, C. A., Robertson, M. W., and Liu, F.-T. (1988) J. Immunol. **141**, 2801-2806.

54. Ho, M. K., and Springer, T. A. (1982) J. Immunol. **128**, 1221-1230.

55. Raz, A., Carmi, P., and Pazerini, G. (1988) Cancer Res. **48**, 645-649.

CHAPTER II

Carbohydrate binding protein 35: molecular cloning and expression of a recombinant polypeptide with lectin activity in Escherichia coli

(recombinant DNA; cDNA clone; lambda gt vector; fusion protein; immunoblot screening; peptide mapping; Northern blotting)

Shizhe Jia, Robert P. Mee, Gerald Morford, Neera Agrwal, Patricia G. Voss, Ioannis K. Moutsatsos and John L. Wang

Department of Biochemistry
Michigan State University
East Lansing, MI 48824

(This chapter has been published in Gene **60**, 197-204 (1987); It has been permitted for reprint from publisher)

SUMMARY

Affinity purified antibodies directed against carbohydrate binding protein 35 (CBP35), a galactose-specific lectin, were used to screen a lambda gt11 expression library derived from mRNA of 3T3 fibroblasts. This screening yielded several putative clones containing cDNA for CBP35, one of which was characterized in terms of its expression of a fusion protein containing b-galactosidase and CBP35 sequences. Limited proteolysis of lysates containing the fusion protein, followed by SDS-polyacrylamide gel electrophoresis and immunoblotting with anti-CBP35, yielded a peptide mapping pattern comparable to that obtained from parallel treatment of authentic CBP35. Such a limited proteolysis followed by affinity chromatography on a column of Sepharose derivatized with galactose also yielded a 30-kDa polypeptide that exhibited carbohydrate-binding activity. This polypeptide can be immunoblotted with anti-CBP35, but not with antibodies directed against b-galacto-sidase. These results indicate that we have identified a cDNA clone for CBP35 that yields a recombinant polypeptide with lectin activity produced in E. coli. Using this cDNA clone as a probe, Northern blot analysis showed an increased expression of the CBP35 gene when quiescent 3T3 cells were activated by the addition of serum growth factors.

INTRODUCTION

Carbohydrate binding protein 35 (CBP35; 35-kDa) is a lectin that binds specifically to galactoseandgalactose-N-acetylglucosamine containing glycoconjugates (Roff and Wang, 1983; Roff et al., 1983). Using a highly specific antibody directed against CBP35, immunoblotting and immunofluorescence studies have shown that the lectin can be found in the nucleus of a cell (Moutsatsos et al., 1986). When serum-starved, quiescent cultures of 3T3 cells were activated by the addition of serum, there was an elevation of the expression of CBP35 (detected by immunoblotting) and an apparent translocation into the nucleus (assayed by immunofluorescence) prior to the onset of DNA synthesis in these synchronized cells (Moutsatsos et al., 1987). Because of these interesting distributional properties of CBP35, we have undertaken molecular cloning studies to obtain a reliable and abundant source of material amenable for our purification scheme, to develop a probe for studying the expression of the CBP35 gene, and ultimately, to obtain information on the primary structure of the protein. In this communication, we report the identification of a cDNA clone for CBP35 from a lambda gt11 expression library. This clone expresses a fusion protein which, upon proteolysis and affinity fractionation, yields a polypeptide that exhibits carbohydrate-binding activity.

MATERIALS AND METHODS

(a) Affinity purification of antibodies against CBP35

CBP35 was purified from mouse lung by affinity chromatography on Sepharose columns derivatized with N-e-aminocaproyl-D-galactosamine (Allen and Neuberger, 1975). The generation and characterization of the speci-ficity of rabbit antiserum directed against CBP35 have been reported previously (Roff and Wang, 1983; Crittenden et al., 1984; Moutsatsos et al., 1986). Antibodies directed against CBP35 were affinity purified by specific adsorption to CBP35 on nitrocellulose membranes following the Smith and Fisher procedure (1984). CBP35 (20 ug) was subjected to SDS-PAGE (Laemmli, 1970). After electrophoresis, the protein was transferred to a sheet of nitrocellulose membrane (200 mA, 3 h, 25°C) (Towbin et al., 1979). Strips of nitrocellulose membrane at the left and right edges of the sheet were excised and immunoblotted with rabbit antiserum directed against CBP35 (1:250 dilution, 3 h, 25°C), followed by HRP-conjugated goat anti-rabbit immunoglobulin (Bio-Rad Laboratories, 1:2000 dilution, 1 h, 25°C). The immunoreactive material was revealed via HRP activity with 4-chloro-1-naphthol and hydrogen peroxide as substrates. Using the positions of migration of CBP35 on these edge strips as guides, a horizontal strip of nitrocellulose membrane, corresponding to the CBP35 band and designated as the CBP35 strip, was excised from the preparative nitrocellulose sheet. This CBP35 strip served as the affinity adsorbent for the purification of anti-CBP35 antibodies. Rabbit antiserum directed against CBP35 was diluted (1:150) in 20 ml Tris buffered saline (20 mM Tris, 0.5 M NaCl, pH 7.5) containing 0.5% Tween 20 and was

incubated with the CBP35 strip overnight at 4°C. The supernatant was saved and reused (see below). The CBP35 strip was washed for 5 min in Tris buffered saline. Antibodies bound to the CBP35 on this strip were eluted by the addition of 500 ul of 5 mM glycine, 0.5 M NaCl, 100 ug/ml bovine serum albumin, 0.5% Tween 20, pH 2.3. This solution was transferred to a tube containing 25 ul of 1 M Na_2HPO_4 to neutralize the pH. The elution was repeated with an additional aliquot of 500 ul of elution solu-tion; this second aliquot was neutralized and combined with the first elution. This sample was designated as affinity purified anti-CBP35. The CBP35 strip can also be reused (see below).The supernatant solution containing the diluted antiserum (1:150) was "boosted" by the addition of 50 ul of fresh antiserum. This was then reincubated with the CBP35 strip overnight at 4°C and the whole procedure repeated. We have found that the CBP35 strip can be used for 10-12 adsorption-elution cycles. This conclusion is based on comparing the titers of the affinity purified anti-CBP35 from the various cycles by immunoblotting individual nitrocellulose strips containing a known amount of CBP35.

(b) Screening of the lambda gt11 expression library

The lambda gt11 cDNA expression library (Clone Tech Laboratories) was prepared using mRNA from mouse 3T3 fibroblasts; it contains about 7.3×10^5 independent clones, with an average insert size of 0.94 kb. E. coli (Y1090) were infected with the lambda phage and grown on agar plates (2×10^5 plaques-/plate) (Young and Davis, 1983). Six positive plaque colonies were identifed and rescreened by immunoblotting with affinity-purified anti-CBP35. The clones were isolated and amplified twice. The recombinant lambda gt11 DNA

was isolated from the positive clones by the method described in Maniatis et al., 1982.

(c) Isolation of recombinant polypeptide with lectin activity from fusion protein of clone 1

The procedures for detecting expression of a fusion protein followed those described previously (Huynh et al., 1984). The fusion protein was detected by immunoblotting using either: (a) affinity-purified rabbit anti-CBP35 plus HRP-conjugated goat anti-rabbit immunoglobulin, or (b) mouse anti-E. coli b-galase (Accurate Chemical Scientific Corp) plus HRP-conjugated goat anti-mouse immunoglobulin. Lysates containing fusion protein were digested with protease S. aureus V-8 (Miles Scientific) (Cleveland et al., 1977). In parallel, authentic CBP35 (Crittenden et al., 1984) was digested. The samples were then loaded on SDS-PAGE. The peptide maps were revealed by immunoblotting with affinity-purified anti-CBP35.

E. coli lysogen cells were suspended in TE buffer (125 mM Tris-HCl, 1 mM EDTA, pH 6.8) containing 2 mM b-mercaptoethanol and were lysed by freezing and thawing. The lysate was treated with deoxyribonuclease I (Sigma; 0.2 mg/ml) and was then centrifuged at 13,000 x g for 2 minutes at 4°C. The pellet fraction was incubated at 37°C for 30 minutes with: (a) TE buffer; (b) TE containing 0.5 M lactose; (c) TE containing 1% Triton X-100; (d) TE containing ribonuclease A (Sigma; 10 ug/ml); (e) TE containing 5 M urea; (f) TE containing 8 M urea; and (g) TE containing 0.1% SDS. The mixtures were then centrifuged at 14,000 x g for 2 minutes and the supernatant and pellet fractions were analyzed by SDS-PAGE and immuno-

blotting with anti-CBP35.

To test material derived from the fusion protein for carbohydrate-binding activity, the lysates of E. coli lysogen were extracted for 30 minutes at 37°C with TE buffer containing 0.1% SDS and 2 mM b-mercaptoethanol. After centrifugation at 13,000 x g for 10 minutes, the pellets were reextracted. Protease V-8 was added to the combined supernatants (total protein concentration 20 mg/ml) to a concentration of 0.1 mg/ml. The mixture was incubated at 37°C for 30 minutes, after which the solution was made 1 mM in PMSF and 1% in Triton X-100. This material was dialyzed at 4°C for overnight against buffer A (75 mM Tris-HCl, 50 mM CaCl2, 2 mM b-mercapto-ethanol, 1 mM PMSF, 1 mM NaN_3, pH 7.2) containing 1% Triton X-100. After dialysis, the mixture was incubated overnight at 4°C with 2 ml of Sepharose derivatized with N-e-aminocaproyl-D-galactosamine (Allen and Neuberger, 1975). The beads were then packed into a column (1.4 x 2.8 cm), washed with 15 ml buffer A containing 1% Triton X-100, and 30 ml of buffer A without Triton X-100 (until the absorbance (280 nm) was less than 0.003). The column was eluted with 15 ml of buffer A containing 0.2 M lactose and the effluent material was analyzed by SDS-PAGE and immunoblotting.

(d) Northern blotting analysis

The cDNA insert of clone 1 was derivatized for detection using the BluGENE system (Bethesda Research Laboratories) by nick translation with biotin-7-dATP. Colorimetric detection of the hybridized probe was accomplished by the addition of strep-tavidin-alkaline phosphatase conjugate and incubation in the substrates nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl-phosphate (Amorese,

1986). Serum-starved, quiescent 3T3 cells were synchronously reactivated by the addition of serum (Moutsatsos et al., 1987); the poly A+ fraction of mRNA derived from quiescent 3T3 cells and cells at various times post serum addition were isolated, electrophoresed, transferred onto nitrocellulose, and probed with the clone 1 cDNA following previously described procedures (Stuart et al., 1985). A cDNA clone for mouse b-2-microglobulin (Stuart et al., 1985) was similarly derivatized for detection by the BluGENE system and this probe was used as an internal control of the amount of RNA electrophoresed in each lane of the gel.

RESULTS AND DISCUSSION

(a) Identification of a cDNA clone for CBP35

Using affinity purified anti-CBP35, we have screened a lambda gt11 expression library of 3T3 cell cDNA. Six positive clones were identified and rescreened. The clones were then isolated and amplified twice. The purity of one of the clones, designated as clone 1, was confirmed upon secondary rescreening, in which all plaques yielded positive reactions (Fig. 1a). The use of normal rabbit immunoglobulin failed to yield any positive reaction. In addition, parallel screening of E. coli infected with gt11 phage (no recombinants) did not show positive reaction with anti-CBP35. These results suggest that lambda gt11 clone 1 contained cDNA insert that coded for sequences specifically immunoreactive with affinity purified anti-CBP35.

DNA from clone 1 was digested with the restriction endonuclease EcoR1 to excise the cDNA insert. The digest was analyzed by agarose gel electrophoresis. Based on this analysis, the size of the cDNA insert was estimated to be approximately 900 base pairs (Fig. 1b).

(b) Expression of a fusion protein encoded by lambda gt11 clone 1 DNA

The 3T3 cell cDNA is inserted near the 3′-end of the b-galase gene in the lambda gt 11 cloning vector (Young and Davis, 1983). The

Figure 1: (a) Specific binding of affinity purified rabbit anti-CBP35 to proteins of lambda gt11 clone 1. This clone, positive upon initial screening of a gt11 expression library of mouse 3T3 fibroblast cDNA, was purified and amplified twice. E. coli (Y1090) were infected with the lambda gt11 clone 1 phage and grown on agar; a nitrocellulose filter, saturated with IPTG, was overlaid on the agar plate for 30 min at 42 °C and overnight at 37° C. It was then immunoblotted with affinity purified anti-CBP35 (1:50 dilution, 3 h 25° C) followed by HRP-conjugated goat anti-rabbit immunoglobulin.

(b) DNA from gt11 clone 1 was digested with EcoR1 (10 U/ug DNA, 37° C, 1 hour) to excise the cDNA insert. The digest was analyzed on 1% agarose gels and was visualized by ethidium bromide staining. Lane 1: DNA digest from gt11 clone 1; Lane 2: DNA (HindIII digest) and phiX174 DNA (HaeIII digest) DNA size markers. The position of migration of the cDNA insert is indicated by an arrowhead and is estimated to be 940 base pairs.

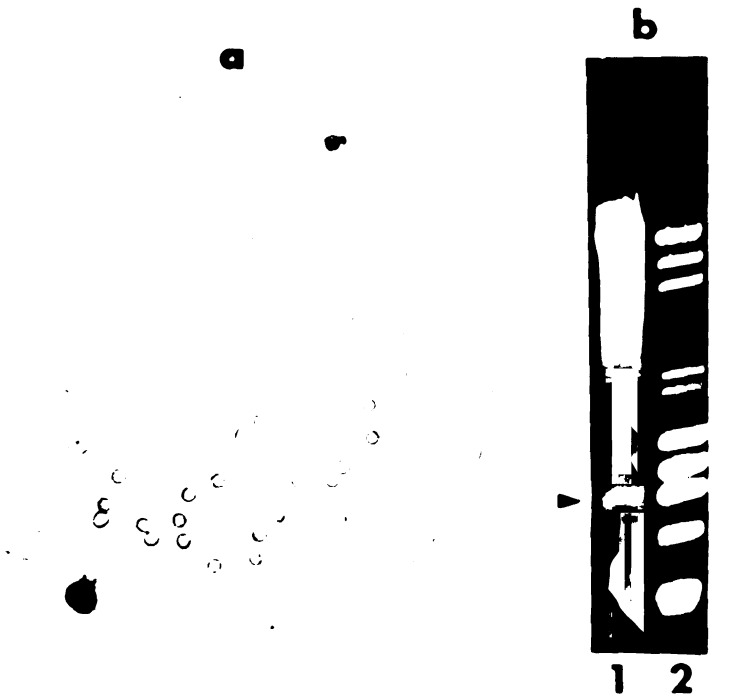


Figure 1

fusion protein encoded by the recombinant lambda gt11 DNA would therefore be composed of b-galase interrupted by the 3T3 cell protein 18 amino acids from the carboxyl terminus. The expression of this fusion protein was examined by infecting E. coli (Y1089) with lambda gt11 clone 1 phage. Lysogens containing lambda gt11 clone 1 DNA were selected on the basis of their growth at 32°C and non-growth at 42°C. After induction with IPTG, the cells were harvested, lysed, and the lysate was analyzed by SDS-PAGE. The lysate of lysogen 1, immunoblotted with anti-b-galase, yielded two major two bands (153 and 150 kD) (Fig. 2a, lane 2). For comparison, E. coli b-galase was also subjected to SDS-PAGE and immunoblotting with anti-b-galase antibodies. This yielded a prominent band at a position corresponding to 116 kD, the subunit molecular weight of b-galase (Fig. 2a, lane 1). These results suggest that in the lysogen 1 lysate, the bands at 153 kDa and 150 kDa represent fusion protein components. This conclusion is supported by immunoblotting of the same bands in lysogen 1 lysate with anti-CBP35 (Fig. 2b, lane 4).

Control experiments were carried out to ascertain the specificity of the immunoblotting reactions. First, anti-CBP35 immunoblots only CBP35 (Fig. 2b, lane 3) but not b-galase (Fig. 2b, lane 1). Second, when lysates of lysogen 1 were prepared in the absence of IPTG, anti-CBP35 failed to detect any polypeptide bands in the immunoblots (Fig. 2b, lane 2). All of these results strongly suggest that the 153 kD and 150 kD bands (Fig. 2a, lane 2 and Fig. 2b, lane 4) represent fusion protein components containing both b-galase and CBP35 sequences recognized by the respective antibodies.

(c) Peptide maps derived from fusion proteins and CBP35

Figure 2: Detection of expression of fusion protein by immunoblotting analysis. E. coli (Y1089) was infected with lambda gt11 clone 1 phage. Lysogens, selected on the basis of their growth at 32°C and nongrowth at 42°C, were incubated at 32°C to an absorbance value (595 nm) of 0.5. The temperature was shifted to 45°C for 20 minutes, then IPTG was added to 10 mM. After further incubation at 38° C for 1 hour, the cells were harvested in TE buffer containing 0.1% SDS, lysed by freezing and thawing, and subjected to SDS-PAGE. (a) Immunoblot analysis with mouse anti-E. coli b-galase. Lane 1: (1) E. coli b-galase; Lane 2: lysate of lysogen 1 derived from lambda gt11 clone 1 induced with IPTG. (b) Immunoblot analysis with affinity purified rabbit anti-CBP35. Lane 1: E. coli b-galase; Lane 2: lysate of lysogen 1 derived from lambda gt11 clone 1 without IPTG induction; Lane 3: authentic CBP35; Lane 4: lysate of lysogen 1 derived from lambda gt11 clone 1 induced with IPTG.

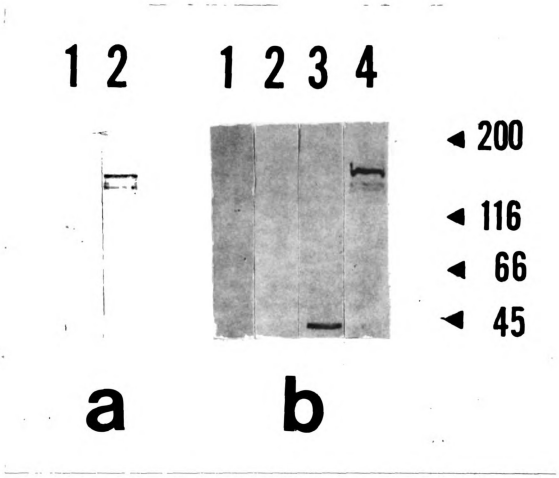


Figure 2

Lysates of lysogen 1 were digested with V-8 protease at 37°C for 30 min and then subjected to SDS-PAGE and immunoblotting with anti-CBP35. This yielded two predominant bands: 33 kD and 30 kD (Fig. 3, lane 2). These bands correspond closely to products of V-8 digestion of authentic CBP35 subjected to SDS-PAGE and immunoblotting in parallel (Fig. 3, lane 3). Neither of these two bands could be immunoblotted with antibodies directed against b-galase. Moreover, analysis of the V-8 digestion products of authentic b-galase by SDS-PAGE and immunoblotting with anti-b-galase did not yield the peptide fragments observed in Figure 3. These results suggest that the cDNA insert of the lambda gt11 clone 1 contains the CBP35 coding sequence.

This conclusion is supported by results of our preliminary sequence analysis of clone 1. Three portions of the nucleotide sequence yielded the following deduced amino acid sequences: (i)Gly-Asn-Asp-Val-Ala-Phe-His-Phe-Asn-Pro-Arg; (ii) Gln-Asp-Asn-Asn-Trp-Gly-Lys; and (iii) Gln-Ser-Ala-Phe-Pro- Phe-Glu-Ser-Gly-Lys, corresponding to three tryptic peptides. These sequences show extensive homology to those of other b-galactoside binding lectins, including bovine heart lectin (Southan et al., 1987) and chicken skin lectin (Hirabayashi et al., 1987). These results confirm that cDNA clone 1 codes for CBP35.

(d) Isolation of recombinant CBP35 from fusion protein

Several experiments were performed to test the possibility that V-8 digestion of the fusion protein may generate a recombinant 3T3 cell protein that retains the carbohydrate-binding capacity of CBP35. To test for carbohydrate-binding activity, it was important to

Figure 3: Comparison of the peptide maps derived from the fusion protein of clone 1 and CBP35. Lysates containing fusion protein (protein concentration 20 mg/ml) were digested with V-8 protease (0.1 mg/ml, 37°C, 30 minutes). Purified CBP35 was digested in parallel. The samples were subjected to SDS-PAGE and immunoblotting analysis with affinity purified anti-CBP35. Lane 1: lysate of lysogen 1 derived from lambda gt11 clone 1; Lane 2: lysate of lysogen 1 derived from lambda gt11 clone 1 after digestion with V-8 protease; Lane 3: CBP35 digested with V-8 protease; and Lane 4: authentic CBP35.

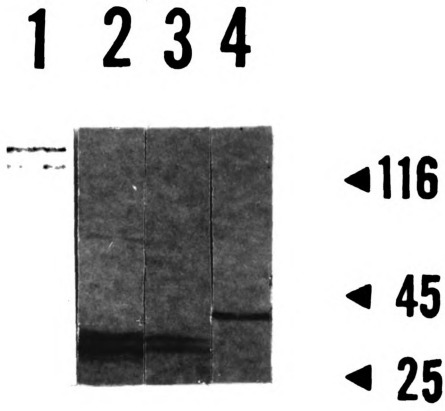


Figure 3

retain the protein in native form or to be able to renature it from buffers containing denaturants. Although the fusion protein was soluble in lysates of lysogen1 prepared in SDS-PAGE buffer, we wished to ascertain its solubility in aqueous buffers without denaturants. Lysates of lysogen 1 were prepared in various buffers and the extraction mixtures were centrifuged and separated into the supernatant and pellet fractions. These fractions were then analyzed for fusion protein by SDS-PAGE and immunoblotting with anti-CBP35. We found that the fusion potein could be solubilized (detected in the supernatant fraction) in TE buffers containing 0.1% SDS or 8 M urea. In contrast, TE buffers containing no denaturants, or TE buffers containing non-ionic detergents (1% Triton X-100) or even low concentrations of urea (5 M) failed to solubilize the fusion protein. Finally, TE buffers containing 10 ug/ml ribonuclease also did not solubilize the fusion protein.

Lysates of lysogen 1 were prepared in TE buffer containing 0.1% SDS and 2 mM b-mercaptoethanol. The supernatant fraction was digested with V-8 protease (0.5 h at 37°C), dialyzed versus Tris buffer containing 1% Triton X-100 (but without SDS), and then fractionated over an affinity column of Sepharose derivatized with N-e-aminocaproyl-D-galactosamine. The bound material was eluted with 0.2 M lactose and analyzed by SDS-PAGE and immunoblotting with anti-CBP35. V-8 digestion of the fusion protein yielded two polypeptides (33 and 30 kDa) that reacted with anti-CBP35 (Fig. 4, lane 2). After purification over the affinity column, however, only the 30 kDa polypeptide was detected by immunoblotting with anti-CBP35 (Fig. 4, lane 3). The position of migration of this polypeptide closely matched that of one of the fragments derived from CBP35 after V-8 digestion (see Fig. 4, lane 3). When antibodies directed against

Figure 4: Immunoblot blot analysis of recombinant polypeptide with lectin activity using affinity purified rabbit anti-CBP35. Lane 1: authentic CBP35; Lane 2: lysate of lysogen 1 derived from lambda gt11 clone 1 after digestion with V-8 protease (0.1 mg/ml, 30 min, 37°C); Lane 3: material that was eluted with 0.2 M lactose when lysate of lysogen 1 derived from lambda gt11 clone 1 was digested with V-8 protease and then purified by affinity chromatography on a column (1.4 x 2.8 cm) of Sepharose derivatized with N-e-aminocaproyl-D-galactosamine.

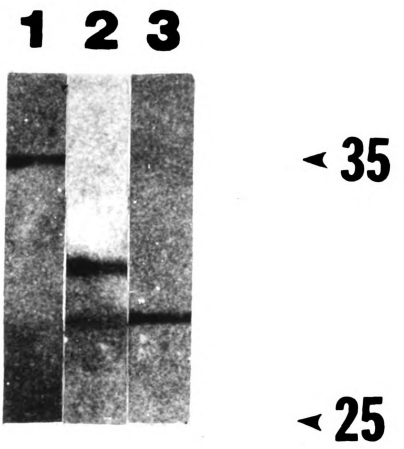


Figure 4

b-galase were used to immunoblot this material, there was no reaction. Thus, it appears that the 30 kDa polypeptide contained little or no b-galase sequences. These results strongly suggest that the 3T3 cell polypeptide contained within the fusion protein carried carbohydrate-binding activity.

The fact that the 3T3 cell portion of the fusion protein, after V-8 digestion, can be isolated by galactose-Sepharose affinity chromatography suggests that we have quite fortuitously come upon a way to purify the eukaryotic polypeptide synthesized in E. coli. Consistent with many reports on the production of eukaryotic polypeptides in E. coli (Marston, 1986), our recombinant polypeptide accumulates intracellularly, but in an aggregated and insoluble form. This fact determines, in part, the strategy for purification. On the one hand, the aggregation of recombinant protein in a discrete form, as dense inclusion bodies (Prouty et al., 1975), allows us to use low speed centrifugation (13,000 x g) to separate it from most of the bacterial cell debris. On the other hand, the recombinant fusion protein has to be solubilized, from the aggregate by the use of strong denaturants such as SDS or urea. Once the fusion protein was solubilized, we then had to liberate the recombinant 3T3 cell polypeptide from the fusion protein. In previous reports, various molecular engineering strategies were devised to place a cleavage site between the carboxyl terminus of the prokaryotic sequence and the amino terminus of the eukaryotic coding sequence (Marston, 1986). In the case of the lambda gt11 cloning vector, the juncture of the b-galactosidase sequence and the cDNA insert is through an EcoRI cleavage site, which contains a glutamic acid residue. V-8 protease cleaves on the carboxyl terminal side of glutamic acid residues in polypeptides (Houmard and Drapeau, 1972). In our fusion

protein, it appears that this site was indeed susceptible to cleavage by the enzyme since antibodies against b-galase failed to immunoblot the carbohydrate-binding 30 kD polypeptide. As a result, lambda gt11 clone 1 appears to be a good source of CBP35 with lectin activity amenable for our purification scheme. The question is then raised whether V-8 digestion and subsequent purification can be applied to other eukaryotic proteins cloned in lambda gt11 expression vectors.

(e) Northern blot analysis of mRNA encoding CBP35

In previous studies, we had shown that when quiescent, serum-starved 3T3 cells were stimulated by the addition of serum, there was an increase in the level of CBP35 as quantitated by immunoblotting analysis (Moutsatsos et al., 1987). This increase occurred within 8 hours after serum addition, well before the synchronized cells entered the first period of DNA synthesis (between 16-24 hours after serum addition). The availability of a cDNA probe for CBP35 provided the opportunity to quantitate the expression of the CBP35 gene at the level of mRNA.

Northern blot analysis was carried out on the poly A+ fraction of mRNA derived from quiescent 3T3 cells and from cells at 8, 16, and 24 hours after serum stimulation. The cDNA probe for CBP35 revealed a mRNA species of about 1.3 kb that is absent or present at very low levels in quiescent cells and increases as a function of time after serum addition (Fig. 5A). The level of mRNA for CBP35 increased within 8 hours after stimulation. The same blot was subjected to a probe for b-2-microglobulin (Fig. 5B), whose mRNA level does not vary substantially during the cell cycle (Kelley et al., 1983).

Figure 5: Northern blot analysis of the poly A+ fraction of mRNA derived from: (1) serum-starved, quiescent 3T3 fibroblasts; (2) 8 hours, (3) 16 hours, and (4) 24 hours after the addition of serum (10%). (A) The blot was probed with the cDNA of clone 1. The numbers on the right indicate the positions of migration of size markers (in kb). (B) The same blot was subjected to a second probe, b-2-microglobulin, as an internal control of the amount of RNA electrophoresed in each lane.

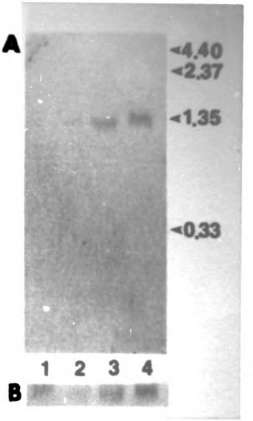


Figure 5

Lane 2 in Fig. 5B showed a less intense b-2-microglobulin signal, suggesting that a less amount of total RNA was loaded, probably due to experimental errors. These results corroborate at the mRNA level previous immunoblotting and immunofluorescence observations on the proliferation dependent expression of CBP35 (Moutsatsos et al., 1987).

The elevated level of the CBP35 mRNA after serum stimulation was observed even in the presence of cycloheximide. This suggests that the stimulation of expression of the CBP35 gene is a "primary" event of growth factor addition, without the requirement for protein synthesis. Moreover, the data suggest that the CBP35 mRNA was "super-induced" (higher level of mRNA in the presence of cycloheximide than in its absence) in a fashion similar to that observed for the oncogenes c-myc and c-fos (Lau and Nathans, 1987). The possible explanation is that the mRNA for CBP35 is stabilized by the presence of cycloheximide.

Finally, the fact that the size of the cDNA insert of clone 1 (1 kb) approximates the size of the mRNA detected on Northern blots (1.3 kb) suggests that the cDNA clone we have identified is nearly full length. This is consistent with the conclusion that the cDNA clone codes for sufficient structural information for carbohydrate binding activity.

ACKNOWLEDGMENTS

We thank Drs. Sue Conrad and Moriko Ito for their advice throughout these studies and Mrs. Linda Lang for her help in the preparation of the manuscript. This work was supported by grants GM-27203 and GM-32310 from the National Institutes of Health. J.L. Wang was supported by Faculty Research Award FRA-221 from the American Cancer Society.

REFERENCES

Allen, A.K. and Neuberger, A.: A simple method for the preparation of an affinity absorbent for soybean agglutinin using galactosamine and CH-Sepharose. FEBS (Fed. Eur. Biochem. Soc.) Lett. 50 (1975) 362-364.

Amorese, D. :Nonradioactive detection of RNA using the BluGENE system. Focus 8 (1986) 6-8.

Cleveland, D., Fischer, S., Kirschner, M. and Laemmli, U. :Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J. Biol. Chem. 252 (1977) 1102-1106.

Crittenden, S.L., Roff, C.F. and Wang, J.L. :Carbohydrate-Binding Protein 35: Identification of the galactose-specific lectin in various tissue of mice. Mol. Cell. Biol. 4 (1984) 1252-1259.

Hirabayashi, J., Kawasaki, H., Suzuki, K. and Kawai, K. :Complete amino acid sequence of 14 kDa b-galactoside-binding lectin of chick embryo. J. Biochem. (Tokyo) 101 (1987) 775-787.

Houmard, J. and Drapeau, G. :Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 3506-3509.

Huynh, T., Young, R. and Davis, R. :Constructing and screening cDNA libraries in lambda gt10 and gt11. In Glover, D. (Ed.), DNA cloning techniques: a practical approach. IRL Press, Oxford, 1984, pp. 49-78.

Kelley, K., Cochran, B.H., Stiles, C.D. and Leder, P.: Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet-derived growth factor. Cell 35 (1983) 603-610.

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

Lau, K.F. and Nathans, D. :Expression of a set of growth related immediate early genes in Balb/c 3T3 cells: coordinate regulation with c-fos or c-myc. Proc. Natl. Acad. Sci. USA 84 (1987) 1182-1186.

Maniatis, T., Fristch, E.F. and Sambrook, J. :Molecular Cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1982.

Marston, F. :The purification of eukaryotic polypeptides synthesized in Escherichia coli. Biochem. J. 240 (1986) 1-12.

Moutsatsos, I.K., Davis, J.M. and Wang, J.L. :Endogenous lectins from cultured cells: subcellular localization of Carbohydrate-binding protein 35 in 3T3 fibroblasts. J. Cell Biol. 102 (1986) 477-483.

Moutsatsos, I.K., Wade M., Schindler, M. and Wang, J.L. :Endogenous lectins from cultured cells: nuclear localization of Carbohy-

drate-binding protein 35 in proliferating 3T3 fibroblasts. Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 6452-6456.

Prouty, W.F., Karnovsky, M.J. and Goldberg, H.L. :Degradation of abnormal proteins in Escherichia coli. Formation of protein inclusions in cells exposed to amino acid analogs. J. Biol. Chem. 250 (1975) 1112-1122.

Roff, C.F., Rosevear, P.R., Wang, J.L. and Barker, R. : Identification of Carbohydrate-binding proteins from mouse and human fibroblasts. Biochem. J. 211 (1983) 625-629.

Roff, C.F. and Wang, J.L. :Endogenous lectins from cultured cells. Isolation and characterization of Carbohydrate-binding proteins from 3T3 fibroblasts. J. Biol. Chem. 258 (1983) 10657-10663.

Smith, D.E. and Fisher, P.A. :Identification, developmental regulation, and response to heat shock of two antigenically related forms of a major nuclear envelope protein in Drosophila embryos:

application of an improved method for affinity purification of antibodies using polypeptides immobilized on nitrocellulose blots. J. Cell Biol. 99 (1984) 20-28.

Southan, C., Aitken, A., Childs, R.A., Abbott, W.M. and Frizi, T. : Amino acid sequence of ┤-galactoside-binding bovine heart lectin. member of a novel class of vertebrate proteins. FEBS (Fed. Eur. Biochem. Soc.) Lett. 214 (1987) 301-304.

Stuart, P., Ito, M., Stewart, C. and Conrad, S.E. :Induction of cellular thymidine kinase occurs at the mRNA level. Mol. Cell Biol. 5 (1985) 1490-1497.

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 (1979) 4350-4354.

Young, R. and Davis, R. :Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. U.S.A. 80 (1983) 1194-1198.

CHAPTER III

CARBOHYDRATE BINDING PROTEIN 35: COMPLEMENTARY DNA SEQUENCE REVEALS HOMOLOGY WITH PROTEINS OF THE hnRNP*

Shizhe Jia and John L. Wang

Department of Biochemistry
Michigan State University
East Lansing, MI 48824

Running Title: Carbohydrate Binding Protein 35

(This chapter has been published, J. Biol. Chem. **263**, 6009-6011 (1988); it has been permitted for reprint from publisher)

FOOTNOTES

* This work was supported by grant GM-27203 from the National Institutes of Health. J.L.W. was supported by Faculty Research Award FRA-221 from the American Cancer Society.

1 The abbreviations used are: CBP, carbohydrate binding protein; hnRNP, heterogeneous nuclear ribonucleoprotein complex.

SUMMARY

The complete nucleotide sequence of a cDNA clone for Carbohydrate Binding Protein 35, a galactose specific lectin identified in the nucleus of mouse 3T3 fibroblasts, has been determined. The deduced amino acid sequence suggests that the protein consists of two domains: (a) an amino terminal portion that is homologous to certain regions of proteins of the heterogeneous nuclear ribonucleoprotein complex (hnRNP); and (b) a carboxyl terminal portion that is homologous to b-D-galactoside specific lectins isolated from a number of animal tissues. This two domain motif is reminiscent of several DNA- and RNA-binding proteins.

INTRODUCTION

Carbohydrate Binding Protein 35 (CBP35; Mr 35,000) is a galactose specific lectin identified in the nuclei of cultured mouse 3T3 fibroblasts (1-3). The level of expression of this lectin and its nuclear localization are dependent on the proliferative state of the cells (4). When quiescent 3T3 cells were stimulated by the addition of serum, immunofluorescence studies with a polyclonal antiserum directed against CBP35 revealed prominent punctate intranuclear staining, suggesting the association of the lectin with subnuclear structures. Monospecific, affinity purified rabbit anti-CBP35 was used to screen a gt11 cDNA library derived from 3T3 fibroblasts. One identified clone, designated as clone 1, expressed a fusion protein which, upon digestion with S. aureus V-8 protease, yielded a polypeptide (Mr 30,000) that had carbohydrate-binding activity (5). We now report the complete nucleotide sequence of this cDNA clone and its deduced amino acid sequence.

MATERIALS AND METHODS

A: DNA cloning

CBP35 cDNA clone was identified as described before (5). cDNA insert was isolated and cloned into M13mp18 in both orientations (17).

B: DNA sequencing

The nucleotide sequences of the insert in M13 clones were determined using the dideoxy chain termination method (18). The universal primer was purchased from New England Biolabs. Six synthetic oligonucleotide primers were synthesized on an Applied Biosystems DNA synthesizer (Model 380 B).

RESULTS AND DISCUSSION

A restriction map and the sequencing strategy for clone 1 is shown in Fig. 1, and the nucleotide sequence of the 883 residue cDNA insert of clone 1 is shown in Fig. 2. The first amino acid of the polypeptide encoded by this cDNA and its translational reading frame were determined on the basis of: (i) the known nucleotide sequence of the EcoRI cleavage site, which represents the juncture of the b-galactosidase sequence and the cDNA insert in the lambda gt11 cloning vector (5); and (ii) homology between the deduced amino acid sequence and several b-D-galactoside specific lectins (see below). The deduced amino acid sequence accounted for 263 amino acids (Fig. 2). Comparison of the sequence derived from clone 1 and another cDNA clone indicates that the sequence shown in Fig.2 lacks only the ATG initiation codon from the full protein coding region (S. Jia and J.L. Wang, unpublished observations). Therefore, the deduced amino acid sequence is missing only the amino terminal methionine of the CBP35 polypeptide. Previous Northern blotting analysis had yielded 1.3 kilobases as the transcript size corresponding to the clone 1 cDNA insert (5). Thus, although clone 1 contains essen-tially all the protein structural information, it nevertheless represents a less than full length cDNA clone in terms of the 5′- and 3′-untranslated regions.

Analysis of the deduced amino acid sequence of clone 1 in terms of hydrophilicity or hydropathy value (6) clearly shows a two-domain structure. Using the numbering system for the amino acid sequence shown in Fig. 2, the amino terminal portion (residues 1-125) exhibited neither a highly hydrophilic nor a hydrophobic region.

Figure 1: Restriction map of the cDNA insert of CBP35 clone 1 and strategy for sequencing. The cDNA insert was subcloned into M13mp18 (17) in both orientations and nucleotide sequences were determined using the dideoxy chain termination method (18). [^{35}S]dATP (New England Nuclear, 500 Ci/mmole) was used in place of [^{32}P]dATP. The arrows indicate the regions that were sequenced using one universal primer and six synthetic oligonucleotide primers. The oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer (Model 380 B).

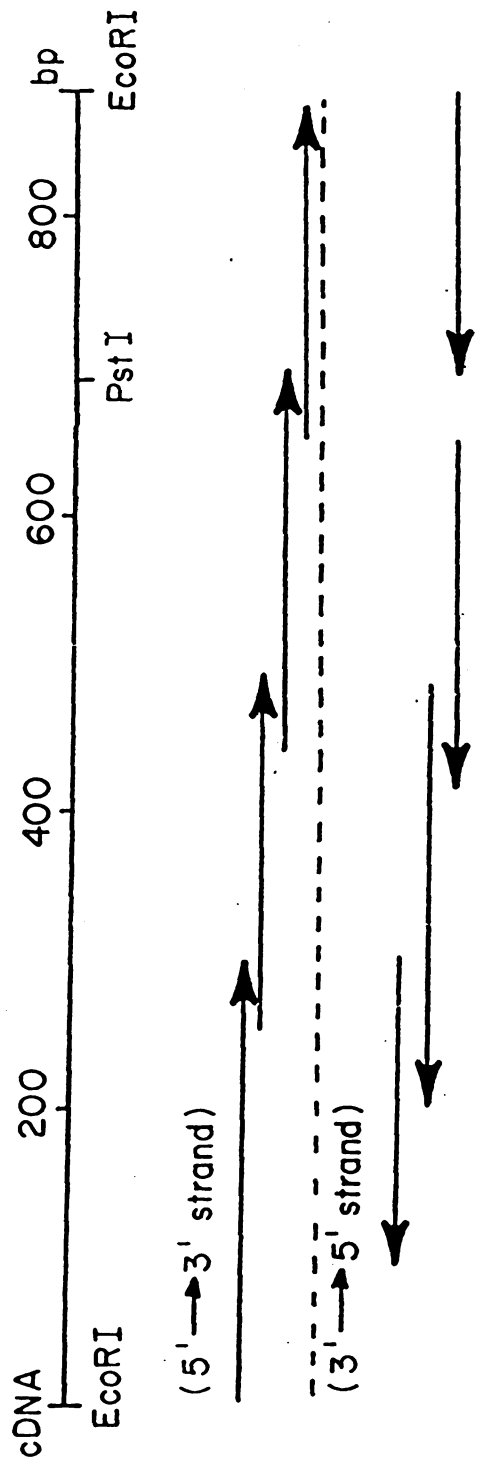


Fig. 1

Figure 2: Nucleotide and deduced amino acid sequence for CBP35 clone 1. The numbers at the right indicate the positions of the nucleotide base pairs (top line) and the amino acids residues (bottom line). The open reading frame in the nucleotide sequence covers 263 amino acid residues. * represents the in-frame stop codon. The amino acid sequence was deduced from the nucleotide sequence using the program of Staden (19).

```
CGA GAC AGC TTT TCG CTT AAC GAT GCC TTA GCT GGC TCT GGA AAC CCA AAC CCT CAA GGA TAT CCG GGT GCA TGG GGG AAC CAG CCT GGG   90
Arg Asp Ser Phe Ser Leu Asn Asp Ala Leu Ala Gly Ser Gly Asn Pro Asn Pro Gln Gly Tyr Pro Gly Ala Trp Gly Asn Gln Pro Gly   30

GCA GGG GGC TAC CCA GGG GCT GCC TAT CCT GGG GCC TAT CCA GGA CAG GCT CCT CCA GGG GCC TAC CCA GGA CAG GCT CCT CCA GGG GCC  180
Ala Gly Gly Tyr Pro Gly Ala Ala Tyr Pro Gly Ala Tyr Pro Gly Gln Ala Pro Pro Gly Ala Tyr Pro Gly Gln Ala Pro Pro Gly Ala   60

TAT CCA GGA CAG GCT CCT CCT AGT GCC TAC CCC GGC CCA ACT GCC CCT GGA GCT TAT CCT GGC CCA ACT GCC CCT GGA GCT TAT CCT GGT  270
Tyr Pro Gly Gln Ala Pro Pro Ser Ala Tyr Pro Gly Pro Thr Ala Pro Gly Ala Tyr Pro Gly Pro Thr Ala Pro Gly Ala Tyr Pro Gly   90

CAA CCT GCC CCT GGA GCC TTC CCA GGG CAA CCT GGG GCA CCT GGG GCC TAC CCC CAG TGC TCT GGA GGC TAT CCT GCT GCT GGC CCT TAT  360
Gln Pro Ala Pro Gly Ala Phe Pro Gly Gln Pro Gly Ala Pro Gly Ala Tyr Pro Gln Cys Ser Gly Gly Tyr Pro Ala Ala Gly Pro Tyr  120

GGT GTC CCC GCT GGA CCA CTG ACG GTG CCC TAT GAC CTG CCC TTG CCT GGA GGA GTC ATG CCC CGC ATG CTG ATC ACA ATC ATG GGC ACA  450
Gly Val Pro Ala Gly Pro Leu Thr Val Pro Tyr Asp Leu Pro Leu Pro Gly Gly Val Met Pro Arg Met Leu Ile Thr Ile Met Gly Thr  150

GTG AAA CCC AAC GCA AAC AGG ATT GTT CTA GAT TTC AGG AGA GGG AAT GAT GTT GCC TTC CAC TTT AAC CCC CGC TTC AAT GAG AAC AAC  540
Val Lys Pro Asn Ala Asn Arg Ile Val Leu Asp Phe Arg Arg Gly Asn Asp Val Ala Phe His Phe Asn Pro Arg Phe Asn Glu Asn Asn  180

AGA AGA GTC ATT GTG TGT AAC ACG AAG CAG GAC AAT AAC TGG GGA AAG GAA GAA AGA CAG TCA GCC TTC CCC TTT GAG AGT GGC AAA CCA  630
Arg Arg Val Ile Val Cys Asn Thr Lys Gln Asp Asn Asn Trp Gly Lys Glu Glu Arg Gln Ser Ala Phe Pro Phe Glu Ser Gly Lys Pro  210

TTC AAA ATA CAA GTC CTG GTT GAA GCT GAC CAC TTC AAG GTT GCG GTC AAC GAT GCT CAC CTA CTG CAG TAC AAC CAT CGG ATG AAG AAC  720
Phe Lys Ile Gln Val Leu Val Glu Ala Asp His Phe Lys Val Ala Val Asn Asp Ala His Leu Leu Gln Tyr Asn His Arg Met Lys Asn  240

CTC CGG GAA ATC AGC CAA CTG GGG ATC AGT CGT GAC ATA ACC CTC ACC AGC GCT AAC CAC GCC ATG ATC TAA GCC AGA AGG GGC GGC ACC  810
Leu Arg Glu Ile Ser Gln Leu Gly Ile Ser Gly Asp Ile Thr Leu Thr Ser Ala Asn His Ala Met Ile *

GAA ACG CCC TGT GTG CCT TAG GAG TGG GAA ACT TTG CAT TTC TCT CTC CTT ATC CTT CTT GTA AGA CAT CCC G                        883
```

Fig. 2

This was quite distinct from the carboxyl terminal portion (residues 126-263), which contained both hydrophobic and hydrophilic regions that are characteristic of many globular proteins. Each of these two domains is homologous to distinct groups of proteins previously identified on the basis of their function/activity. A region containing 76 amino acids (residues 138-214) was found to be homologous to a number of b-D-galactoside specific lectins (7-11) (Fig. 3). For bovine heart lectin (10), chicken skin lectin (7,8), and electric eel lectin (11), the complete amino acid sequences over this 76-residue stretch were available. The extents of homology between CBP35 with these lectins were: (a) 34% with lectin from electric eel; (b) 36% with bovine heart lectin; and (c) 38% with chicken skin lectin. Several peptide sequences are highly conserved in all the lectins sequenced so far. These include the sequence His-Phe-Asn-Pro-Arg-Phe-Asn (residue 171-177) and the sequence Trp-Gly-Lys-Glu-Glu-Arg-GlnSer-Ala-Phe-Pro-Phe (residues 194-205). The latter sequence, containing a tryptophan and two glutamic acid residues, correspond to the peptide hypothesized to be present in the b-D-galactoside binding site of the eel lectin (11).

Previous searches of data bases with the bovine heart lectin sequence (10) and with the sequences of two hepatoma clones (9) failed to reveal any significant homologies with other proteins. As a result of these analyses, it was proposed that the soluble vertebrate lectins with b-D-galactoside binding activity represent a new protein family. CBP35 now joins this protein family on the basis of its sequence homology with these lectins. Moreover, the present sequence results also provide the structural basis for the observation that the clone 1 fusion protein, upon V-8 protease digestion, yielded a polypeptide with carbohydrate-binding activity

Figure 3: Comparison of the amino acid sequence of CBP35 clone 1 (residues 127 through 214 in the numbering system of Figure 2) with the amino acid sequences of six other b-D-galactoside binding lectins. BHL, bovine lung lectin (10), CSL, chicken skin lectin (7,8), HLL, peptides from human lung lectin (9), HEP 1 and HEP 2, sequences deduced from two cDNA clones derived from a human hepatoma library (9), and EEL, lectin from electric eel (11). Dashes indicate gap introduced for optimal alignment; complete gaps are positions for which no sequence information is available; * denotes uncertain residue assignments. Sequence identities between CBP35 clone 1 and the other lectins are highlighted by boxes. The sequences were compared using the FASTP program (16).

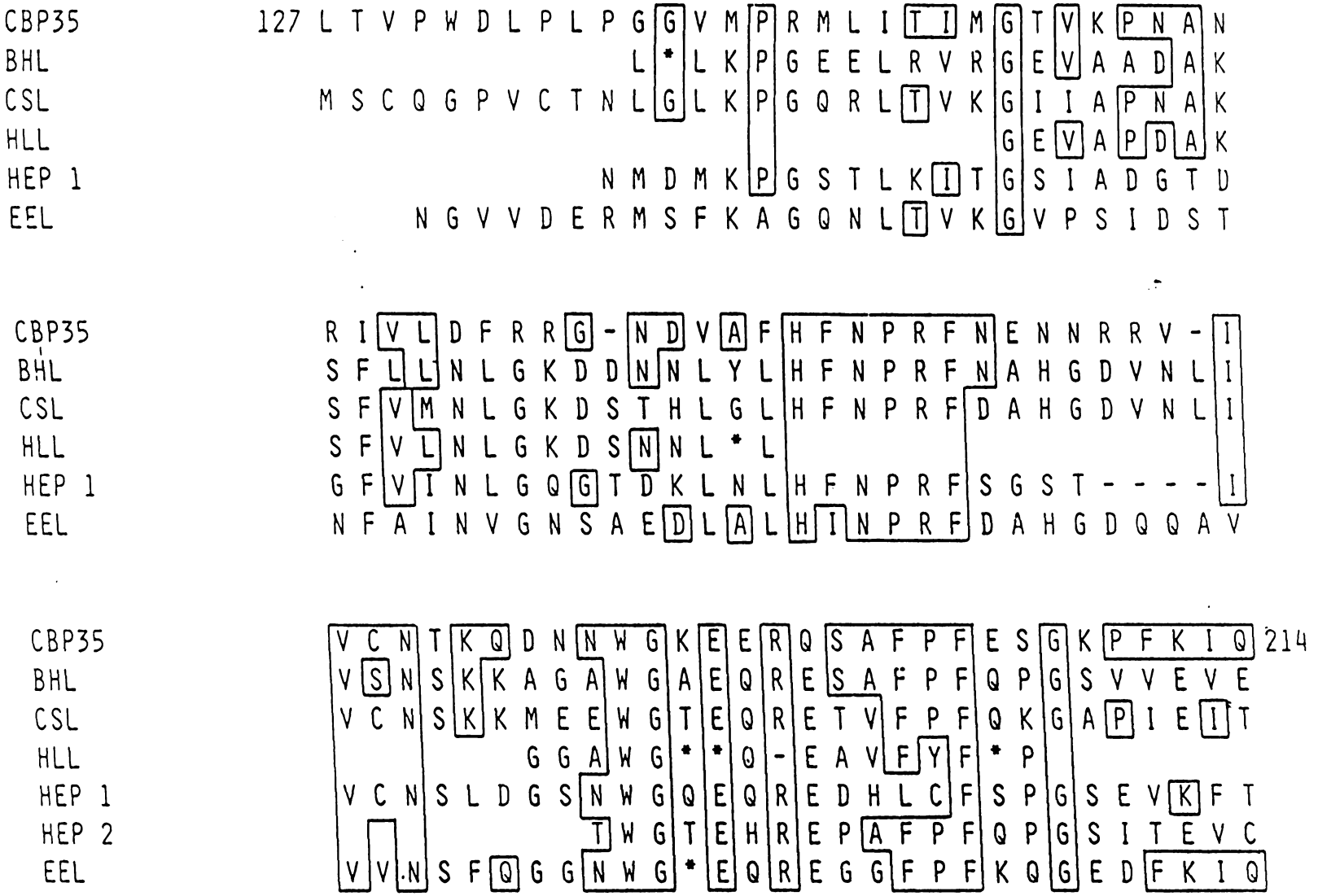

CBP35 127 L T V P W D L P L P G G V M P R M L I T I M G T V K P N A N
BHL L * L K P G E E L R V R G E V A A D A K
CSL M S C Q G P V C T N L G L K P G Q R L T V K G I I A P N A K
HLL G E V A P D A K
HEP 1 N M D M K P G S T L K I T G S I A D G T D
EEL N G V V D E R M S F K A G Q N L T V K G V P S I D S T
CBP35 R I V L D F R R G - N D V A F H F N P R F N E N N R R V - I
BHL S F L L N L G K D D N N L Y L H F N P R F N A H G D V N L I
CSL S F V M N L G K D S T H L G L H F N P R F D A H G D V N L I
HLL S F V L N L G K D S N N L * L
HEP 1 G F V I N L G Q G T D K L N L H F N P R F S G S T - - - - I
EEL N F A I N V G N S A E D L A L H I N P R F D A H G D Q Q A V
CBP35 V C N T K Q D N N W G K E E R Q S A F P F E S G K P F K I Q 214
BHL V S N S K K A G A W G A E Q R E S A F P F Q P G S V V E V E
CSL V C N S K K M E E W G T E Q R E T V F P F Q K G A P I E I T
HLL G G A W G * * Q - E A V F Y F * P
HEP 1 V C N S L D G S N W G Q E Q R E D H L C F S P G S E V K F T
HEP 2 T W G T E H R E P A F P F Q P G S I T E V C
EEL V V N S F Q G G N W G * E Q R E G G F P F K Q G E D F K I Q

Fig. 3

(5).

While the amino acid sequence toward the carboxyl terminal half of CBP35 is homologous to the sequences of other lectins, the sequence at the amino terminal end also showed interesting features. First, the sequence between residues 40-112 showed eight internal sequence homologies (Fig. 4): (I) residues 40-48; (II) residues 49-57; (III) residues 58-66; (IV) residues 67-75; (V) residues 76-84; (VI) residues 85-93; (VII) residues 94-102; and (VIII) residues 104-112. Each of these homologous regions consists of a 9-residue repeat, with a consensus sequence of Pro-Gly-Ala-Tyr-Pro-Gly, followed by three additional amino acids. As a result, this stretch of the sequence is characterized by a high proportion of Pro (27%) and Gly (24%).

In addition, the sequence at the amino terminal portion also showed homology with the amino acid sequences of several proteins identified as polypeptides of the heterogeneous nuclear ribonucleoprotein complex (hnRNP) (Fig. 5). These include: (a) 25 identities over 108 residues with a glycine-rich protein (GRP33) of the hnRNP of brine shrimp Artemia salina (12); (b) 18 identities over 71 residues with the deduced amino acid sequence of clone DL-4, identified from a human hepatoma cDNA library on the basis of it expression of a fusion protein reactive with chicken antibodies directed against bovine hnRNP proteins (13); and (c) 11 identities over 42 residues with human hnRNP protein C1 (14). There was no apparent homology between the sequences of CBP35 and the rat hnRNP protein A1 (15) when subjected to the same analysis with the FASTP program (16).

None of the hnRNP proteins contained any copies of the striking

9-residue repeat sequence (Fig. 4) found in CBP35. However, the extent of homology between CBP35 and the hnRNP proteins shown in Fig.6 was 25%, comparable to the level of homology between the hnRNP proteins themselves (12-15). More over, based on sequence-scrambling comparison of 20 random sequences, the FASTP program of Lipman and Pearson (16) yielded the following statistical significance for the observed homologies: (a) The aligned score between CBP35 and GRP33 was 1.52 standard deviations above the mean score; (b) The aligned score between CBP35 and hnRNP protein C1 was 0.96 standard deviations above the mean score; and (c) The aligned score between CBP35 and DL-4 sequences was 0.03 standard deviations above the mean score. For comparison, the aligned score between hnRNP proteins C1 and A1 was 1.04 standard deviations above the mean score.

Finally, CBP35 shares with a number of hnRNP proteins what appears to be one of their typical features, i.e. the presence of distinct domains with non-uniform distribution of Gly and Pro residues in the polypeptide chains. In the case of CBP35, 86% of the 36 Gly residues and 87% of the 39 Pro residues are in the amino terminal domain of the molecule. For both GRP33, the hnRNP protein from brine shrimp (12) and rat hnRNP protein A1 (15), approximately 77% of the total Gly residues are located within the carboxyl terminal 124 amino acid residues. An unequal distribution of Pro has also been observed for human hnRNP protein C1; all of the Pro residues were found in the amino terminal half of the polypeptide chain (14). Therefore, the homology of CBP35 to the sequences of hnRNP proteins and the conservation of certain structural features in distinct domains suggest that the lectin might be one of the hnRNP proteins.

Figure 4: Alignment of the amino acid residues showing internal sequence homology. The amino terminal domain of CBP35 clone 1 shows a repetitive sequence, each consisting of nine amino acids. Numbers at the left and right indicate the amino acid residue in the numbering system of Figure 2.

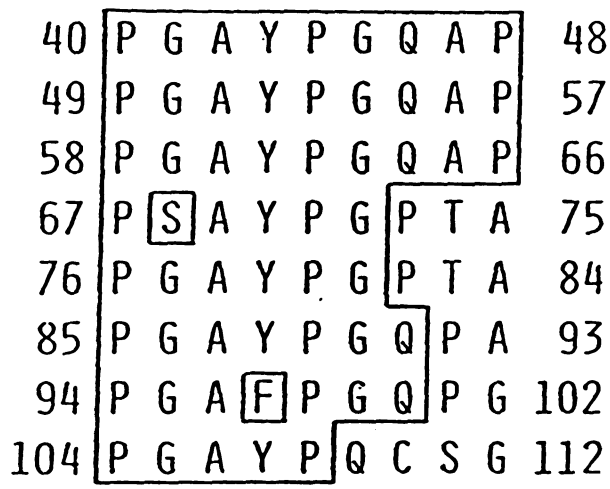

40 P G A Y P G Q A P 48
49 P G A Y P G Q A P 57
58 P G A Y P G Q A P 66
67 P S A Y P G P T A 75
76 P G A Y P G P T A 84
85 P G A Y P G Q P A 93
94 P G A F P G Q P G 102
104 P G A Y P Q C S G 112

Fig. 4

Figure 5: Comparison of the deduced amino acid sequence of CBP35 clone 1 (residues 14 through 120 in the numbering system of Figure 2) with the amino acid sequences of hnRNP proteins. GRP33, glycine-rich protein of the hnRNP of Artemia salina (12), DL-4, sequence deduced from a cDNA clone derived from a human hepatoma library (13), and human hnRNP C1 protein (14). Dashes indicate gaps introduced for optimal alignment. Sequence identities between CBP35 clone 1 and the hnRNP proteins are highlighted by boxes. The sequences were compared using the FASTP program (16).

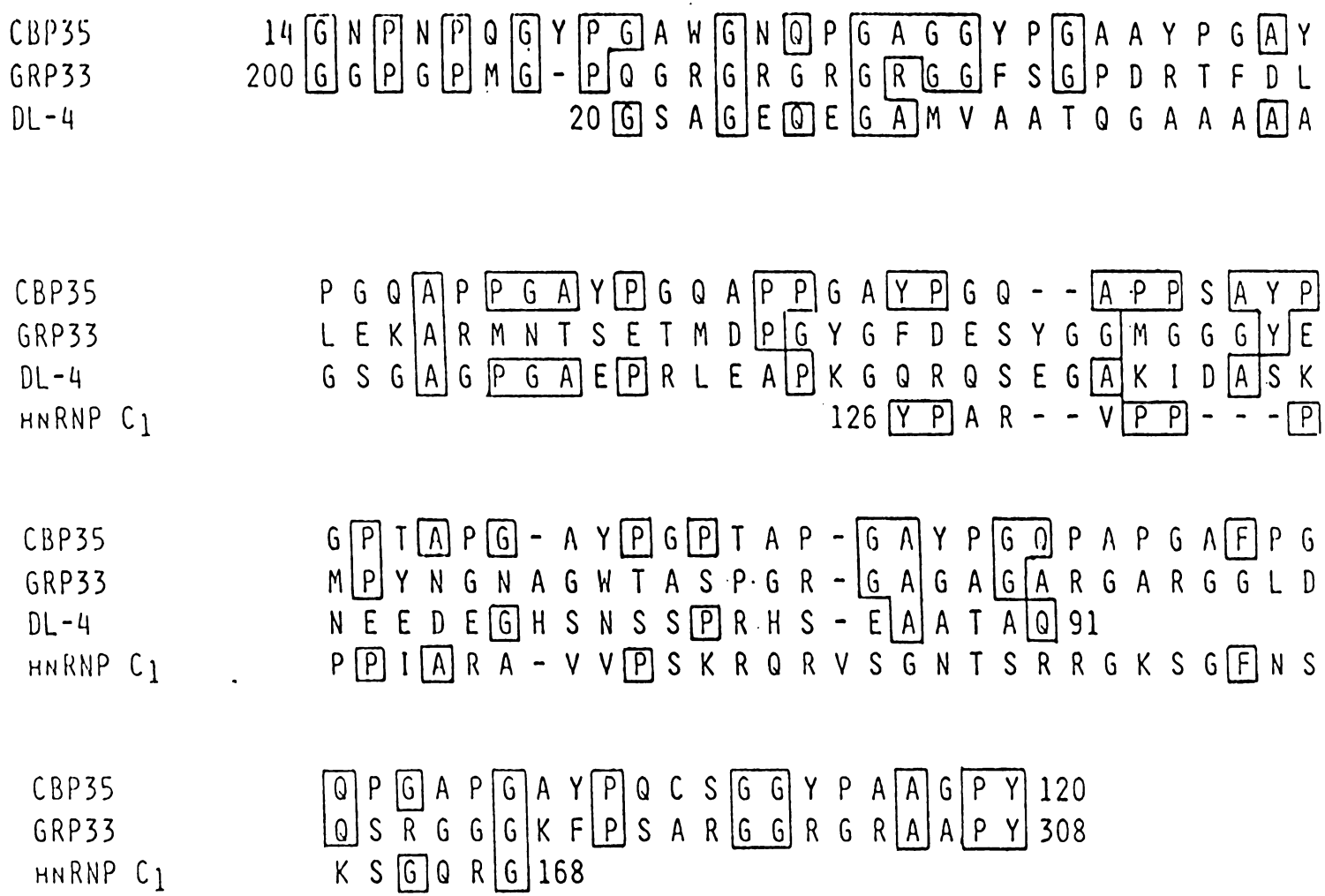

CBP35 14 G N P N P Q G Y P G A W G N Q P G A G G Y P G A A Y P G A Y
GRP33 200 G G P G P M G - P Q G R G R G R G R G G F S G P D R T F D L
DL-4 20 G S A G E Q E G A M V A A T Q G A A A A A
CBP35 P G Q A P P G A Y P G Q A P P G A Y P G Q - - A P P S A Y P
GRP33 L E K A R M N T S E T M D P G Y G F D E S Y G G M G G G Y E
DL-4 G S G A G P G A E P R L E A P K G Q R Q S E G A K I D A S K
HNRNP C1 126 Y P A R - - V P P - - - P
CBP35 G P T A P G - A Y P G P T A P - G A Y P G Q P A P G A F P G
GRP33 M P Y N G N A G W T A S P G R - G A G A G A R G A R G G L D
DL-4 N E E D E G H S N S S P R H S - E A A T A Q 91
HNRNP C1 P P I A R A - V V P S K R Q R V S G N T S R R G K S G F N S
CBP35 Q P G A P G A Y P Q C S G G Y P A A G P Y 120
GRP33 Q S R G G G K F P S A R G G R G R A A P Y 308
HNRNP C1 K S G Q R G 168

Fig. 5

This notion is supported by recent observations that have identified CBP35 in hnRNP (20). Nucleoplasm derived from 3T3 cells was fractionated on a cesium sulfate gradient (1.25-1.75 g/ml); immunoblotting analysis with anti-CBP35 localized the lectin in fractions with densities of 1.30-1.35 g/ml. This range of densities corresponds to the density of hnRNP on cesium sulfat gradients. Conversely, affinity chromatography on the basis of saccharide-binding could also be used to isolate hnRNP. When nucleoplasm derived from [35S]methionine-labeled 3T3 cells was fractionated on a column of Sepharose derivatized with N-e-aminocaproyl-D-galactosamine, the bound and eluted fraction yielded CBP35 as well as a set of polypeptides whose molecular weights matched those reported for the core particle of hnRNP. Moreover, the complex isolated in the saccharide affinity column also contained RNA. These results, along with the present structural data, both suggest that CBP35 is a component of hnRNP.

REFERENCES

1: Roff, C.F., Rosevear, P.R., Wang, J.L., and Barker,R. (1983) Biochem. J. 211, 625-629.

2: Roff, C.F., and Wang, J.L. (1983) J. Biol. Chem. 258, 10657-10663.

3: Moutsatsos, I.K., Davis, J.M., and Wang, J.L. (1986) J. Cell Biol. 102, 477-483.

4: Moutsatsos, I.K., Wade, M., Schindler, M., and Wang, J.L. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 6452-6456.

5: Jia, S., Mee, M.P., Morford, G., Agrwal, N., Voss, P.G., Moutsatsos, I.K., and Wang, J.L. (1987) Gene 60, 197-204.

6: Hopp, T.P., and Woods, K.R. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 3824-3828.

7: Ohyama, Y., Hirabayashi, J., Oda, Y., Ohno, S., Kawasaki, H., Suzuki, K., and Kasai, K. (1986) Biochem. Biophys. Res. Commun. 134, 51-56.

8: Hirabayashi, J., Kawasaki, H., Suzuki, K., and Kasai, K. (1987) J. Biochem. (Tokyo) 101, 775-787.

9: Gitt, M.A., and Barondes, S.H. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 7603-7607.

10: Southan, C., Aitken, A., Childs, R.A., Abbott, W.M., and Feizi, T. (1987) FEBS Lett. 214, 301-304.

11: Paroutaud, P., Levi, G., Teichberg, V.I., and Strosberg, A.D. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 6345-6348.

12: Cruz-Alvarez, M., and Pellicer, A. (1987) J. Biol. Chem. 262, 13377-13380.

13: Lahiri, D.K., and Thomas, J.O. (1986) Nucleic Acids Res. 14, 4077-4094.

14: Swanson, S., Nakagawa, T.Y., LeVan, K., and Dreyfuss, G. (1987) Mol. Cell Biol. 7, 1731-1739.

15: Cobianchi, F., SenGupta, D.N., Zmudzka, B.Z., and Wilson, S.H. (1986) J. Biol. Chem. 261, 3536-3543.

16: Lipman, D.J., and Pearson, W.R. (1985) Science 227, 1435-1441.

17: Messing, J. (1983) Methods in Enzymol. 101, 20-78.

18: Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.

19. Staden, R. (1979) Nucleic Acids Res. 6, 2601-2610.

20. Laing, J. G., and Wang, J. L. (1988) Biochem. 27, 5329-5334.

CHAPTER IV

Nucleotide Sequence of the Murine Gene for Carbohydrate Binding Protein 35*

Shizhe Jia and John L. Wang

Department of Biochemistry
Michigan State University
East Lansing, MI 48824

Running Title: Genomic Sequence of Carbohydrate Binding Protein 35

(This chapter, with some revisions and additions, will be submitted for publication)

FOOTNOTES

* This work was supported by grants GM-38740 and GM-27203 from the National Institutes of Health.

[1] The abbreviations used are: CBP35, Carbohydrate Binding Protein (M_r ~ 35,000); kb, kilobase(s); bp, base pairs; IVS, intervening sequence; SRE, serum responsive element.

[2] Heng-Yin Yang, unpublished results.

SUMMARY

We have isolated a cloned segment of the gene for mouse Carbohydrate Binding Protein 35 (CBP35). Nucleotide sequence analysis of the cloned gene, as well as the results of genomic Southern blot hybridization, revealed that the gene is unique and spans 9 kilobases of genomic DNA. The mRNA for the lectin is encoded by five exons separated by four introns. Intron I is located in the 5′ transcribed but untranslated region of the primary transcript. Examination of the nucleotide sequence 5′ to the transcription initiation site revealed characteristic TATA and CCAAT boxes. A putative serum-responsive element has also been identified about 200 nucleotides upstream from the site for initiation of transcription. This may serve as a binding site for a class of transcription factors responsible for the serum-stimulated expression of the CBP35 gene.

INTRODUCTION

Carbohydrate Binding Protein 35 (CBP35; M_r 35,000), a galactose-specific protein found in many cells and tissues (1,2), belongs to the S-type family of lectins (3). In mouse 3T3 fibroblasts, this protein has been identified in several subcellular locations, predominantly in the cytoplasm and nucleus (4). When quiescent 3T3 cells were stimulated by the addition of serum, there was an increase in the overall level of CBP35, as well as a dramatic rise in the amount of the polypeptide in the nucleus (5).

A cDNA clone for CBP35 has been isolated and sequenced (6). The deduced amino acid sequence showed that the polypeptide consisted of two domains. The NH_2-terminal half was proline- and glycine-rich and showed some structural similarity to proteins of the heterogeneous nuclear ribonucleoprotein complex (hnRNP). The carboxyl terminal half was homologous to many other β-galactoside specific lectins (7-15).

On the basis of structural information, a number of other proteins, identified independently in different laboratories and for distinct reasons, have been found either to be identical to CBP35 or to be a homologue of CBP35 in another species/cell type. These include: (a) L-34, a metastasis-associated lectin isolated from murine fibrosarcomas (15); (b) Mac-2, a cell surface marker linked to macrophage differentiation (16); (c) LBP, the major non-integrin Laminin Binding Protein of murine macrophages (17); and (d) εBP, an IgE-binding protein identified in rat basophilic leukemia cells (18,19). Although only L-34 was originally isolated on the basis

of carbohydrate-binding activity, all the other proteins listed above have now been shown to bind saccharides (15-19).

The majority of the CBP35 of mouse 3T3 fibroblasts is intracellular (4). Although a large proportion of L-34, Mac-2, and εBP are also found inside their respective cells, they have been studied, for the most part, because of their extracellular localization (15,16,20). For these reasons, it was important to compare the structural features of the proteins that might lead to the possibility of dual localization. Thus far, the coding regions of the mature proteins do not offer hints on the mechanism of externalization (6,15,16,18). Nor do the cDNA sequences implicate an obvious signal sequence for targeting to the lumen of the endoplasmic reticulum, a classical endomembrane pathway for secretion. In their studies on Mac-2, Cherayil *et al.* (16) demonstrated the possible existence of two alternatively spliced Mac-2 cDNAs, one of which has the potential to encode a colinear, NH_2 terminally extended Mac-2 protein containing a signal peptide-like sequence. An analysis of the genomic region encoding these cDNAs would provide confirmation of these postulated splicing events. In the present paper, we report the structure of the murine CBP35 gene, paying particular attention to the region 5′ to the translation initiation codon. In the course of these studies, we also found a possible structural basis for the observed serum-stimulated increase in expression of the CBP35 gene.

EXPERIMENTAL PROCEDURES

Screening of the Genomic Libraries

The cDNA for CBP35 (6) was labeled with [^{32}P]dCTP using a random priming kit (Boehringer Mannheim) and the method of Feinberg and Vogelstein (21). The specific activity of the labeled cDNA was ~ 10^8 cpm/μg DNA. This probe was used to screen a λ EMBL4 library derived from mouse liver chromosomal DNA (Clone Tech Laboratory) using procedures described previously (22). Three positive clones were identified. Since these three clones appeared to have the same inserts, only one clone, designated as λG_3, was characterized by restriction mapping and DNA sequence analysis (see Fig. 1).

A 1.7 kb PstI-SalI fragment at the left end of the λG_3 clone (highlighted by a hatched rectangle in Fig. 1) was isolated and labeled with [^{32}P]dCTP as described above. This was then used as a hybridization probe to screen a λ FIXII library, also derived from mouse liver chromosomal DNA (Strategene). The λGH clone was identified and isolated from this screening.

Cloning and Sequencing

The 6.4 kb SalI-EcoRI fragment of λG_3 clone (Fig. 1) was first cloned into the PUC-13 plasmid. After digestion with PstI and HindIII, the individual fragments were subcloned into the sequencing vectors M13mp18 and M13mp19 (23). The 1.7 kb SalI-PstI fragment and parts of the 2.5 kb PstI-HindIII fragment were sequenced using the dideoxy method (24) (Sequenase Kit, United States Biochemical Co.).

Figure 1: Restriction map and sequencing strategy for the murine CBP35 gene. The scale at the top indicates the length in kilobases. Exons are indicated by dark lines/boxes and are numbered I-V. The intervening sequences are designated IVS and are numbered 1-4. The numbers between restriction enzyme sites indicate the size of the restriction fragment. The length and direction of sequencing reactions for the λGH and λG_3 clones are shown by the arrows. The hatched rectangle in the λG_3 clone indicates the fragment that was used as a hybridization probe to identify the λGH clone.

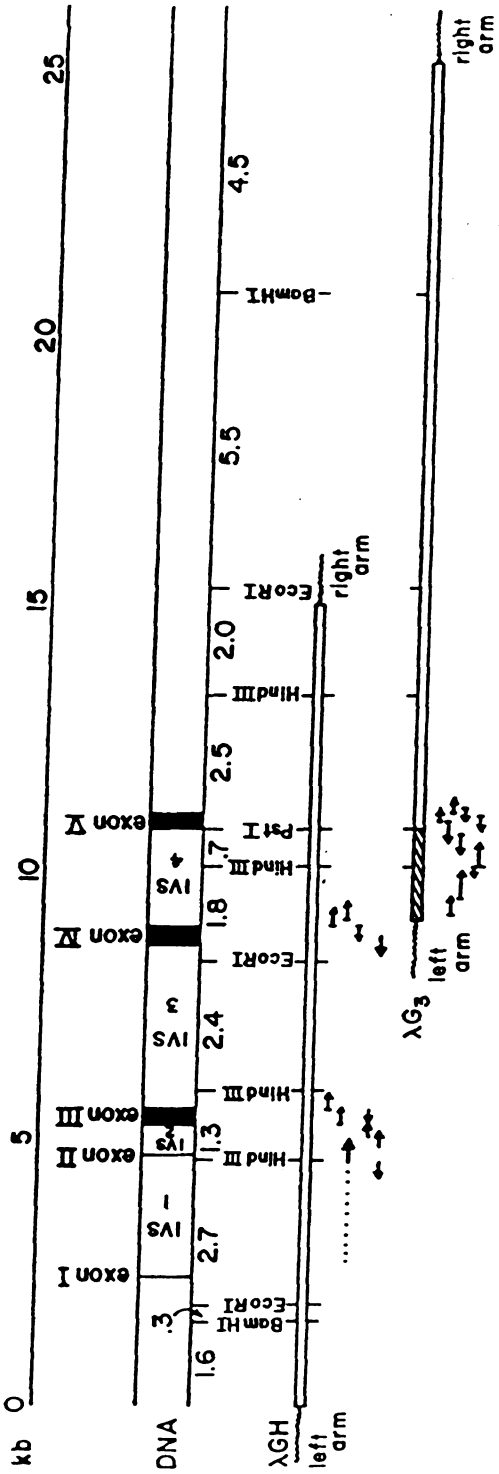

kb
0
5
10
15
20
25
DNA
λGH
left arm
right arm
BamHI
EcoRI
Hind III
Pst I
exon I
exon II
exon III
exon IV
exon V
IVS 1
IVS 2
IVS 3
IVS 4
1.6
.3
2.7
1.3
2.4
1.8
.7
2.5
2.0
5.5
4.5
λG3

Figure 1

The 6.4 kb EcoRI-EcoRI fragment and the 6.7 kb EcoRI-SalI fragment from the λGH clone (Fig. 1) were cloned into the PUC-18 plasmid. After HindIII digestion, these fragments were further subcloned into M13mp18 and M13mp19 vectors (23). Both single-stranded and double-stranded DNA sequencing analyses were performed with the Sequenase Kit. Eight oligonucleotide primers, synthesized on the basis of the cDNA sequence (6), were used in carrying out the sequencing reactions. The oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer (Model 380B) in the Macromolecular Structure Facility at Michigan State University. The areas and directions of the sequencing analysis are shown by the arrows in Figure 1.

Directional Deletion of a 2.7 kb Fragment for DNA Sequencing

The method used followed that described in Sambrook *et al.* (25). The 2.7 kb EcoRI-HindIII fragment of λGH clone (Fig. 1) was digested by Nuclease S1 (Boehringer Mannheim) to blunt end the DNA fragment. This was then cloned into the SmaI site of the M13mp18 vector. The orientation of the insert was determined by restriction analysis. A clone with the original HindIII end toward the SalI site of the M13mp18 multiple cloning site was selected. Approximately 10 μg of M13 replicative form DNA were isolated from this clone and were digested with SalI and PstI. Nuclease S1 (40 U; Boehringer Mannheim) and Exonuclease III (200 U; Boehringer Mannheim) were then added to the double-digested DNA at 37°C. Samples were withdrawn from this digestion mixture at 90, 135, 180, and 225 seconds. The samples were placed at 30°C for 30 minutes. Then, the Klenow fragment of DNA polymerase I (10 U; Boehringer Mannheim) was added

to each sample, along with the dNTPs (0.5 mM each) and incubated at 25°C for 30 minutes to create blunt ends. The DNA samples were finally ligated and transformed into *E. coli*. Twenty-five clones were selected from each transformation and sequenced using universal primer and the Sequenase Kit. A total of about 96 clones were analyzed.

Primer Extension Assay

Quiescent, serum-starved cultures of 3T3 fibroblasts were stimulated by the addition of serum (5). After 18 hours, cytoplasmic RNA was isolated following the protocol of Henikoff (26). The cells were isolated and resuspended on ice for 5 minutes in 10 mM Tris Buffer (pH 8.6) containing 0.14 M NaCl, 1.5 mM $MgCl_2$, 0.5% Triton X-100, and 1000 U/ml of Rnasin. The nuclei were pelleted by centrifugation at 500 x g for 5 minutes. The supernatant was then made 2% in sodium dodecyl sulfate and 50 µg/ml in proteinase K. After 1 hour at 37°C, the supernatant was extracted with phenol: chloroform (v/v, 1:1) once and the RNA was collected by ethanol precipitation. The DNA in the sample was removed by digestion with 2 µg/ml deoxyribonuclease I at 37°C for 1 hour.

The general method for the primer extension assay has been described (25). A 22- nucleotide primer, 5′-CGAAAAGCTGTCTGCCATTTTC-3′ was synthesized and $^{32}PO_4$-labeled at the 5′ end with T4 polynucleotide kinase (2 U/µl) at 37°C for 30 minutes. The labeled primer (10^7 cpm/µg) was isolated by ethanol precipitation and was then allowed to hybridize for 12 hours at 30°C to 30 µg of cytoplasmic RNA isolated from 3T3 cells. Reverse transcriptase (60 U; Boehringer Mannheim) was added at 37°C for two hours. The radioactive product

of the extension reaction was electrophoresed on a 10% polyacrylamide sequencing gel, followed by autoradiography. The length of the extended cDNA was estimated by comparison to a sequencing reaction using a 20-nucleotide primer.

Southern Analysis of Genomic DNA

Mouse liver nuclei were isolated from a homogenate of 15 grams of liver tissue by centrifugation at 500 x g for 5 minutes. The nuclei were suspended in 10 mM Tris (pH 8.0) containing 0.14 M NaCl, 1.5 mM $MgCl_2$, 0.5% sodium dodecyl sulfate and 100 µg/ml proteinase K at 37°C for 16 hours. DNA was isolated by ethanol precipitation and purified by CsCl gradient centrifugation. The DNA was then digested with EcoRI, HindIII and BamHI. The initial digestion used 20 U enzyme per µg of DNA and was carried out for 20 hours. An additional 10 U of enzyme per µg of DNA was added and incubated for 4 more hours. The digested DNA was separated by agarose gel electrophoresis (0.8%), subjected to Southern blotting (27) using ^{32}P-labeled CBP35 cDNA (10^8 cpm/µg) as a hybridization probe.

RESULTS

Isolation and Characterization of Genomic Clones

A λ EMBL4 library derived from mouse liver chromosomal DNA was screened using the cDNA for CBP35 (6) as a hybridization probe. Of the 10^6 plaques screened, three gave reproducible positive signals. One of these, designated as λG_3, was further cloned by lower density rescreening and plaque purification. Digestion of λG_3 with PstI and sequence analysis in both directions (Fig. 1) revealed the single PstI site in the nucleotide sequence of the cDNA (6). Comparison of the genomic sequence with the cDNA sequence defined exon V, corresponding to the 3′ end of the cDNA clone.

The 1.7 kb fragment, derived from the "left" end of λG_3 by PstI digestion, was sequenced in its entirety[2], without finding any additional sequence matching the cDNA. This suggested that the 1.7 kb fragment was inside an intron of the genomic DNA. Thus, this 1.7 kb fragment of λG_3 (highlighted by a hatched rectangle in Fig. 1) was used as a hybridization probe to search for other genomic clones from the λ FIXII library, also derived from mouse liver chromosomal DNA. Clone λGH was selected from such a screening.

Analysis of restriction enzyme cleavages and of the partial nucleotide sequence showed that the λGH and λG_3 clones overlapped over a region of approximately 6 kb, corresponding to the "right" end of λGH and the "left" end of λG_3 (Fig. 1). When the cDNA for CBP35 (6) was digested with PstI, a fragment corresponding to the 3′ end (nucleotides 696-883) of the cDNA was isolated. This

Figure 2: Southern blot hybridization of mouse genomic DNA with the cDNA clone for CBP35. Chromosomal DNA was isolated from mouse liver nuclei. The DNA was digested with restriction enzymes, separated on agarose gels (0.8%) and subjected to Southern blot analysis with ^{32}P-labeled CBP35 cDNA (10^8 cpm/μg). Panel A: lane 1, BamHI digest; lane 2, HindIII digest; lane 3, EcoRI digest; and lane 4, undigested DNA. In panel B, the same EcoRI digest sample that was used for lane 3 of panel A was electrophoresed for a longer time to better resolve the bands of a doublet.

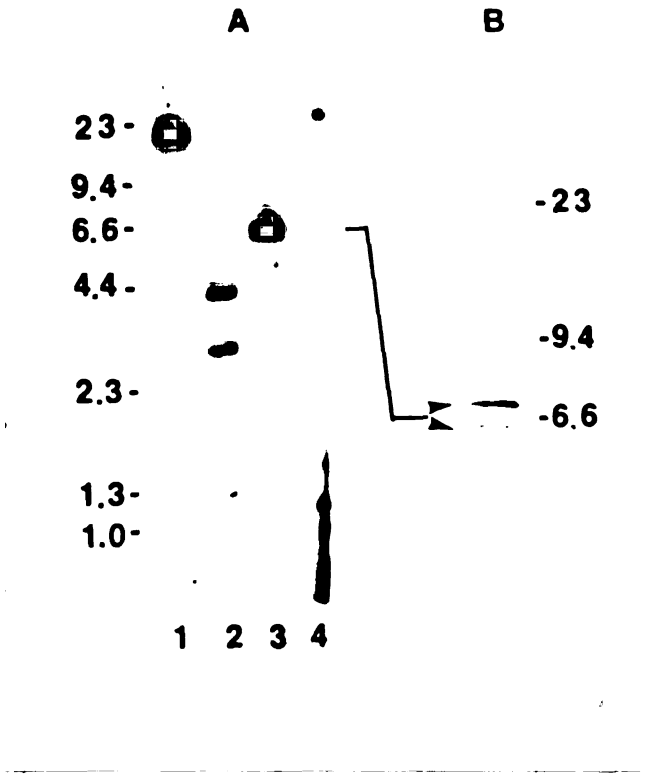


Figure 2

fragment hybridized to both λGH and λG_3 clones. In contrast, the other fragment of the cDNA generated by PstI (nucleotides 1-695) hybridized to the λGH clone, but not the λG_3 clone. Therefore, it appears that the overlapping region between λGH and λG_3 contained the 3′ end of the cDNA, while the 5′ end of the cDNA was found in the remainder of the λGH clone (Fig. 1). Nucleotide sequence and primer extension analyses (see below) indicate that the gene for CBP35, including several consensus sequences for 5′ regulatory elements, can be entirely mapped onto the λGH clone.

DNA isolated from nuclei of mouse liver was digested with several restriction enzymes, separated by gel electrophoresis and then subjected to Southern blot analysis with the cDNA for CBP35. A single fragment (19-20 kb) was observed after BamHI digestion (Fig. 2A, lane 1). HindIII digestion yielded fragments of 4.2 kb, 3.2 kb, and 1.3 kb (Fig. 2A, lane 2). Although the EcoRI digested material yielded a broad band in the experiment shown in Figure 2A (lane 3), it actually contained two fragments, which could be resolved into a doublet upon longer electrophoresis of the same material (Fig. 2B). The positions of migration of the bands corresponded to DNA molecules of ~ 6.4 kb and ~ 7.0 kb. The lengths of each of these genomic DNA fragments that hybridized with the cDNA probe were in good agreement with those derived from the genomic clones isolated above (Fig. 1). No other bands could be observed. These results indicate that the CBP35 gene is a single gene in the normal mouse genome.

Sequence Analysis of the Genomic Clones

Using a strategy similar to that described for the analysis of the

PstI fragment of clone λG_3 (highlighted by a hatched rectangle in Fig. 1), nucleotide sequences were determined for parts of the λGH clone (see Fig. 1). Figure 3 reports the nucleotide sequences of the 5′ and 3′ flanking regions, the exons, and the exon/intron boundaries of the CBP35 gene. The beginning of exon I was identified by carrying out directional deletion on the 2.7 kb EcoRI-HindIII fragment of λGH, cloning each of the resulting fragments, and sequencing approximately the first 100 nucleotides of each of the 90 or so clones obtained (highlighted by dotted line in Fig. 1). This revealed a sequence CTTCCG, fitting the consensus for a mRNA cap site (rectangle highlighted by a wavy underline in Fig. 3). On the basis of this landmark, the initiation site of transcription was deduced to be an adenine residue (circled and labeled +1 in Fig. 3).

The partial sequences of the 2.7 kb EcoRI-HindIII fragment (highlighted by dotted line in Fig. 1) also revealed a 2.3 kb intron (IVS1 in Fig. 3). On the 3′ side of this IVS1, exon II starts with the sequence GAAAATGG. This ATG triplet (solid triangle in Fig. 3) falls within the consensus translation initiation sequence A/GNN-ATG-G (28). On the basis of comparing the nucleotide sequence obtained in this study with that of the cDNA (6), the present assignment of methionine to this initiation codon revealed: (a) the cDNA clone coded for the entire CBP35 polypeptide chain except for the NH_2-terminal Met residue; and (b) the second amino acid of the polypeptide should be Ala rather than the Arg that was reported previously (6). These conclusions are consistent with the results of the nucleotide and amino acid sequences reported for Mac-2 (16), L-34 (15), and εBP (18).

Given these assignments of the start of exon I and the translation

Figure 3: Nucleotide sequence of the murine CBP35 gene. The numbering system, shown on the left, is based on the transcription initiation site; this is the A residue highlighted by a circle and whose position is labeled +1. The nucleotide sequences of the introns, denoted IVS 1-4, are not shown except for the 5′ and 3′ ends to highlight the conservation of the donor and acceptor splice site consensus sequences. The sequence CCAATTAAGG, highlighted by a rectangle with a double underline, represents a putative Serum Response Element. The sequence CCAAT, highlighted by a rectangle with a single solid underline, and the sequence AATATATAT, highlighted by a rectangle, represent sequences that fit the consensus sequences and locations of CCAAT and TATA boxes, respectively. The sequence CTTCCG, highlighted by a rectangle with a wavy underline, represents a putative mRNA cap site. The translation initiation codon, ATG, is highlighted by a triangle. The open reading frame extends to a termination codon, TAA, highlighted by an inverted triangle. The sequence AATAAA, highlighted by a rectangle with a dotted underline, represents the polyadenylation signal.

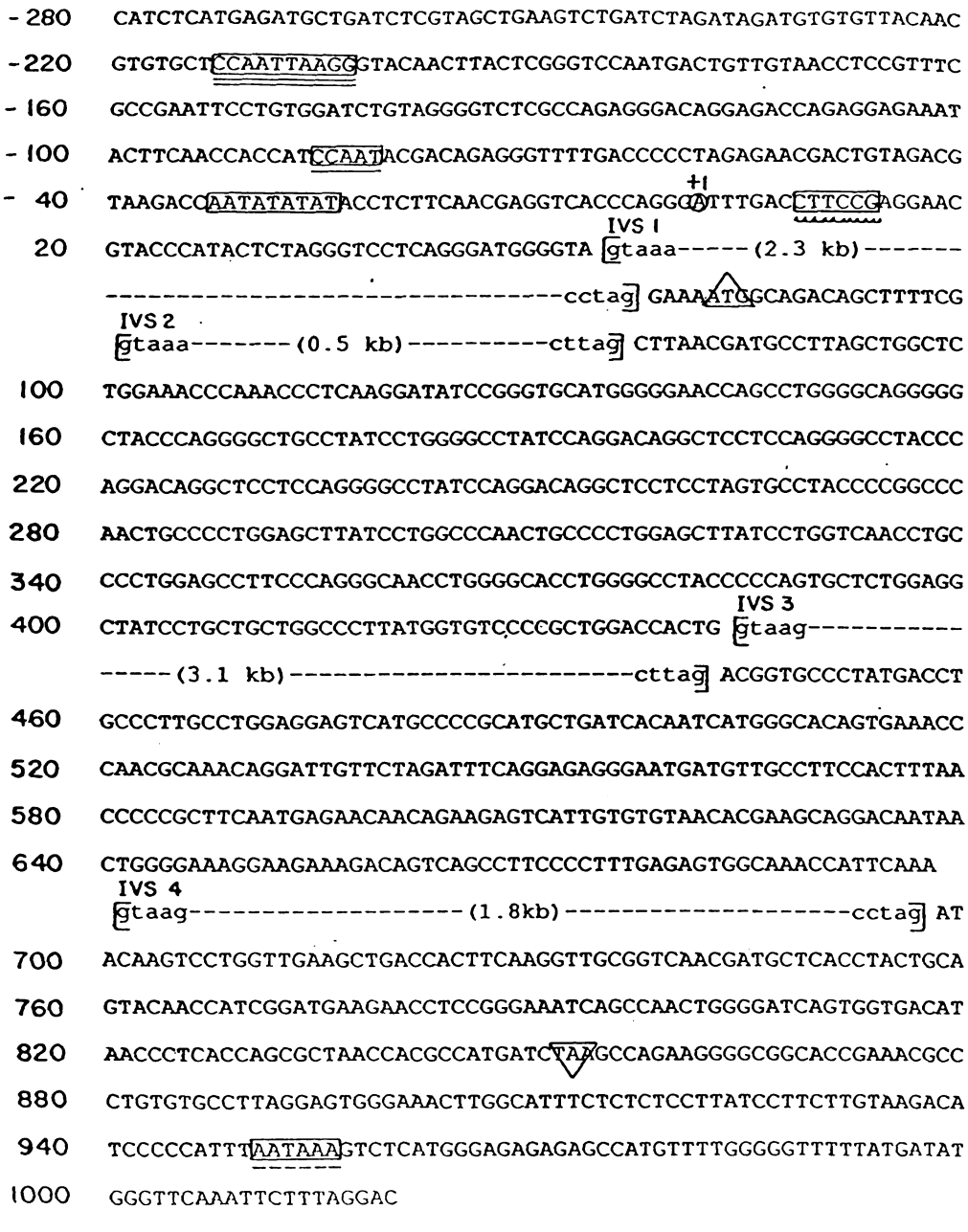

-280 CATCTCATGAGATGCTGATCTCGTAGCTGAAGTCTGATCTAGATAGATGTGTGTTACAAC
-220 GTGTGCTCCAATTAAGGGTACAACTTACTCGGGTCCAATGACTGTTGTAACCTCCGTTTC
-160 GCCGAATTCCTGTGGATCTGTAGGGGTCTCGCCAGAGGGACAGGAGACCAGAGGAGAAAT
-100 ACTTCAACCACCATCCAATACGACAGAGGGTTTTGACCCCCTAGAGAACGACTGTAGACG
-40 TAAGACCAATATATATACCTCTTCAACGAGGTCACCCAGGGATTTGACCTTCCGAGGAAC
+1
IVS 1
20 GTACCCATACTCTAGGGTCCTCAGGGATGGGGTAgtaaa-----(2.3 kb)-------
cctag GAAAATGGCAGACAGCTTTTCG
IVS 2
gtaaa-------(0.5 kb)----------cttag CTTAACGATGCCTTAGCTGGCTC
100 TGGAAACCCAAACCCTCAAGGATATCCGGGTGCATGGGGGAACCAGCCTGGGGCAGGGGG
160 CTACCCAGGGGCTGCCTATCCTGGGGCCTATCCAGGACAGGCTCCTCCAGGGGCCTACCC
220 AGGACAGGCTCCTCCAGGGGCCTATCCAGGACAGGCTCCTCCTAGTGCCTACCCCGGCCC
280 AACTGCCCCTGGAGCTTATCCTGGCCCAACTGCCCCTGGAGCTTATCCTGGTCAACCTGC
340 CCCTGGAGCCTTCCCAGGGCAACCTGGGGCACCTGGGGCCTACCCCCAGTGCTCTGGAGG
IVS 3
400 CTATCCTGCTGCTGGCCCTTATGGTGTCCCCGCTGGACCACTGgtaag----------
-----(3.1 kb)----------cttag ACGGTGCCCTATGACCT
460 GCCCTTGCCTGGAGGAGTCATGCCCCGCATGCTGATCACAATCATGGGCACAGTGAAACC
520 CAACGCAAACAGGATTGTTCTAGATTTCAGGAGAGGGAATGATGTTGCCTTCCACTTTAA
580 CCCCCGCTTCAATGAGAACAACAGAAGAGTCATTGTGTGTAACACGAAGCAGGACAATAA
640 CTGGGGAAAGGAAGAAAGACAGTCAGCCTTCCCCTTTGAGAGTGGCAAACCATTCAAA
IVS 4
gtaag----------(1.8kb)----------cctag AT
700 ACAAGTCCTGGTTGAAGCTGACCACTTCAAGGTTGCGGTCAACGATGCTCACCTACTGCA
760 GTACAACCATCGGATGAAGAACCTCCGGGAAATCAGCCAACTGGGGATCAGTGGTGACAT
820 AACCCTCACCAGCGCTAACCACGCCATGATCTAAGCCAGAAGGGGCGGCACCGAAACGCC
880 CTGTGTGCCTTAGGAGTGGGAAACTTGGCATTTCTCTCTCCTTATCCTTCTTGTAAGACA
940 TCCCCCATTTAATAAAGTCTCATGGGAGAGAGAGCCATGTTTTGGGGGTTTTTATGATAT
1000 GGGTTCAAATTCTTTAGGAC

Figure 3

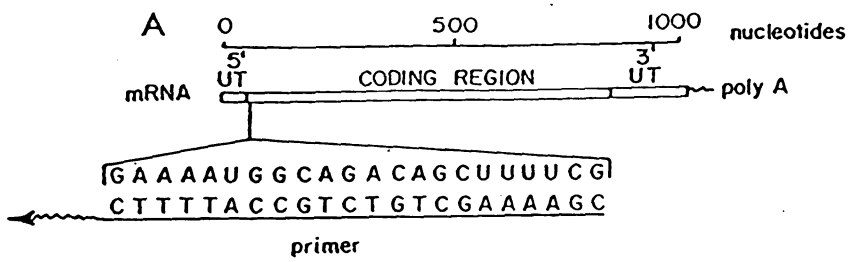
initiation codon in exon II, there must necessarily be an intron in the 5′ transcribed but untranslated region of the primary transcript, a somewhat unusual occurrence. This conclusion is supported by preliminary experiments that indicate an oligonucleotide probe synthesized on the basis of a nucleotide sequence in IVS1 failed to hybridize to the 1.3 kb mRNA, previously identified with our cDNA clone (29,30). The sequence results also predict that the 5′ untranslated region of the CBP35 mRNA will be 57 nucleotides long. A primer extension experiment was performed to ascertain this length. A 22-mer oligonucleotide complementary to the entire exon II (Fig. 3) was synthesized, labeled with ^{32}P, and was used as the primer for reverse transcription of the mRNA (Fig. 4A). The results showed that the product of the extension reaction contained 72 nucleotides (Fig. 4B), consistent with the notion that the transcription initiation site was located ~ 54 bases upstream from the initiation ATG codon in the mRNA.

Structural Features of the CBP35 Gene

The CBP35 gene spans ~ 9 kb of genomic DNA and contains five exons (Fig. 1): (a) exon I, 53 bp; (b) exon II, 22 bp; (c) exon III, 366 bp; (d) exon IV, 255 bp; and (e) exon V, ~ 323 bp (the exact size of last exon is not known; see below). These exons are interrupted by four introns: (a) IVS1, ~ 2.3 kb; (b) IVS2, ~ 0.5 kb; (c) IVS3, ~ 3.1 kb; and (d) IVS4, ~ 1.8 kb. The RNA donor and acceptor splice sites are characterized by GTAAA/G ---- CC/TTAG sequences (Fig. 3), conforming to the GT - AG rule (31).

The Met initiation codon is located in exon II (residues 58-60, solid triangle, Fig. 3), characterized by the consensus initiation

Figure 4: Identification of the transcription initiation site of the murine CBP35 gene by primer extension assay. A) The primer was synthesized to be complementary to the sequence of exon II (Figure 3), including four nucleotides of the 5′ untranslated region, the AUG initiation codon, and 15 additional nucleotides of the coding region. B) The product of the primer extension reaction was separated on a 10% polyacrylamide gel, followed by autoradiography. Lane 1, product of the extension reaction using primer synthesized as shown in panel A. Lane 2, product of the extension reaction using tRNA as a control. The four lanes marked G, C, T, and A on the right are size markers from a sequencing reaction. The numbers on the right indicate the length of the nucleotide that would migrate to the position on the gel. The number on the left highlights the length of the nucleotide determined for the primer extended product.


A
0
500
1000
nucleotides
5'
3'
mRNA
UT
CODING REGION
UT
poly A
GAAAAUGGCAGACAGCUUUUCG
CTTTTACCGTCTGTCGAAAAGC
primer

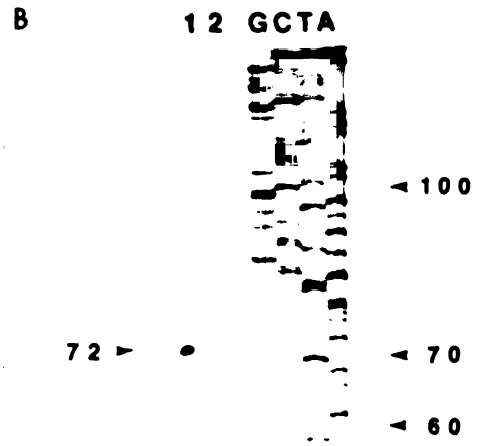

B
12 GCTA
100
72
70
60

Figure 4

sequence A/GNN-ATG-G. From this initiation codon through the termination codon at positions 851-853 (inverted triangle, Fig. 3), there is a translation open reading frame coding for a polypeptide of 264 amino acids. Thus, the coding sequence accounts for 795 nucleotides. As discussed above, the 5′ untranslated region of the mRNA contains 57 nucleotides. In the 3′ untranslated region, there is an AATAAA sequence (rectangle highlighted by dotted underline in Fig. 3), characteristic of the consensus polyadenylation signal (32), located 103 nucleotides downstream from the translation termination codon. Although we have an additional 64 nucleotides of sequence information, in terms of the genomic sequence, we have not determined the precise site of polyadenylation. Assuming the typical distance between the polyadenylation signal and the site of poly A addition to be ~ 30 nucleotides, the size of exon V is estimated to be 290-323 bp.

As indicated previously, the 5′ end of the CBP35 gene was an adenine residue (circled and labeled +1 in Fig. 3), deduced to be the initiation of transcription site on the basis of a consensus sequence for mRNA capping, CTTCCG, located six residues downstream (rectangle highlighted by a wavy underline in Fig. 3). At -34 through -26, a TATA box-like sequence, AATATATAT (rectangle in Fig. 3), was found. This agreed well with the customary location of such a promoter sequence (31,33). Approximately 80 nucleotides upstream from the adenine residue described above, the sequence CCAAT (-86 to -82) was found (rectangle highlighted by a single solid underline in Fig. 3).

In previous studies, we had demonstrated that the expression of the gene for CBP35 was stimulated upon addition of serum to serum-

starved, quiescent cultures of 3T3 fibroblasts (29,30). This stimulation resulted in an increase in the nuclear transcription of the CBP35 gene and in the accumulation of its mRNA, early in the activation process. Moreover, both of these increases were not dependent on *de novo* protein synthesis, inasmuch as they occurred even in the presence of cycloheximide. From studies on the genomic sequences of several serum-stimulated genes, a regulatory element designated as Serum Responsive Element (SRE) had been identified (34,35). The SRE is characterized by the consensus sequence, C-C-A-A/T-A-T-A/T-A/T-G-G, to which certain transcription factors bind. A search for a possible SRE in the genomic sequence of CBP35 derived from the present study resulted in a candidate; this sequence CCAATTAAGG is located at positions -213 through -204 (rectangle highlighted by a double underline in Fig. 3).

Comparisons of the Sequences of CBP35 with L-34, Mac-2, and εBP

We had previously shown (19) that the amino acid sequences of CBP35 and εBP were identical in 223 out of the 262 positions compared (85% identity). Analyses of antibody cross-reactivity and demonstration of carbohydrate-binding activity in εBP established that CBP35 and εBP were mouse and rat homologous, respectively. Since that report, the sequences of two other mouse proteins have been published, L-34 (15) and Mac-2 (16). The nucleotide sequences of CBP35 and L-34 are compared in Figure 5, starting at the 5′ end with the translation initiation codon through all the nucleotide sequence information available at the 3′end. There were very few differences between CBP35 and L-34 at the nucleotide level. However, because the sequences are best aligned by a few insertions/deletions (Fig. 5), the reading frame changes between the proteins. This results in

Figure 5: Comparison of the nucleotide sequence of the coding region of the CBP35 gene with the corresponding sequence reported for L-34. The sequence for CBP35, in capital letters, was taken from the cDNA sequence (reference 6) and was revised on the basis of the present work on the genomic sequence. The sequence for L-34 was taken from reference 15. A dot at any position indicates that the sequence of L-34 was identical to that of CBP35; where the sequence of L-34 differed from that of CBP35, the L-34 sequence is shown in lower case lettering. Certain deletions in either the CBP35 sequence (hyphen) or the L-34 sequence (underline) were included in this comparison in order to maximize the alignment of identities.

```
CBP35 ATGGCAGACAGCTTTTCGCTTAACGATGCCTTAGCTGGCT
L-34  ..........cg..........................
      CTGGAAACCCAAACCCTCAAGGATATCCGGGTGCATGGGG
      ........................................
      GAACCAGCCTGGGGCAGGGGGCTACCCAGGGGCTGCCTAT
      ........................................
      CCTGGGGCCTATCCAGGACAGGCTCCTCCAGGGGCCTACC
      ........................................
      CAGGACAGGCTCCTCCAGGGGCCTATCCAGGACAGGCTCC
      ........................................
      TCCTAGTGCCTACCCCGGCCCAACTGCCCCTGGAGCTTAT
      ..................._.....................
      CCTGGCCCAACTGCCCCTGGAGCTTATCCTGG-TCAACCT
      ..................................c.....
      GCCCCTGGAGCCTTCCCAGGGCAACCTGGGGCACCTGGGG
      ........................................
      CCTACCCCCAGTGCTC-TGGAGGCTATCCTGCTGCTGGCC
      ........_........c.......................
      CTTATGGTGTCCCCGCTGGACCACTGACGGTGCCCTATGA
      ........................................
      CCTGCCCTTGCCTGGAGGAGTCATGCCCCGCATGCTGATC
      ........................................
      ACAATCATGGGCACAGTGAAACCCAACGCAAACAGGATTG
      ........................................
      TTCTAGATTTCAGGAGAGGGAATGATGTTGCCTTCCACTT
      ........................................
      TAACCCCCGCTTCAATGAGAACAACAGAAGAGTCATTGTG
      ........................................
      TGTAACACGAAGCAGGACAATAACTGGGGAAAGGAAGAAA
      ........................................
      GACAGTCAGCCTTCCCCTTTGAGAGTGGCAAACCATTCAA
      ........................................
      AATACAAGTCCTGGTTGAAGCTGACCACTTCAAGGTTGCG
      ........................................
      GTCAACGATGCTCACCTACTGCAGTACAACCATCGGATGA
      ........................................
      AGAACCTCCGGGAAATCAGCCAACTGGGGATCAGTGGTGA
      ........................................
      CATAACCCTCACCAGCGCTAACCACGCCATGATCTAAGCC
      ........................................
      AGAAGGGGCGGCACCGAAAC--GCCCTGTGTCCTTAGGAT
      ..................cg....................
      GGGAAACTTGGCATTTCTCTCTCCTTATCCTTCTTGTAAG
      ........................................
      ACATCCCCCATTTAATAAAGTCTCATGG-GAGAGAGAGCC
      ............................t....a.ag
      ATGTTTTGGGGGTTTTTATGATATGGGTTCAAATTCTTT

      GGAC
```

Figure 5

clusters of differences at the level of the amino acid sequences, as summarized in Table I.

In the study on the Mac-2 antigen (16), it was reported that the sequence was identical to that of CBP35 in the coding region, with the exception of the second amino acid (Ala for Mac-2 and Arg for the previous report on CBP35). This difference was attributed to a sequencing error (16). Our present genomic nucleotide sequence confirms this notion and, therefore, our revision of the second amino acid to an Ala residue makes the CBP35 and Mac-2 coding sequences identical. We also infer, then, that the Mac-2 sequence differs from the sequence of L-34 at the same positions as summarized in Table I. In the coding regions of all four structures (CBP35, Mac-2, L-34, and εBP), there was no sequence obviously corresponding to a signal peptide for sequestration into the endoplasmic reticulum (36). No potential N-linked carbohydrate attachment sites (Asn-Xxx-Ser/Thr) were found.

In contrast to the high degree of identity/homology observed in the coding regions of CBP35, L-34, Mac-2, and εBP, the nucleotide sequence of the CBP35 5′ untranslated region is quite distinct from those of the corresponding region in the mRNA of L-34, Mac-2, and εBP (Fig. 6). These latter three mRNAs share a high degree of identity for the sequences available in the 5′ untranslated regions. The reason for this difference between the CBP35 mRNA and the other three mRNAs is not clear. Oligonucleotides synthesized on the basis of specific sequences in L-34 and in Mac-2 mRNA (Fig. 6) failed to hybridize to the 6.4 kb EcoRI-EcoRI fragment and to the 2.7 kb EcoRI-HindIII fragment of the λGH clone (Fig. 1).

Table I: Differences in the Amino Acid Sequences of CBP35 Versus L-34

Residue Number	CBP35[a]	L-34[b]
4	Ser	Thr
92	Gln	Ser
93	Pro	Thr
110	Gln	Ser
111	Cys	Ala
112	Ser	Pro
219	Glu	Ala
221	Asp	Glu
222	His	Pro

[a] The sequence for CBP35 was taken from reference (6) and was revised by adding NH_2-terminal Met and by substituting Ala for Arg in the second residue.

[b] The sequence for L-34 was taken from reference (15).

Figure 6: Comparison of the nucleotide sequences of the 5′ untranslated region of L-34, Mac-2, εBP, and CBP35. The numbering system, shown on the right, is based on the ATG translation initiation codon; thus, the A residue immediately upstream in the 5′ untranslated region is numbered -1. The reference structure is the sequence reported for L-34 (15), which extends 144 nucleotides 5′ to the translation initiation codon. The end points of the sequence information available from the Mac-2 (16) and εBP (18) cDNA clones are indicated by the brackets and their position numbers. In those positions for which sequence information is available for Mac-2 and εBP, a blank space indicates that the sequence is identical to that of L-34. At any position, only nucleotides that differed from the L-34 reference are shown. Certain deletions (hyphens) were introduced to maximize the alignment of identities. The ^ sign between residues -4 and -5 in the Mac-2 sequence indicates the position where a 27 nucleotide insert has been found in the 5′ untranslated region of certain Mac-2 cDNAs. The shaded region denotes the sequence on the basis of which oligonucleotide probes specific to the 5′ untranslated region of L-34 and Mac-2 mRNAs were synthesized. The nucleotide sequence of the 5′ untranslated region of the CBP35 mRNA, deduced from the genomic sequence obtained in this study, is considerably different from those of L-34, Mac-2, and εBP, and is displayed in its entirety.

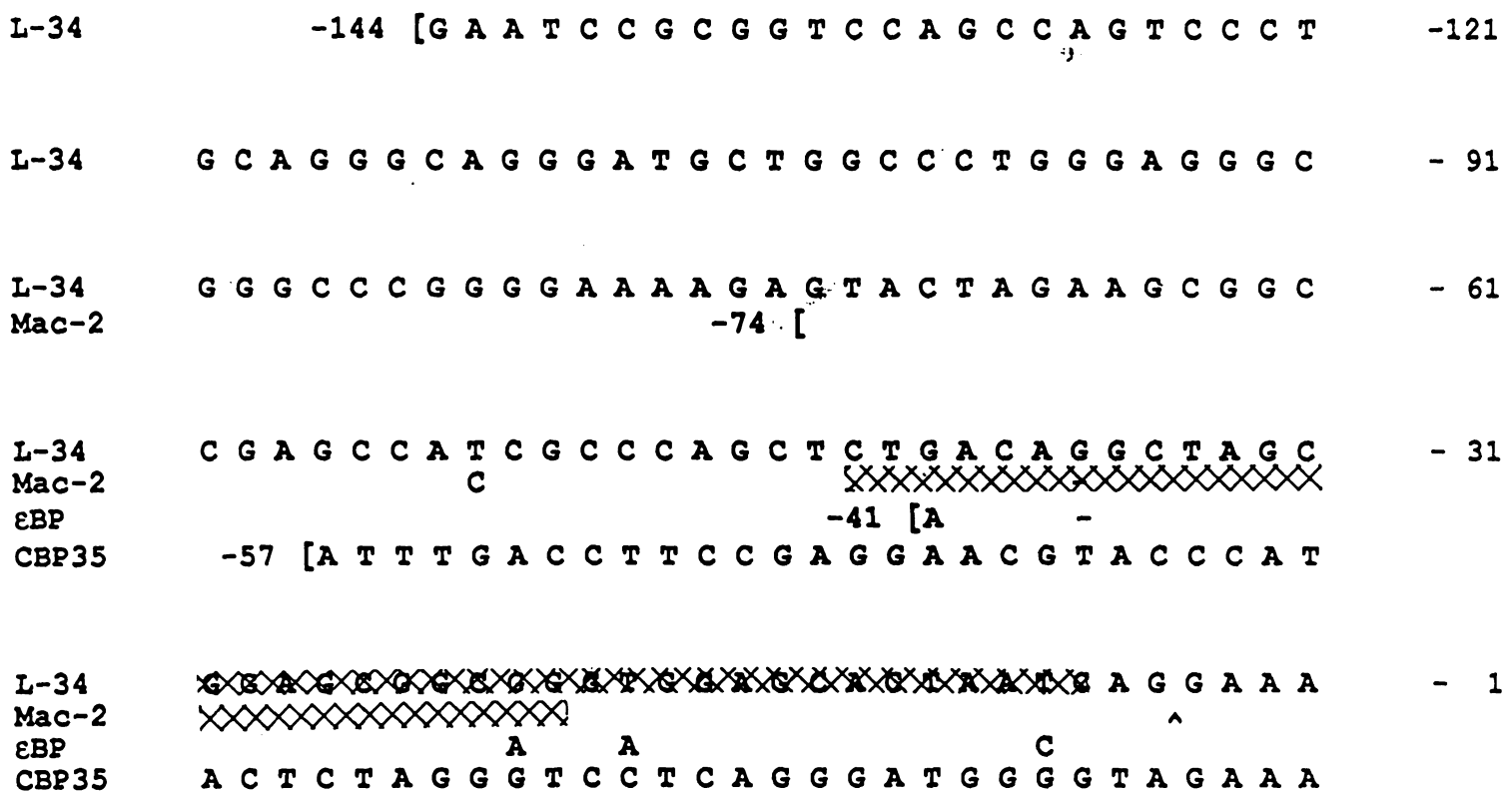

L-34 -144 [G A A T C C G C G G T C C A G C C A G T C C C T -121
L-34 G C A G G G C A G G G A T G C T G G C C C T G G G A G G G C - 91
L-34 G G G C C C G G G G A A A A G A G T A C T A G A A G C G G C - 61
Mac-2 -74 [
L-34 C G A G C C A T C G C C C A G C T C T G A C A G G C T A G C - 31
Mac-2 C
εBP -41 [A -
CBP35 -57 [A T T T G A C C T T C C G A G G A A C G T A C C C A T
L-34 A G G A A A - 1
Mac-2
εBP A A C
CBP35 A C T C T A G G G T C C T C A G G G A T G G G G T A G A A A

Figure 6

DISCUSSION

The results documented in this study indicate that one functional murine CBP35 gene spans ~ 9 kb of genomic DNA. Five exons, separated by four intervening sequences, have been identified. The nucleotide sequences of exons II through V provide the amino acid sequence for the coding region and are in good agreement with the structure derived from the cDNA clone for CBP35 (6), as well as the corresponding information for L-34 (15), Mac-2 (16), and εBP (18). The complete nucleotide sequence of IVS4 has been determined.[2] In addition, partial sequences of IVS1 were obtained by sequencing the first 100 nucleotides, or so, of directionally deleted fragments of the 2.7 kb EcoRI-HindIII fragment of λGH. The sequences of these introns failed to show any identity/homology to each other or to any structures in the gene bank.

Besides confirming the coding region structure, the genomic sequence also provided important information on the regulatory elements upstream from the site of initiation of transcription of the CBP35 gene. Both a TATA box-like sequence and a CCAAT box were found, at positions expected for these structures on the basis of a knowledge of previously characterized genes (31,33). The genomic sequence also revealed a candidate for a SRE-like structure and, therefore, a possible binding site for serum-responsive transcription factors. Such SRE and associated factors have been defined in certain genes, such as c-fos and β-actin, that are activated in quiescent cells upon serum addition (34,35). Addition of serum to serum-deprived 3T3 fibroblasts increased the expression of CBP35 as shown by: (a)

analysis at the single cell level by immunofluorescence (5), (b) analysis at the protein level by Western blotting (5); (c) analysis at the mRNA level by Northern blotting (29,30); and (d) analysis of the transcription of the CBP35 gene in nuclear run-off experiments (30). The sequence in CBP35, CCAATTAAGG, differs from the consensus SRE sequence, C-C-A-A/T-A-T-A/T-A/T-G-G (34,35), at one position. An A residue in position 5 of the 10 nucleotide consensus is replaced by a T residue in the CBP35 gene.

Of particular interest in this genomic sequence is the 5′ untranslated region of the mRNA. This is because of the fact that, although CBP35, L-34, Mac-2, and εBP are all predominantly intracellular proteins, there is also evidence that L-34 and Mac-2 are found on the outside of the cell. In this context, the situation is very similar to that reported for probasin, a rat prostatic protein that was found both in secretions and in the nucleus of prostatic epithelial cells (37). Only one probasin mRNA could be detected by primer extension and S1 nuclease protection assays. *In vitro* translation of this mRNA demonstrated that a protein containing a signal sequence (36) and a protein lacking a signal sequence were synthesized by initiation at different AUG codons. Since the coding regions of CBP35, L-34, Mac-2, and εBP were all the same, it was hoped that the 5′ untranslated region might offer a hint concerning the mechanism of dual localization of the protein, intracellularly and extracellularly.

Indeed, Cherayil *et al.* (16) have demonstrated the possible existence of two alternatively spliced Mac-2 cDNAs, one of which can use CTG (amino acid Leu) as an initiation codon and can encode an additional 13 amino acids before the classical Met translation

initiator. This NH_2-terminally extended segment contains uncharged residues and could possibly serve as a signal peptide (36). Our present genomic sequence results have not generated sufficient information either to confirm or to refute the hypothesized alternative splicing events. This is because, unlike the high degree of identity/homology observed in the coding regions of CBP35, L-34, Mac-2, and εBP, the sequence of the 5′ untranslated region of the CBP35 mRNA, predicted from the presently available genomic information, is totally different from those reported for L-34, Mac-2, and εBP. It does not appear that the sequences in the 5′ untranslated regions of these latter cDNAs are in our λGH clone, inasmuch as oligonucleotide probes synthesized according to these sequences failed to hybridize to the clone.

In light of these results, it is important to consider whether the genomic clone that we have sequenced actually codes for a functional CBP35 gene. The evidences that are brought to bear include: (a) The amino acid sequence is identical to those reported for Mac-2; no substitutions were found in the coding region. (b) All exon/intron boundary sequences are in good agreement with the GT-AG rule. (c) The relative positions of the putative TATA box and CCAAT box correspond well with similar sequences of other genes. (d) A polyadenylation signal was found in the 3′ untranslated region. (e) There is a putative SRE-like sequence, which may account for the observed behavior of the CBP35 gene upon serum stimulation of quiescent cells. (f) Most importantly, we have obtained preliminary evidence that an oligonucleotide complementary to nucleotide residues 44-53 and residues 54-63 (i.e. it bridges exon 1 and exon 2 in Figure 3) hybridizes to the 1.3 kb mRNA identified by our cDNA clone in Northern blots (29,30). This is consistent with the

results of our primer extension assay and lends strong support to the notion that the 2.3 kb IVS1 is spliced out in the processing of the mRNA.

REFERENCES

1. Roff, C.F., and Wang, J.L. (1983) J. Biol. Chem. 258, 10657-10663.
2. Crittenden, S.L., Roff, C.F., and Wang, J.L. (1984) Mol. Cell. Biol. 4, 1252-1259.
3. Drickamer, K. (1988) J. Biol. Chem. 263, 9557-9560.
4. Moutsatsos, I.K., Davis, J.M., and Wang, J.L. (1986) J. Cell Biol. 102, 477-483.
5. Moutsatsos, I.K., Wade, M., Schindler, M., and Wang, J.L. (1987) Proc. Natl. Acad. Sci. USA 84, 6452-6456.
6. Jia, S., and Wang, J.L. (1988) J. Biol. Chem. 263, 6009-6011.
7. Gitt, M.A., and Barondes, S.H. (1986) Proc. Natl. Acad. Sci. USA 83, 7603-7607.
8. Southan, C., Aitken, A., Childs, R.A., Abbott, W.M., and Feizi, T. (1987) FEBS Lett. 214, 301-304.
9. Hirabayashi, J., Kawasaki, H., Suzuki, K., and Kasai, K. (1987) J. Biochem. (Tokyo) 101, 775-787.
10. Paroutaud, P., Levi, G., Teichberg, V.I., and Strosberg, A.D. (1987) Proc. Natl. Acad. Sci. USA 84, 6345-6348.
11. Hirabayashi, J., and Kasai, K. (1988) J. Biochem. (Tokyo) 104, 1-4.
12. Clerch, L.B., Whitney, P., Hass, M., Brew, K., Miller, T., Werner, R., and Massaro, D. (1988) Biochemistry 27, 692-699.
13. Wilson, T.J.G., Firth, M.N., Powell, J.T., and Harrison, F.L. (1989) Biochem. J. 261, 847-852.
14. Couraud, P-O., Casentini-Borocz, D., Bringman, T.S., Griffith, J., McGrogan, M., and Nedwin, G.E. (1989) J. Biol. Chem.

264, 1310-1316.

15. Raz, A., Pazerini, G., and Carmi, P. (1989) Cancer Res. 48, 3489-3493.

16. Cherayil, B.J., Weiner, S.J., and Pillai, S. (1989) J. Exp. Med. 170, 1959-1972.

17. Woo, H-J., Shaw, L.M., Messier, J.M., and Mercurio, A.M. (1990) J. Biol. Chem. 265, 7097-7099.

18. Albrandt, K., Orida, N., and Liu, F-T. (1987) Proc. Natl. Acad. Sci. USA 84, 6859-6863.

19. Laing, J.G., Robertson, M.W., Gritzmacher, C.A., Wang, J.L., and Liu, F-T. (1989) J. Biol. Chem. 264, 1907-1910.

20. Gritzmacher, C.A., Robertson, M.W., and Liu, F-T. (1988) J. Immunol. 141, 2801-2806.

21. Feinberg, A.P., and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.

22. Woo, S.L.C. (1979) Methods Enzymol. 68, 389-401.

23. Messing, J. (1983) Methods Enzymol. 101, 20-78.

24. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

25. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

26. Henikoff, S. (1984) Gene (Amst.) 28, 351-362.

27. Southern, E.M. (1975) J. Mol. Biol. 98, 503-508.

28. Kozak, M. (1986) Cell 44, 283-292.

29. Jia, S., Mee, R.P., Morford, G., Agrwal, N., Voss, P.G., Moutsatsos, I.K., and Wang, J.L. (1987) Gene (Amst.) 60, 197-204.

30. Agrwal, N., Wang, J.L., and Voss, P.G. (1989) J. Biol. Chem. 264, 17236-17242.

31. Breathnach, R., and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.

32. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature (London) 263, 211-214.

33. Cordon, J., Wasylyk, B., Buchwalder, A., Sasson-Corsi, P., Kedinger, C., and Chambon, P. (1980) Science 209, 1406-1414.

34. Treisman, R. (1986) Cell 46, 567-574.

35. Gustafson, T.A., Miwa, T., Boxer, L.M., and Kedes, L. (1988) Mol. Cell Biol. 8, 4110-4119.

36. Briggs, M.S., and Gierasch, L.M. (1986) Adv. Protein Chem. 38, 109-180.

37. Spence, A.M., Sheppard, P.C., Davie, J.R., Matuo, Y., Nishi, N., McKeehan, W.L., Dodd, J.G., and Matusik, R.J. (1989). Proc. Natl. Acad. Sci. USA 86, 7843-7847.

CHAPTER V

CONCLUDING REMARKS

When this thesis project was initiated, CBP35 had been purified and characterized in terms of its carbohydrate-binding specificity. There were also new and potentially interesting data pertaining to the proliferation-dependent expression of the protein and its sub-nuclear localization. In quiescent 3T3 fibroblasts, there was a low level of CBP35, predominantly in the cytoplasm; stimulation of the same cells resulted in an increased level of the protein and an apparent translocation into the nucleus. On the basis of these results, molecular cloning of CBP35 was undertaken: (a) to obtain a reliable and abundant source of the material for chemical characterization (e.g. post-translational modification); (b) to develop a probe for studying the expression of the CBP35 gene; and (c) to obtain information on the primary structure of the protein.

The identification of the cDNA clone for CBP35 (Chapter II of the thesis) provided the key stepping stone to the accomplishment of the above three objectives. Recombinant CBP35 has been expressed in E. coli, purified to homogeneity in milligram quantities, and has been particularly important in determining the isoelectric variants of CBP35, which in turn have shed light on the regulation of nuclear versus cytoplasmic localization of the lectin (1). The cDNA clone for CBP35 has also served as an invaluable probe for the analysis of the regulation of the expression of CBP35 gene, in terms of rate of transcription, accumulation of mRNA, and "super induction" in the presence of cycloheximide (2).

The amino acid sequence of CBP35, deduced from the nucleotide sequence of the cDNA clone (Chapter III of the thesis), indicated that the polypeptide could be delineated into two domains: an NH_2-terminal domain that was proline- and glycine-rich and a COOH-terminal CRD. This result provided the first example of a S-type lectin having two domains, a structural motif that had been well-demonstrated in the C-type lectins (3). Thus, the primary structure of CBP35 added an important dimension to the picture of the S-type lectins: that the CRD of a S-type lectin can be coupled to another functional domain (such as has been found for C-type lectins). In the particular case of CBP35, the effector domain showed a highly unusual internal repeat of 9-amino acid residues (Pro-Gly-Ala-Tyr-Pro-Gly-Xxx-Xxx-Xxx), as a result of which there is an uneven distribution of Pro and Gly residues in the CBP35 polypeptide. Both features (internal repeat and uneven distribution of Pro and Gly) have been found in other nuclear proteins. Of particular interest were the available structures of the hnRNP core proteins because immunochemical studies parallel to my structural determination had implicated the associa-tion of CBP35 with nuclear hnRNP (4).

Upon completion of the sequence studies on the cDNA clone, there were discussions on carrying out mutational analysis on CBP35 to address certain questions related to structure, activity, and/or localization. For example, deletional mutation might be carried out to express only the NH_2-terminal domain and to test whether this determines the fate of localization of the polypeptide (cytoplasm, nucleus, RNP complex). Alternatively, the COOH-terminal domain might be expressed to test whether it formed dimers, since all the 14-16 kD group of S-type lectins had been reported to dimerize in non-denaturing buffers while lectins of the 29-35 kD group were

reported to remain monomers.

The focus of my research, however, remained at the structural level, mainly as a result of two developments in the field. First, structural studies, antisera cross-reactivity, and tests of carbohydrate-binding activity on Mac-2, and eBP showed that these other proteins, studied for their possible cell surface function, were identical to CBP35 (5, 6). Second, analysis of cDNA sequences 5′ to the position where our cDNA clone began failed to identify an obvious signal sequence for the endomembrane pathway for secretion. Thus, the question in the field at this juncture was: how to explain the dual (intracellullar and extracellular) localization of CBP35, L34, Mac-2, and eBP, all which are identical (homologous) in coding region sequences. One possibility was the existence of two (or more) alternatively spliced RNAs, one of which has the potential to code for a signal sequence containing polypeptide(6).

The genomic sequence analysis (Chapter IV of the thesis) was undertaken with the hope of shedding light or providing confirmation of these splicing events. The presently available genomic sequence information does not, however, provide sufficient information to either confirm or refute the hypothesis. This must remain as a challenge for future studies.

References

1. Cowles, E. A., Agrwal N., Anderson, R. L., and Wang, J. L. (1990) J. Biol. Chem. in press.

2. Agrwal, N., Wang, J. L., and Voss, P. G. (1989) J. Biol. Chem. **264**, 17236-17242.

3. Drickamer, K. (1988) J. Biol. Chem. **263**, 9557-9560.

4. Laing, J. G., and Wang, J. L. (1988) Biochem. **27**, 5329-5334.

5. Laing, J. G., Robertson, M. W., Gritzmacher, C. A., Wang, J. L., and Liu, F.-T. (1989) J. Biol. Chem. **264**, 1907-1910.

6. Cherayil, B. J., Weiner, S. J., and Pillai, S. (1989) J. Exp. Med. **170**, 1959-1972.