This is to certify that the

thesis entitled

Identification of splicing factors that interact with galectin-3 using yeast two hybrid system

presented by

Jung W. Park

has been accepted towards fulfillment
of the requirements for

M.S. degree in Clinical Lab Science

Ronald Patterson

Major professor

Date _____

IDENTIFICATION OF SPLICING FACTORS THAT INTERACT WITH GALECTIN-3 USING YEAST TWO HYBRID SYSTEM

By

Jung W. Park

A DISSERTATION

Submitted to
Michigan State University
in partial fulfullment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Clinical Laboratory Science

1997

ABSTRACT

IDENTIFICATION OF SPLICING FACTORS THAT INTERACT WITH GALECTIN-3 USING YEAST TWO HYBRID SYSTEM

By

Jung W. Park

Pre-mRNA splicing in mammalian cells involves removal of intervening sequences from newly transcribed precursor mRNAs in two transesterification steps and is an essential step in the formation of most eukaryotic mRNAs. Previously, galectin-3, also known as CBP35, was identified as a required splicing factor. To characterize the mechanism of galectin-3's involvement in splicing in human HeLa cells, the sensitive *in vivo* screening technique called 'yeast two hybrid system' or 'trap system' was used. In this system, a target protein, recombinant murine galectin-3, fused to the Gal4 DNA-binding domain (DNA-BD) traps potential interacting proteins fused to the Gal4 activation domain (AD) expressed from HeLa cDNA library. To date, two different genes representing known human proteins have been isolated and sequenced as potential positive ligands. hDHFR and ribosomal protein L17 are the two proteins that have been identified using the GenBank database. Several verification methods and independent biochemical techniques could either prove or disprove whether the interactions are genuine. Since neither protein is known to be involved in pre-mRNA splicing, further screening using recombinant human galectin-3 might enhance the probability of identifying relevant splicing factors.

Dedication

This is dedicated to my wife, Sung-hee, whose love and support, two beautiful kids, Daniel and Elizabeth, and at last, but not least, my parents, Young and Min for their prayers and supports through all these years.

Acknowledgments

I would like to appreciate my committee members, Dr. J. Gerlach, Dr. L. Mendoza, Dr. S. Conrad for their time, guidance, and effort. I also would like to acknowledge the friendship and valuable discussions I have received from Kwi-suk Kim, Xin-yu Tan, Sue Dagher, Sheldon Leung. Finally, I would deeply thank my mentor, Dr. R. Patterson for his patience, understanding, encouragement, support, and guidance. Without them, it would not been possible.

Table of Contents

List of Figures

Introduction

Galectin Biology

Galectins (1) are mammalian lectins that specifically bind to β-galactoside sugars. These lectins, formerly known as S-type or S-lac lectins, are found in many eukaryotic tissues ranging from lower invertebrates such as sponges to mammals including humans (2). Members of the galectin family share two distinct properties: similar amino acid sequence and specific affinity for ß-galactoside residues (Gal(ß1 --> 4) GlcNAc). These proteins have been implicated in diverse processes such as cell regulation (3), cell adhesion (4), T-cell maturation (5), neoplastic transformation (6), and immune responses (7) (for galectins review, see ref. 8,9). To date, there have been 9 mammalian galectins discovered which are named by consecutive numbering (galectin-1 to galectin-9). Of these, galectin-1 and galectin-3 have been studied most rigorously in different organisms and tissues, leading to multiple nomenclatures for each galectin. For example, galectin-1 has been found in human, bovine, rat, and mouse,

and was previously designated as L-14-I, L14, RL-14.5, galaptin, MGBP, and other names (10). For the purpose of uniform normanclature, a simple consecutive numbering of the galectins was proposed (1).

Interestingly, all members of the galectin family lack both nuclear localization sequences and signal peptides, yet several galectins are found in the nucleus, cytoplasm, and extracellular spaces in various tissue types (11). There is evidence that some galectins are externalized to the cell surface and extracellular matrix via a non-classical secretory pathway (12). Galectin-1 is found remaining in the cytoplasm until it is externalized to cell surface during differentiation in myoblasts (13). Galectin-3 was first identified as a major macrophage cell surface antigen, Mac-2 (14), implicated in protein-carbohydrate interaction with laminin (15). Also, galectin-3 has been identified as a component of ribonucleoprotein (RNP) particles in 3T3 fibroblast nuclei (16) even though galectin-3 does not contain a nuclear localization sequence. Metastasis of malignant cells was also reported to be related to the presence of galectins on the cell surface (17). However, the mechanism for various localizations of galectins is not understood at this time.

Galectins can be classified into three subgroups according to their amino acid sequence homology and polypeptide structure. The first group is the prototype and includes galectin-1 and galectin-2, which are

homodimers, with subunit molecular masses of ~14 kDa (about 130 amino acids)(figure 1). Galectin-1 and galectin-2 share ~40% sequence homology with the most similarity in the series of ß-sheets which contain the carbohydrate recognition domain. Galectin-5 and galectin-7 are other prototype galectins that have been identified. Galectin-1 is found in the cytoplasm as well as extracellular matrix (18) and has been shown to interact with laminin through protein-carbohydrate interaction between the carbohydrate recognition domain (CRD) of galectin-1 and the carbohydrate residues of laminin. Although the precise functions of the prototype galectins are not yet fully established, there is evidence for their involvement in regulation of cell growth (19), cell adhesion (20), embryogenesis (21) and apoptosis (22).

The members of the second group (tandem repeat) include galectin-4, galectin-6, galectin-8 and galectin-9. They are monomers with a molecular mass of ~35 - 36 kDa composed of two highly similar tandem repeats of 140 amino acids each (figure 1). Galectin-4 and galectin-8 share ~38% identity within their two homologous carbohydrate recogniation domains linked by a short ~30 amino acid linker peptide (23). The functional significance of the redundancy of two homologous CRDs is still under investigation. Although the biological role of tandem repeat galectins is not known at this time, they presumably are involved in similar processes as galectin-1 or

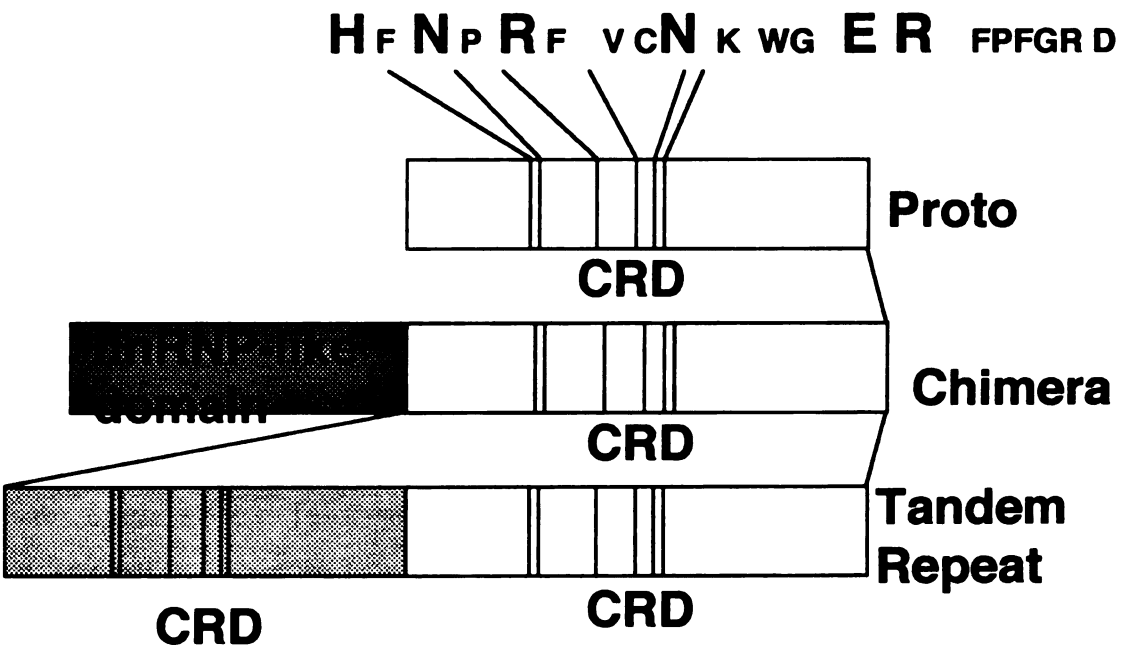
galectin-2 considering their similar amino acid sequence, localization, and regulated level of expression (24).

The sole example of the final group (chimera) is galectin-3 (figure 1). Galectin-3 is a monomeric ~29 -35 kDa protein, composed of two distinctive domains: an N-terminal half that i) is proline and glycine rich with pro-gly-ala-tyr-pro-gly-X-X-X repeats characteristic of the collagen gene superfamily, and ii) has limited homology to a heterogeneous nuclear ribonucleoprotein (hnRNP), and a C-terminus homologous to the carbohydrate recognition domain of the other galectins. Galectin-3, previously known as CBP35, Mac-2, IgE-binding protein, CBP30, L-29, LR-29, and L-34 according to its localization in different tissue types or its functional properties, is reported to bind laminin (25), be a modulator of cell-cell and cell-matrix interaction (26), be a modulator of immune responses (27), bind to LPS (28), inhibit apoptosis (29), and cause metastasis (30). However, it is not clear how one protein can be involved in so many different cellular functions.

Galectin-3 has been a focus of our research due to its nuclear localization. Treatment of permeabilized mouse 3T3 fibroblasts and human HeLa cells with ribonuclease A released galectin-3, whereas treatment with DNase I did not. In addition, mouse 3T3 galectin-3 fractionated with hnRNP and snRNP complexes in Cs_2SO_4 gradients with a

Figure 1. Schematic representation of the three types of galectins (proto-, chimera- and tandem-repeat). The proto-type is composed of a single lectin domain. The chimera-type is composed of two parts; a C-terminal half containing the carbohydrate recognition domain (CRD) and an N-terminal half of unknown function. The tandem-repeat-type is composed two CRD domains. Conservative residues among galectins are shown in single letters above the proto-type structure. Large capital letters denote essential residues for carbohydrate binding.

Figure 1
Three Types of Galectins (proto-, chimera-, and tandem-repeat)



buoyant density of between 1.28 and 1.32 g/ml, which is a distinguishing characteristic of RNA-protein complexes (31). In quiescent mouse 3T3 cells, galectin-3 is phosphorylated and localized mainly in the cytoplasm. However, in proliferating cells, this lectin is predominantly in the nucleus in both phosphorylated and unphosphorylated forms (32). Therefore, these data as well as other lines of evidence imply the functional significance of nuclear localization of galectin-3. RNP complexes are known to interact with precursor mRNAs in RNA splicing. Although galectin-3's precise role in the nucleus has to be established, its involvement in pre-mRNA splicing is suggested and this may be important for the regulation of cell growth.

Pre-mRNA splicing

Splicing, removal of intervening sequences from precursor mRNA (pre-mRNA), is one of the essential post-transcriptional processes of most eukaryotic mRNAs. Understanding of the biochemistry involved in splicing has been successful just over 10 years due to the development of efficient and faithful *in vitro* splicing systems. The mechanism of pre-mRNA splicing has been elucidated to two separate transesterification reactions of the pre-mRNA substrate. Step 1 involves a nucleophilic attack on the phosphodiester bond at the 5' intron-exon junction from a 2' ribose hydroxyl group of an adenosine residue at the branch point in the intron. This generates reaction

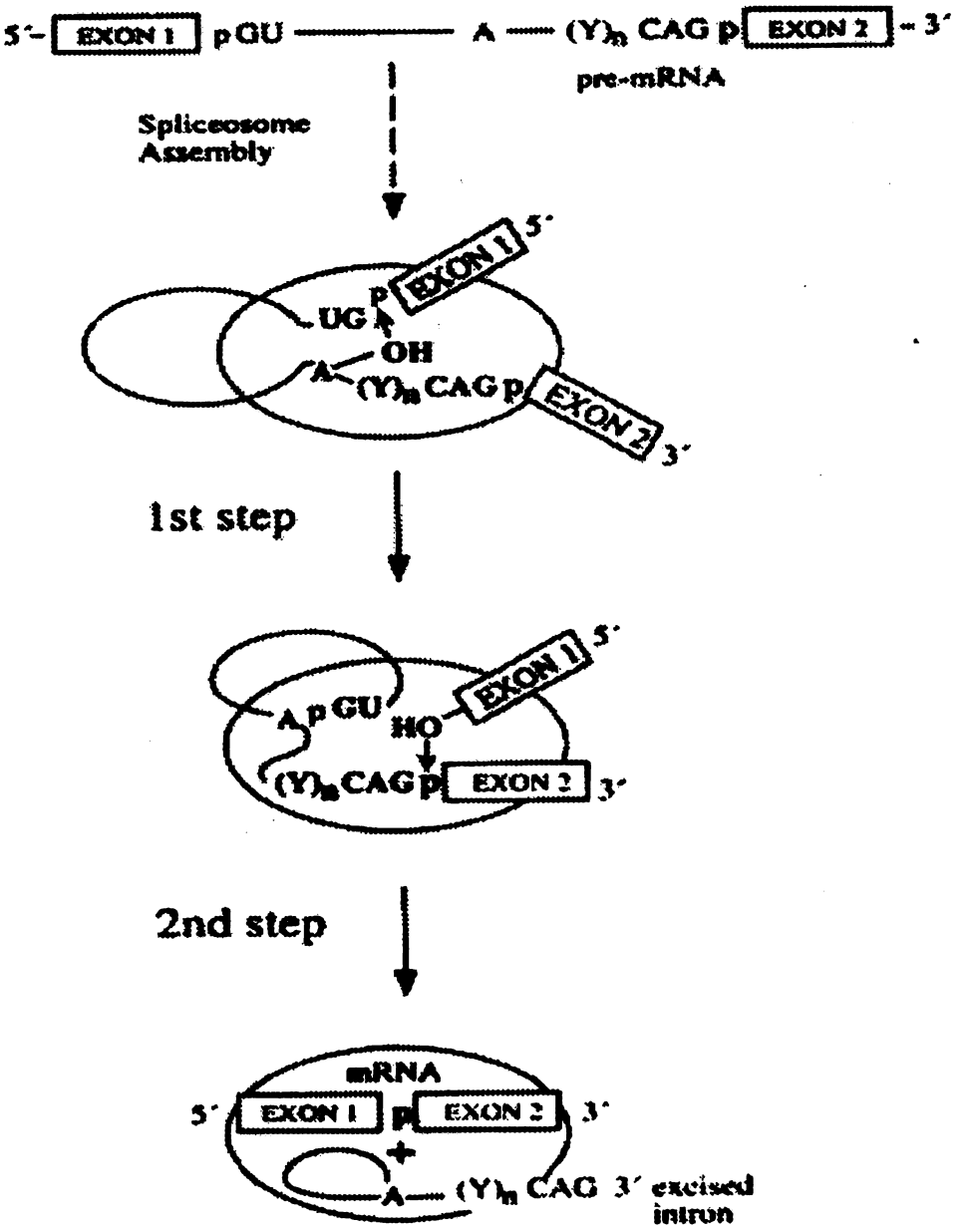
intermediates corresponding to a free 5′ exon and a lariat attached to the 3′ exon. In step 2, a second nucleophilic attack occurs from free 3′ ribose hydroxyl group of the 5′ exon on the phosphodiester bond at the 3′ intron-exon junction. This results in the formation of products corresponding to ligated exons (mRNA) and fully excised intron lariat (see ref. 33a,33b for splicing review) (figure 2).

Despite the fact that splicing can be achieved by two simple chemical reactions, the splicing machinery requires locating cleavage sites, forming a lariat structure, and forming ligated products accurately. These steps involve at least 5 small nuclear RNAs (U1, U2, U4, U5, U6 snRNAs), a number of other proteins and complicated mechanisms. The complex macromolecular structures performing these reactions are termed spliceosomes. Genetic studies involving *in vitro* mutagenesis have shown the base-pairing interactions between snRNA and pre-mRNA (33c,33d). These interactions between snRNA and pre-mRNA have been also identified by UV cross-linking at specific stages of spliceosome assembly (34a,34b).

The mechanistic similarities between splicing of nuclear pre-mRNA involving spliceosomes and group II introns, which are self-catalytic, support the idea that RNA-RNA interactions play a central role, and maybe even catalyze the splicing reactions (35). Although RNAs might

Figure 2. The mechanism of spliceosome mediated pre-mRNA splicing. Exon sequences are enclosed in boxes. The phosphate at the 3′ terminus of the intron, shown in bold, forms the phosphodiester bond that links the spliced exons together in the mRNA product. The two catalytic transesterifications take place following a complex pathway of spliceosome assembly, indicated by a dotted arrow.

Figure 2
A Schematic Diagram of Splicing Involving Two Transesterification Reactins



be directly involved in the catalysis of pre-mRNA splicing, many known snRNP proteins and numerous other non-snRNP proteins (approximately 70 - 80) might play critical roles in positioning, stabilizing, and disrupting RNA-RNA interactions in the spliceosome (36). Only a few of the splicing factors have been identified and characterized (see ref. 37 for review).

Identification of Galectin-3 as a Required Splicing Factor

In a recent report, galectin-3 was identified as a required splicing factor in HeLa cell nuclear extracts (38). It was demonstrated that saccharide ligands (lactose) with high affinity to galectin-3 inhibit the *in vitro* splicing reaction. Moreover, nuclear extracts depleted of galectins by lactose-sepharose affinity chromatography failed to support splicing, but splicing activity was restored upon reconstitution with mouse recombinant galectin-3 purified from an *E. coli* expression system. In parallel studies, the formation of spliceosome complexes was analyzed by native gel electrophoresis which allows isolation of native splicing complexes. Spliceosomes are multicomponent complexes containing precursor mRNA, splicing intermediates, snRNPs (U1, U2, etc) and many non-snRNP splicing co-factors (38a). This electrophoretic separation of complexes involved in pre-mRNA splicing produces at least 4 bands: H complexes, pre-commitment, commitment complexes (E), and active spliceosomes (A,B,C).

Nuclear extracts depleted of galectins by lactose-sephrose column chromatography form only H-complexes suggesting that galectin-3 is required for the formation of active spliceosomes. The addition of recombinant galectin-3 to the depleted extract resulted in the formation of active spliceosomes as well as restoration of splicing activity.

In contrast, the removal of galectin-3 by immunodepletion using monoclonal anti-galectin-3 antibody did not affect splicing activity. This unexpected finding appears to be related to the redundancy of galectins in the nucleus. It is known that HeLa nuclear extracts contain at least one more galectin (galectin-1) other than galectin-3. Passage of HeLa nuclear extracts over lactose-sepharose affinity column or adding galectin-specific saccharide ligands to nuclear extracts abolished the splicing activity by removing either physically or functionally all galectins from the extract. Monoclonal anti-galectin-3 antibody, however, specifically depletes galectin-3 while galectin-1 remains physically and functionally intact in the splicing extract. This finding, therefore, suggests that other galectins in HeLa nuclear extracts, specifically galectin-1, not removed by monoclonal anti-Mac-2, substitute for galectin-3 during splicing.

How galectins participates in the splicing pathway has not been established. Identifying a splicing factor that interacts with nuclear galectins could provide a major stepping stone toward understanding the mechanism and role

of the galectins during splicing. One of the difficulties searching for galectin-3 ligand(s) that results from the fact that only basal levels of galectin-3 are necessary for splicing to occur. Therefore, the detection of galectin-3 in active spliceosome complexes becomes difficult even with very sensitive detection methods such as western blotting. These data might indicate i) the amount of galectin-3 bound to the splicing complexes is too low to be detected by biochemical techniques available currently, ii) the interaction might be too transient to detect galectin-3 in spliceosomes, iii) galectin-3 interacts with protein(s) which is (are) not a component of spliceosomes, iv) or galectin-3 interacts with proteins that might be constituents of spliceosomes, but these proteins interact either before ligands join the splicing complexes or after ligands are released from the splicing complexes. Finally, if the interactions are too weak, galectin-3 could be separated from the complexes during native gel electrophoresis. The identification of *bona fide* splicing factors as ligand(s) for galectin-3 would strengthen the model that galectin-3 associates with splicing complexes and participates in the splicing pathway *in vivo*. As described below, the yeast two hybrid system is an *in vivo* approach used to identify ligands for a specific protein in which weak and/or transient interactions occur. This system was used to identify spliceosomal components that serve as ligand(s) for the nuclear galectins.

Yeast Two Hybrid System: A Screening Technique for Protein-Protein interaction *in vivo*

The yeast two hybrid approach takes advantage of the modular domain structure of eukaryotic transcription factors. Many eukaryotic transcription activators have at least two distinct functional domains, one that directs binding to specific DNA sequences and one that activates transcription (39). This modular structure is best illustrated by yeast experiments showing that the DNA-binding domains and activation domains can be exchanged from one transcription factor to the next and retain function. For example, when the DNA-binding domain of the yeast transcription factor Gal4 is replaced with the DNA binding domain of the bacterial repressor LexA, the resulting hybrid protein activates transcription of genes containing upstream LexA binding sites (40). Similarly, when the DNA binding domain of Gal4, which by itself does not activate transcription, is fused to activation domains from other proteins the resulting hybrid proteins activate transcription of reporters with upstream Gal4 binding sites (41).

A crucial inference of the modular nature of transcription activators is that the DNA-binding and activation domains need not be covalently attached to each other for activation to occur. This was first demonstrated by Ma and Ptashne (42) with a Gal4 derivative that contained the DNA-binding domain as well as a domain that

interacts with another yeast protein, Gal80, but that lacked the activation domain. When this derivative was expressed in yeast it did not activate transcription of a reporter gene containing upstream Gal4 binding sites. However, when it was co-expressed with a second, hybrid protein, consisting of Gal80 fused to an activation domain, interaction between the Gal4 DNA-binding derivative and the Gal80-activation domain hybrid resulted in activation of the reporter gene.

The general utility of the modularity of transcription factors was demonstrated by Fields and Song (43) who showed that yeast transcription could be used to assay the interaction between two proteins if one of them was fused to a DNA-binding domain and the other was fused to an activation domain. In their experiment, one of the hybrid proteins contained the DNA-binding domain of Gal4 fused to the yeast protein Snf1, and the other contained the activation domain of Gal4 fused to another yeast protein, Snf4. When Snf1 and Snf4 interacted they brought together the DNA-binding and activation domains, so that the two hybrid proteins bound to Gal4 binding sites upstream of a lacZ reporter gene and activated its transcription. Thus, the interaction between Snf1 and Snf4 was assayed by production of β-galactosidase.

The interaction trap or matchmaker is an extension of the two hybrid system (44). It consists of three critical components. First, it uses a vector for expression of a

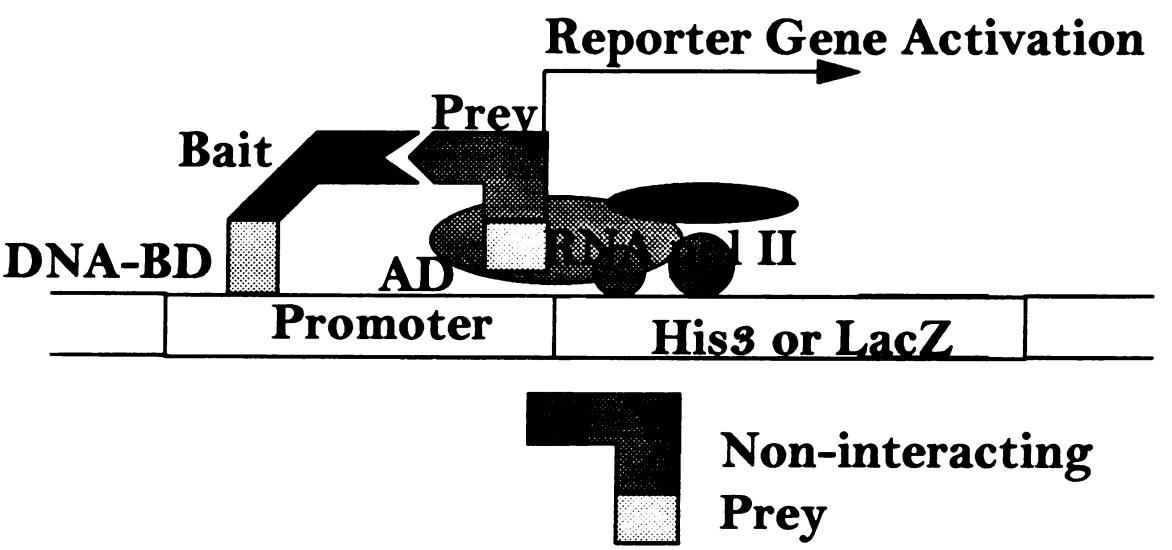
protein of interest fused to the Gal4 DNA binding domain (DNA-BD). Because the goal of the interaction trap or trap system is to find proteins that interact with the protein fused to Gal4 DNA-BD, this hybrid is referred to as 'bait'. Second, the trap uses a yeast strain with two reporter genes. One reporter is a yeast His3 that has its normal upstream regulatory sequences replaced with Gal4 responsive elements. Transcription of His3 gene under the control of the Gal4 responsive elements can be measured by the ability of the strain to grow in the absence of Histidine. The Gal4 responsive element-His3 gene construct is integrated into the yeast chromosome. The other reporter gene is LacZ, which is controlled by a different promoter and provides a secondary assay of activation by the bait and activation-tagged proteins interacting with it. The LacZ reporter also provides some quantitative information about the interaction. Third, the interaction trap uses a plasmid that directs the expression of cDNA library-encoded proteins fused at their amino termini to a transcriptional activator domain (AD).

When the Gal4 DNA-BD and Gal4 AD are separately expressed in yeast, Gal4 transcriptional activator is not active. However, when the two hybrid genes containing two potential interactors are introduced into a yeast host strain, interaction between two proteins contained in each fusion protein will bring the Gal4 DNA-binding domain and the Gal4 transcriptional activation domain in proximity,

resulting in the restoration of Gal4 transcriptional activity. Functionally active Gal4 transcriptional activator would then interact with Gal4 responsive elements, and in turn, induce the expression of the reporter genes, i.e. His3 and LacZ (figure 3).

Figure 3. A schematic representation of the yeast two hybrid system. When two fusion proteins (bait fused to Gal4 DNA binding domain and a potential interactor for bait fused to Gal4 activation domain) are expressed in the nucleus of the host cell, the Gal4 transcriptional activator is functionally restored. Interaction between the bait and ligand (shown as a match between the bait and its prey) is required to activate the reporter gene. Non-interacting prey shown below will not interact with the bait, thus will not activate the reporter gene.

Figure 3
Schematic Diagram of Yeast Two Hybrid System



Materials and methods

Expression Plasmids

pAS2-1 is used to generate a fusion of the Gal4 DNA-BD (amino acids 1-147) and a protein of interest cloned into the multiple cloning sites (MCS) in the correct orientation and reading frame. pAS2-1 is derived from $pAS1^{CYH2}$ and carries the wild type $CYH^{S}2$, which confers cyclohexamide sensitivity in transformed cells. pAS2-1 contains a unique EcoRI site in the MCS and a neutral, short peptide to completely eliminate the autonomous activation activity of pAS2. The unique restriction sites in the MCS are PstI, SalI, BamHI, SmaI/XmaI, EcorI, SfiI, NcoI, and NdeI. pyBG3 (pAS2-1 / murine galectin-3) and pyBG1 (pAS2-1 / human galectin-1) were generated by ligating the coding region for galectin-3 (an EcoRI fragment) and galectin-1 (a BamHI fragment) into pAS2-1 purchased from Clonetech. The inserted direction of orientation was determined by restriction analysis with HaeIII and SacI, respectively. cDNA library expression vector, pGAD GH containing Gal4 activation domain fused with human HeLa cDNAs as EcoRI fragments in MCS, Leu2 for growth selection, and SV40 large T antigen nuclear localization signal just upstream of Gal4 AD was purchased from Clonetech.

Western Blotting

The expression of Gal4 DNA-BD-galectin-3 fusion protein in yeast was detected by western blotting using monoclonal anti-mouse Mac-2 antibody. The yeast cell lysate

was prepared by boiling and vortexing with 0.5 μm acid-washed glass beads in 1X sample buffer containing final concentration of 2 % SDS, 8 M urea to protect the lysates from proteolytic degradation. Harvested cells at an OD_{600} of 0.2-0.4 are mixed with 1X sample buffer and acid-washed beads, boiled for 2 minutes, and vortexed for 10-15 seconds (5X). The cell lysates expressing the galectin-3 fusion protein were separated on 12.5 % SDS-polyacrylamide gel along with the cell lysates containing no plasmid. The separated proteins were transferred to polyvinylidene difluoride (PVDF) membranes (Bio-Rad) in a transfer buffer (25 mM Tris / 193 mM glycine / 20 % methanol). After the protein transfer, the membranes were blocked with 5 % nonfat dry milk dissolved in phosphate-buffered saline (PBS; 10 mM sodium phosphate / 150 mM NaCl, pH 7.2) containing 0.05 % Tween 20 (PBS-T).

After the blocking, the membranes were incubated with the rat monoclonal anti-Mac-2 (anti-galectin-3) antibody (freshly diluted 1 : 5000 in 5 % nonfat dry milk in PBS-T) at room temperature for 2 hours, and washed five times in PBS-T for 15 minutes each. The membranes were incubated with horseradish peroxidase-conjugated goat anti-rat antibodies (Pierce) diluted 1 : 3000 in PBS-T for 30 minutes at room temperature and washed five times for 15 minutes each in PBS-T. The membranes were exposed to X-ray film after incubating the membranes with the enhanced

chemiluminescent reagents from Amersham according to the manufacturer's protocol.

Yeast Strains

Y190 was derived from Y153 by plating on YPD containing 10 mg/ml cycloheximde medium and by selecting for cycloheximide resistance (45). Y190 is (leu2-3, 112, ura3-52, trp1-901, his3-Δ200, ade2-101, gal4Δgal80Δ URA3GAL-lacZ, LYS GAL-HIS3, cyh^R).

Library Transformation

Y190 containing pAS2-1-galectin-3 was sequentially transformed with pGAD GH containing a human HeLa cDNA library as previously described (46). The transformation efficiency of both bait fusion and cDNA library fusion plasmids was determined by counting the number of colonies grew on SD (yeast nitrogen base without amino acid) medium lacking tryptophane and leucine. Transformants were selected on SD / - Trp / -Leu / -His / 25 mM 3-AT. Selected transformants were subsequently tested for β-galactosidase production by colony lift filter assay as described previously (47). Clones which grew on SD/-Trp/-Leu/-His and produced a positive β-galactosidase assay were considered positive and recovered into *E. coli* by the smash and grab method as described (48). Loss of the pAS2-1-galectin-3 plasmid was selected by streaking on SD lacking leucine to isolate clones containing only the cDNA plasmid.

Colony-lift Filter Assay

Fresh colonies grown to 1-3 mm in diameter are transfered to a sterile dry filter (whatman #5) by placing a filter onto the plate and then lifting it carefully with forceps. The filter with colonies is transfered into a pool of liquid nitrogen, and submerged completely for 10 seconds. The frozen filter is thawed at room temperature momentarily and then placed, colony side up, on top of another filter paper pre-soaked with Z buffer containing X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside). Filters are incubated at 30°C or room termperature. The colonies turning blue color within 8 hrs were considered potential positives and used for further screening.

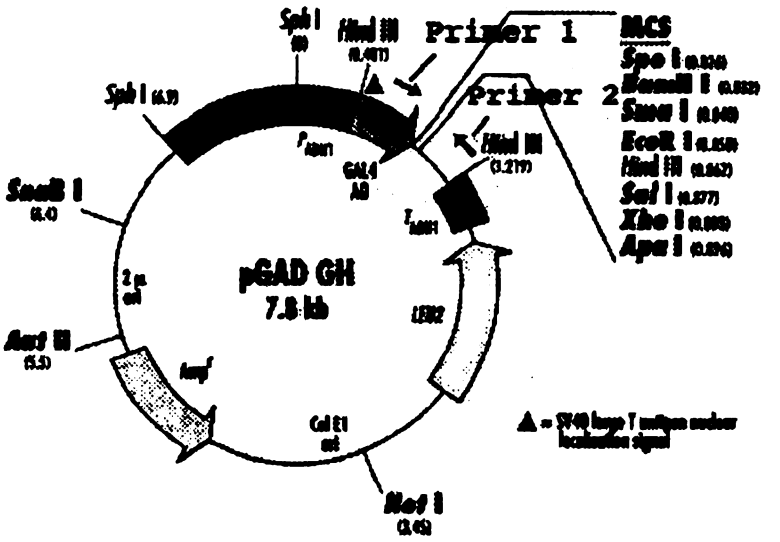
Identification of Positive Clones

The amplification of the cDNA library inserts in library-encoding plasmids from potential positive clones was as follows; for each candidate clone, a single colony was inoculated in 5 ml SD/-Trp/-Leu/-His/+25 mM 3-AT culture and grown until the culture is saturated (20 hr with shaking at 250 rpm). Yeast DNA extract, including plasmids, was prepared by vortexing with 0.2 ml of lysis solution (2 % Triton X-100, 1 % SDS, 100 mM NaCl, 10 mM Tris (pH8.0), 1.0 mM EDTA), 0.2 ml of phenol : chloroform : isoamylalcohol (25 : 24 : 1) and 0.3 g of 500 μm acid-washed glass beads. The cDNA library inserts were amplified by PCR using primers 5'-TACCACTACAATGGATG-3' and

5'-TAGCATCTGTGACTTTTTGG-3' (figure 4). Forward primer (5' to 3') was targeted to anneal to the vector sequence (pGAD GH) about 200 bps upstream from the first EcoRI site where cDNA library was cloned. Reverse primer (3' to 5') was targeted in the multiple cloning region flanking the second EcoRI site. PCR products were separated on 1.0% agarose gels and their sizes were determined by comparison to λ DNA cut with HindIII. PCR products were purified using Promega Wizard PCR purification kit according to manufacturer's protocols and sequenced by dideoxy termination method by the MSU sequencing facility.

Figure 4. A diagram showing the library expression vector, pGAD GH, with primer sites indicated. Elsewhere, activation domain, nuclear localization sequences are denoted as AD and NLS respectively.

Figure 4
A diagram showing the library expression vector, pGAD GH



Results

Phenotype Verification of Reporter Yeast Strains

Two phenotypically similar yeast strains (Y190 and CG1945) are available for the yeast two hybrid system from Clonetech. I chose Y190 over CG1945 for the following reasons. First, Y190 exhibits higher LacZ promoter strength than CG1945. Thus, Y190 expresses a higher level of β-galactosidase activity than CG1945. Second, CG1945 exhibits a clumping phenomenon which results in lower transformation efficiency, less reproducibility, and other technical difficulties. The phenotype of Y190 (leu2-3, 112, ura3-52, trp1-901, his3-D200, ade2-101, gal4Δ, gal80Δ, URA3 :: $GAL1_{UAS}$-$GAL1_{TATA}$-lacZ, LYS :: $GAL1_{UAS}$-$HIS3_{TATA}$-HIS3, cyh^R) was verified before the transformation with 'bait' plasmid. Y190 is deficient for trp and leu and can not grow on minimal medium (Synthetic Dropout Medium (SD)) lacking those nutrients unless functional TRP1 and LEU2 genes are introduced (figure 5). However, Y190 is leaky for HIS3 expression for unknown reasons, and therefore, the presence of 3-amino-1,2,4-triazole (3-AT) is required to suppress the background growth on His- medium. A titration experiment was performed to determine the lowest amount of 3-AT required to achieve the minimum number of colonies growing on SD/-His medium. 25 mM 3-AT was necessary to

suppress the background growth of Y190 (figure 6). It is worth noting that Y190 transformed with the DNA-BD / bait hybrid plasmid alone produced positive results for β-galactosidase assay. This autonomous activation of the LacZ gene implies that the bait fusion protein also might activate His3 which has an identical promoter. Apparently, the presence of 25 mM 3-AT was sufficient to suppress the growth from autonomous activation of His3 gene by the bait protein as well as the background growth from leaky His3 production.

Construction of Bait Hybrid Plasmids

To make a plasmid that directs the synthesis of the Gal4 fusion or 'bait' protein, the coding region for recombinant murine galectin-3 was inserted into pAS2-1 as described in materials and methods (designated as pyBG3; yeast Bait Galectin-3). pAS2-1 is a multicopy yeast plasmid containing the yeast 2 μ origin of replication and the selectable marker gene Trp1, as well as Gal4 DNA-BD flanked by the yeast ADH1 promoter and terminator. Since subcloning was not directionally performed using two different restriction sites, the orientation of the insert had to be determined. In this subcloning, galectin-3 cDNA was inserted in the correct reading frame so that recombinant galectin-3 was correctly translated. Bait proteins expressed from this plasmid contain amino acids 1 to 147 of Gal4 transcription activator, which include the

Figure 5. Phenotypic verification of the yeast reporter strains. CG-1945 and Y190 were grown on minimal synthetic dropout medium lacking different amino acids. Both strains grew on SD / -His medium due to leaky expression of His3 gene.

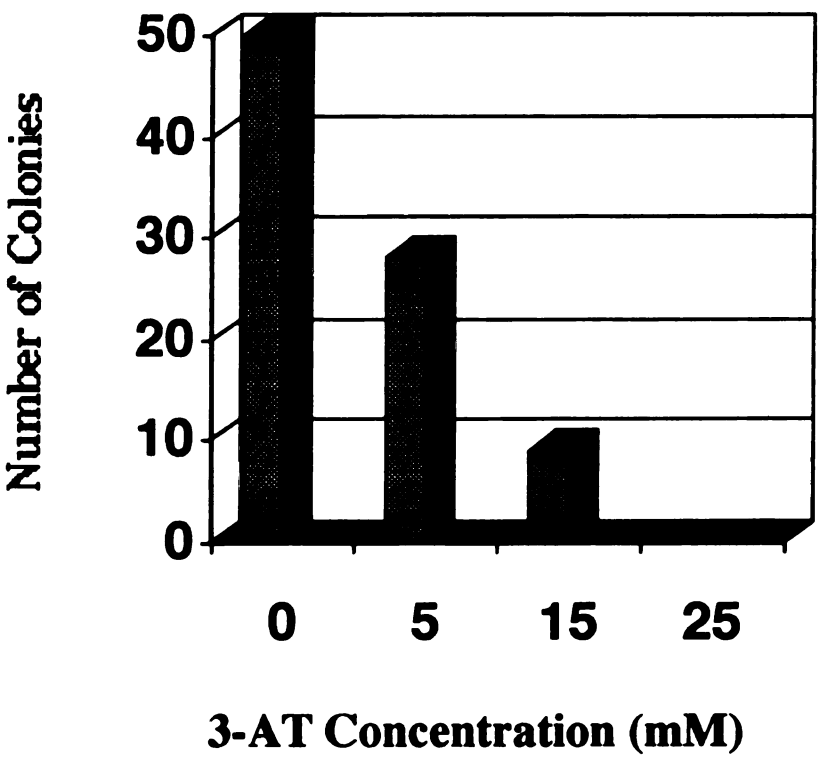
Figure 5
Growth of The Yeast Host Strains on Various Types of Media

	SD/-Trp	SD/-Leu	SD/-HisSD/-Ura	YPD
CG1945	-	-	- [a]	+	+
Y190	-	-	- [b]	+	+

[a] In the presence of 5 mM 3-AT
[b] In the presence of 25mM 3-AT

Figure 6. Determination of 3-AT concentration required to suppress background growth of Y190 with pyBG3 on SD /-Trp /-His. Y190 containing pyBG3 was grown on plates containing 0, 5, 15, 25 mM 3-AT, and the number of colonies on each plate was determined.

Figure 6
Determination of 3-AT Concentration To Suppress The Background Growth of Y190 With pYBG3 on SD/ -Trp/-His



DNA-binding domain. Transformants will constitutively express the protein of interest with Gal4 DNA-BD at its amino terminus. Since it contains a yeast nuclear localization signal, the bait protein will enter the nucleus. After strain Y190 was transformed with pyBG3, the transformants were grown on medium lacking tryptophan to maintain plasmid pyBG3. The expression levels afforded by the ADH1 promoter are generally sufficient to provide occupancy of Gal4 responsive elements upstream of the reporter genes.

Recombinant human galectin-1 cDNA was also subcloned into the pAS2-1 expression vector and the fusion plasmid was named pyBG1 (data not shown). The Gal4 DNA-BD / galectin-1 fusion protein needs to be tested for its stable expression in Y190 transformed with pyBG1. Polyclonal anti-galectin-1 antibody raised from rabbit will be used to detect the galectin-1 from Y190 cell lysates.

DNA-BD /galectin-3 fusion protein was stably expressed

The DNA-BD / galectin-3 (pyBG3) construct was transformed into strain Y190 using the small-scale yeast transformation protocol. Transformants selected on SD / -Trp were prepared for western blotting using a monoclonal anti-galectin-3 antibody(anti-Mac2). The expected fusion protein with molecular weight of 51 kDa was detected while no protein was detected from strain Y190 containing no expression plasmid (figure 7). The overexpression of this fusion protein in yeast cells did not produce toxic

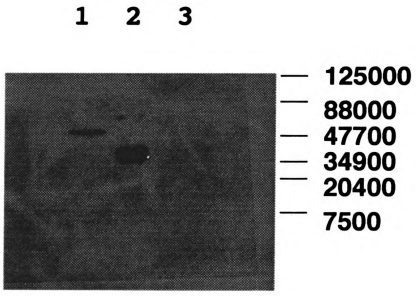
effects. The growth curve of cells transformed with the Gal4 DNA-BD / galectin-3 was compared to that of cells transformed with the Gal4 DNA-BD vector alone. A doubling time of ~3 hrs were observed in both cases during logarithmic phase of growth (data now shown).

Galectin-3 fused to Gal4 DNA-BD exhibits cryptic transcriptional activation

To test whether the galectin-3 fusion protein autonomously activates reporter genes without interacting with its ligand(s), Y190 transformed with pyBG3 was first plated on minimal medium lacking tryptophan and histidine without 3-AT. As expected, pyBG3 transformants grew normally on SD / -Trp / -His due to the background production of histidine. The autonomous transcriptional activation of His3 gene by galectin-3 fusion protein was not in suspect at the time. When pyBG3 transformants were, however, subjected to β-galactosidase assay using a simple colony-lift filter technique, all colonies unexpectedly turned blue indicating the autonomous transcriptional activation by Gal4 DNA-BD / galectin-3. For faster screening (simple blue/white) and extremely sensitive detection of β-galactosidase, a colony-lift filter assay was preferably used to screen the large sized library. In this assay, X-gal is used as a substrate for β-

Figure 7. Western analysis of Gal4 DNA binding domain / galectin-3 fusion protein expressed in Y190. Y190 cell lysates prepared from 5 ml cultures were subjected to SDS-PAGE using 12.5 % acrylamide, and electrotransferred to PVDF membrane. The fusion protein was detected using anti-Mac-2 (anti-galectin-3), horseradish peroxidase-conjugated goat anti-mouse IgG, and chemiluminescent substrate. Lane: 1, cell lysates prepared from Y190 containing pyBG3; 2, recombinant galectin-3 (~ 1 μg); 3, cell lysates prepared from Y190 containing no plasmid.

Figure 7
Western Blotting of Y190 Cell Extract Against Anti-Galectin-3



galactosidase and the time taken to turn blue is described as a function of the LacZ expression level in inversely proportional fashion.

Screening of HeLa cDNA libraries fused to Gal4 AD for the interaction with galectin-3 fused to Gal4 DNA-BD resulted in unexpectedly large number of positive clones

The DNA-BD / galectin-3 and GAL4 AD / HeLa cDNA library hybrid plasmids were transformed into strain Y190 sequentially. The two-hybrid cDNA libraries are constructed into the AD vector rather than the DNA-BD vector because fusing random proteins to a DNA-BD would produce a much larger percentage of hybrids that might function as transcriptional activators on their own, than if the library-encoded proteins are fused to an AD.

The autonomous transcriptional activation of the His reporter gene was completely suppressed in the presence of 25 mM 3-AT. However, LacZ activation could not be suppressed by this simple method. In many cases in which such autonomous activation was observed, especially using transcription factors as bait, these types of baits are avoided. Otherwise, the deletion of activating region from the target protein without affecting the overall structure could salvage the problem of autonomous activation. In an attempt to resolve this problem for further screening, a semi-quantitative assay was designed to delineate potential true LacZ positives. First, the time taken for cells transformed with pyBG3 to turn blue was monitored at

different concentrations of X-gal and compared to the time taken for cells transformed with pCL1 which expresses full length Gal4 transcriptional activator (figure 8). On a plate containing 100 patches of different His positive clones, a clone transformed with pyBG3 alone and a clone transformed with pCL1 were grown together with the rest of clones, then subjected to the colony-lift filter assay. The time taken to turn blue for each colony was monitored and compared to the time taken for the clone transformed with pyBG3. Clones turning blue faster than the colony transformed with pyBG3 were considered potential positives and each positive was arbitrarily classified 1,2,or 3 depending on the kinetics of turning blue (1 - slowest to 3 - fastest).

A total of 2.8 X 10^7 clones from a library scale transformation were screened(figure 9). Robust colonies that grew larger than 1 mm after 1 week on SD / -Trp / -Leu / -His / +25 mM 3-AT were considered potential positives and subjected to the β-galactosidase filter lift assay and further screening. Out of 2.8 X 10^7 transformants, ~2700 (0.0083 %) were larger than 1 mm after 1 week in the presence of 25 mM 3-AT. The number of LacZ+ positives was estimated from the total number of LacZ+ positives out of 300 His+ positives (180 colonies). Each LacZ+ colony was arbitrarily designated as 1-3 according to blueness as a function of time, which in turn can be indirectly

Figure 8. Semi-quantitative assay for β-galactosidase expression. Expression levels of β-galactosidase for Y190 containing pCL1 which expresses wild type Gal4 transcriptional activator and Y190 containing pyBG3 were compared in the presence of different amounts of X-gal. X-axis is defined as the X-gal amount (in μg) applied for colony-lift filter assay. Y-axis is defined as the time (in minutes) required for more than 90 % of Y190 colonies to turn blue in the same assay. Y190 containing pyBG3 grown in the presence of 20 μg X-gal did not turn blue after a prolonged incubation (more than 8 hours).

Figure 8
Semi-quantitative Assay for β - galactosidase Activity

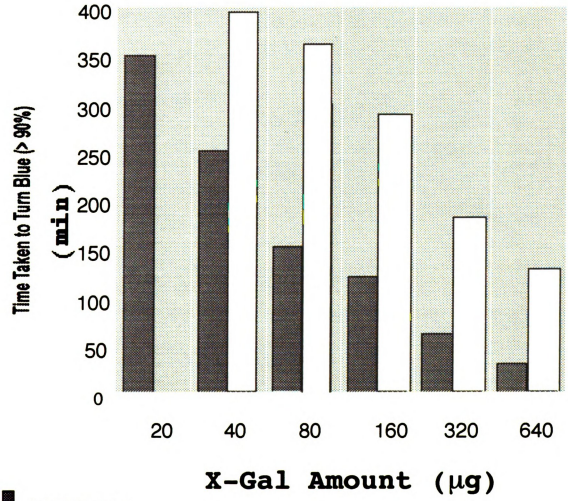


Figure 9. Summary of data from a cDNA library screen using the yeast two hybrid system.

Figure 9
Summary of Yeast Two Hybrid System

Transformation Efficiency	**1 . 111 X 10^5 cfu / mg**	**------**
Total Number of Clones Screened	**2 . 8 X 10^7 cfu**	**------**
Number of His+	**~ 2700**	**0.0083%**
Number of His+ /LacZ+	**~ 1620**	**0.0049%**
Minimum Number of Positive Clones to be Sequenced	**57**	

Estimated Number of Protein Messages Per Cell : 30,000 - 40,000

interpreted as an expression level of β-galactosidase. The arbitrary designation of each colony was a helpful indicator to predict the strength of interaction between galectin-3 and its ligand. The theoretical number of His+/ LacZ+ clones to be isolated and identified as potential positives while covering the whole library (1 X 10^6) was estimated from the total number of clones screened.

Isolation and identififcation of two potential positive clones

PCR amplification of DNA isolated from positive clones in the library screening using two primers described in materials and methods resulted in the PCR products that correspond to the potential interacting clones. 25 PCR products out of 50 appeared to have an apparent size of ~500 bps, one PCR product with ~3000 bps, one PCR product with ~1300 bps, and two PCR products with less than 200 bps. The rest of the clones (21 clones) did not produce PCR products. Interestingly, PCR amplification of a few clones produced multiple fragments in nearly equal amounts (figure 10).

Six individual clones with PCR product sizes of 500 bps (five) and 1300 bps (one) have been sequenced and their sequences were identified using the Genebank database. Clones containing the 500 bp insert were identified as the large subunit ribosomal protein L17, and the clone with the 1300 bp insert was identified as an exon6 sequence of human

Figure 10. PCR analysis of positive clones. A positive control denoted as C or + is the PCR product of the multiple cloning site flanked by about 100 base pairs downstream of the MCS in the Gal4 activation domain expression vector, pGAD GH. A negative control is the PCR performed with Y190 cells containing no plasmid. Sizes of standard marker fragments in base pairs are shown at side of figures. Also, the number above in 10a/b and also at bottom in 10c indicate the positive clone designation.

Figure 10a
PCR Products of DNA Extracts from Postive Clones

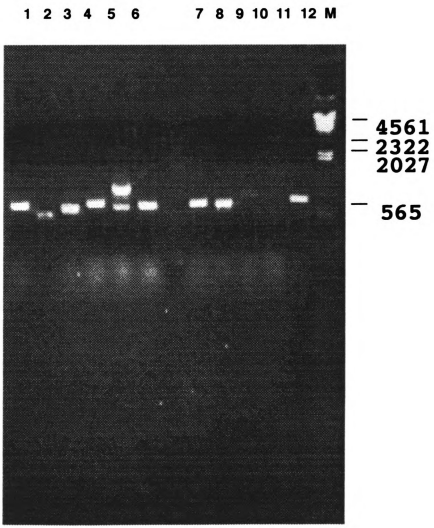


Figure 10b

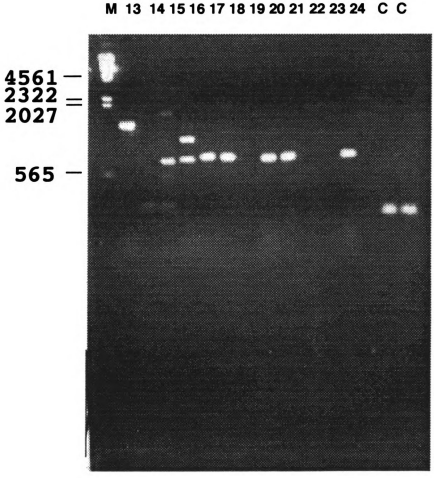
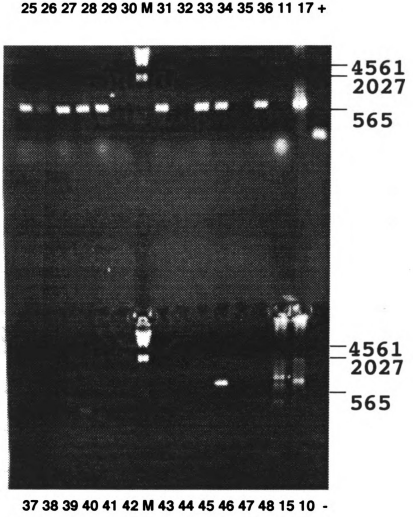


Figure 10c


25 26 27 28 29 30 M 31 32 33 34 35 36 11 17 +
4561
2027
565
4561
2027
565
37 38 39 40 41 42 M 43 44 45 46 47 48 15 10 -

dihydrofolate reductase (hDHFR) gene. At the present time, the biological significance of these two interactions, especially in the context of pre-mRNA splicing, is not apparent.

However, further verifications of these interaction followed by *in vitro* coprecipitation would demonstrate that these interactions actually occur *in vivo*. A more detailed discussion will be given in the following section. How the autonomous transcriptional activation by 'bait' protein alone might have affected the whole screening process is of concern. To circumvent this autonomous activation of transcription by the galectin-3 fusion protein, a galectin-1 fusion protein which lacks the amino terminus of galectin-3, yet retains splicing activity will be tested for its autonomous transcriptional activation and used as a 'bait' if galectin-1 does not autonomously activate transcription of the reporter genes. In case of both galectin-1 and galectin-3 activating without Gal4 activation domain, the partial deletion of the gene might be necessary to use them as baits although doing so may produce the undesirable effects.

Discussion

Possible ligands for galectin-3

There are four possible candidate molecules for the nuclear ligand(s) for galectin-3: 1) Galectin-3 might directly interact with pre-mRNA altering the structural conformation of pre-mRNA or its intermediates for proper catalytic reaction. Galectin-3 lacks RNA binding motifs, although a recent report suggests that galectin-3 has RNA-binding affinity (49).

2) Another possible candidate for galectin-3 ligand(s) would be nuclear carbohydrates or glycoproteins containing the ß-galactoside structure. However, there has not been any carbohydrate structure with detectable affinity for galectin-3 found in the nucleus. The only nuclear proteins known to be glycosylated include nucleoporins (50a) and some transcription factors (50b) with O-linked GlcNAc which has no affinity for galectin-3.

3) Galectin-3 might interact with snRNPs or non-snRNP proteins that mimic the lac structure. A dodecapeptide containing the consensus sequence Tyr-Pro-Try was found to bind to the plant lectin convanavalin A (Con A) with an affinity comparable to that of a known carbohydrate ligand (51). Also, the possibility that some proteins are expressed as distinct compartment-specific isoforms, have distinct endogenous biological actions within each

compartment, and are regulated in a compartment-specific manner as a function of physiological state has emerged as another theme of biology (52). The importance of the CRDs which are highly conserved across the galectin family, for binding to glycoligands is well established. However, depending on the availability of ligand(s) and specific compartmentalization, galectin-3 could interact with proteins other than glycoproteins carrying out a distinct biological role.

4) Finally, galectin-3 may interact with a nuclear protein or any structure that binds at a site other than carbohydrate recognition domain (CRD) of galectin-3. The limited sequence homology of the N-terminus of galectin-3 to an hnRNP protein could implicate a possible interaction between the N-terminal half of galectin-3 and an hnRNP protein. Mutations in other functionally unique regions of galectin-3 such as Pro-Gly-Tyr rich repeats can provide more information to determine if such interactions exist.

Although all four possible candidates may be potential classes of ligand(s), a nuclear protein mimicking a specific carbohydrate structure could be the most plausible candidate for the following reasons. First, UV cross-linking has been used as a strategy to identify specific protein-pre-mRNA interactions in splicing extracts (53). In this protocol, the splicing extract is exposed to UV light at various times during the course of splicing to induce covalent cross-linking between radioactively labeled

pre-mRNA and proximally interacting protein(s) and followed by digestion of RNA by RNase treatment. The cross-linked RNA-protein complexes were separated by SDS-PAGE and visualized by autoradiography. However, UV cross-linking experiments resulted primarily in the detection of hnRNP proteins and a few other proteins such as Polypyrimidine Track Binding Protein, U2AF, and 200 KDa U5 SnRNP (54). Galectin-3 was not detected by this method (Dagher, unpublished data).

Second, no one has yet found the presence of a glycoconjugate with affinity for galectins in the nucleus. Although O-GlcNAc has no affinity for galectin-3, it has been shown that GlcNAc residues of nucleoporins are galactose acceptors for galactosyltransferase using an *in vitro* assay (55). The resulting product, N-acetyllactosamine is a natural, high affinity ligand for galectin-3. The presence of such nuclear glycoconjugates has yet to be identified *in vivo*.

Finally, galectin-1, which contains only the carbohydrate recognition domain, has been shown to replace galectin-3's role in splicing in galectin-depleted splicing extracts (Patterson et al. TIGG article). Also, the C-terminal half of galectin-3 containing only the CRD was shown to reconstitute the splicing in galectin-depleted splicing extracts with 4-5 fold less efficiency than wild type galectin-3 (in press). Galectin-1's ability to substitute for galectin-3's role in splicing indicates the

C-terminal half of galectin-3 may be sufficient for splicing. Also, structural studies that have shown a highly homologous amino acid sequence and identical folding conformation between galectin-1 and the C-terminus of galectin-3 suggest that the C-terminal half might be responsible for its involvement in splicing. The functional role of the N-terminal half in terms of splicing, however, is not clear. Whether it has a supporting role for the optimal activity of C-termimal half or was introduced during evolution with no specific purpose remains to be determined. It is also possible to speculate that galectin-3 attains its role as a splicing factor via a novel mechanism involving the N-terminal half, which may be different than the mechanism of galectin-1 in splicing.

Use of Yeast Two Hybrid System to Identify Galectin Ligands that are involved in splicing

In an effort to identify ligands, affinity column chromatography with galectin-3 immobilized to sepharose beads to attract the ligand(s) was employed. However, this technique has not succeeded in identifying the ligand for galectin-3. For instance, passing the entire HeLa nuclear extract over a galectin-3 column under various salt conditions resulted in detection of no bound protein (high salt) or numerous proteins (low salt). As mentioned earlier, this might imply that the interaction of galectin-3 with a possible protein ligand is weak or transient *in*

vivo, and the same interaction can not be maintained in the *in vitro* system.

The failure to identify galectin-3 nuclear ligand(s) using conventional biochemical techniques led to the utilization of the yeast two hybrid system. This unbiased screening method is very powerful because the bait protein, galectin-3, forms its native structure *in vivo*. Also, the overexpression of both galectin-3 and its potential ligand(s) *in vivo* coupled with strong reporter genes could detect even a weak or transient interaction. Even more powerful, the yeast two hybrid system can be further explored to pinpoint a specific region of interaction between two proteins by introducing site-directed mutations to one or both proteins.

However, there are some inherent barriers in the yeast two hybrid system that might actually prohibit the identification of cellular ligands. First of all, the Gal4 domains might interfere with the ability of the test protein to interact (56). Although galectin-3 is stably expressed, the DNA-BD could potentially interfere with the overall structure or a confined region. It is known that the hydrophobic amino acid residues of extreme N-terminal and C-terminal ends are important for proper folding of galectin-1 and subsequent dimerization (57). Therefore, linkage between the C-terminal end of the Gal4 DNA-BD and the N-terminal end of galectin-3 could be a detrimental factor for interaction if this linkage alters the

hydrophobic or hydrophilic interaction of certain domains or destabilizes the overall structure of the protein. Another potential drawback is that the host organism may not provide the proper post-translational modifications required for native folding or interactions. There are only two known post-modifications for galectin-3 to date; phosphorylations on ser-6 and ser-12. The fact that recombinant murine galectin-3, which is not phosphorylated at all could restore splicing activity in the galectin-depleted NEs suggests that phosphorylation of galectin-3 may not be important for splicing activity. However, it may affect the interaction with some ligands if those ligands interact with phosphorylated galectin-3 during a specific physiological state or in different compartments of cells.

Autonomous transcriptional activation

The observation that galectin-3 fused to Gal4 DNA-BD activated reporter genes without the activation domain leads to speculation that galectin-3 is a possible transcriptional activator. In general, the DNA-BD/bait protein could induce reporter gene expression without an AD/library protein if the bait protein has a transcriptional activation domain. This is especially likely if the target protein is a transcription factor. However, other proteins that are not normally involved in transcription are sometimes capable of activating transcription. This unexpected result, along with recent findings that implicate a coupled mechanism of

transcription and pre-mRNA processing (58), presents an exciting potential model that galectin-3 could be involved in the coupling of transcription and splicing. RNA polymerase II colocalizes with some splicing factors via the phosphorylated Carboxyl Terminal Domain (CTD) in discrete nuclear regions (59). Splicing factors, Sm snRNPs and SR proteins have been shown to coprecipitate with CTD of RNA polymerase II in a phosphorylation-dependent manner indicating the phosphorylated CTD of RNA polymerase II may associate with splicing factors.

Repeated heptapeptides with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (60) of the CTD of PolII are somewhat comparable to the hexapeptide expressed from a diverse peptide library, composed of Tyr-Pro-Ser repeats in its amino acid composition. Ser and Tyr in each respective peptide has a hydroxyl group and thus potentially can be phosphorylated and linked by Proline. As stated earlier, those hexapeptides have been identified as ligands for the plant lectin concanavalin A by screening a large, diverse peptide library expressed on the surface of filamentous phage (51). Considering the different carbohydrate specificity between galectin-3 and Con A, the structure of serine, proline rich peptide differing slightly with tyrosine instead of serine, could be a natural substrate for galectin-3. Therefore, the possible isolation of CTD of Pol II or similar peptides composed of those amino acids as a galectin-3 ligand using the yeast two hybrid system could

provide a strong evidence for the association of transcription and splicing machineries.

Other relevant biological ligands for galectin-3 could be identified in the yeast two hybrid screening such as bcl-2, laminin, a kinase, cell growth regulators, or even collagenase based on the previous suggestions of galectin-3's biological roles and subcellular localization. For example, galectin-3 has been suggested to inhibit apotosis in Jurkat E6-1, a human leukemia cell line. Bcl-2, a well-characterized repressor of apoptosis, was shown to be coprecipitated with galectin-3 immobilized to sepharose. Also, galectin-3 (35 kDa) and Bcl-2 (29 kDa) exhibit a low, but significant, sequence similarity including the highly conserved asparagine-tryptophan-glycine-arginine (NWGR) among galectin family (28 % identity and 48 similarity)(29).

Interpretation of potential interactions between galectin-3 and DHRF and rL17

What is the significance of galectin-3's interaction with ribosomal protein L17? Characterization of rL17 has been little done to date. There is evidence that the level of rL17 is elevated in human colon carcinoma cell lines implying its significance for cell growth. However, it is unknown how rL17 enters the nucleus and binds to ribosome complexes. It has been demonstrated that a large subunit ribosomal protein L5 is composed of at least two distinct functional domains, RNA binding domain and nucleolar transport domain for constitutively expressed 5 S RNAs.

Transport of 5 S RNA is accompanied by rL5 and is inhibited when the nucleolar transport domain is mutated (61). It appears rL5 has to be localized in the nucleoplasm at some point between when the protein is translated and when it is localized in the nucleolus for ribosome assembly. You can raise questions how many ribosomal proteins including rL17 are involved in the transport processing of 5 S RNA although there is no evidence rL17 interacts directly with rL5. The possibility that rL17 starts its journey from the cytoplasm to nucleolus via the nucleoplasm serving as a docking molecule for 5 S RNA and transporting it into the nucleolus can not be ruled out. Galectin-3 might coordinate this transport process by directly binding to the complex and dissociating before the 5 S RNA compex enters the nucleolus or transiently binds to the complex to optimize the process.

Interestingly, ribosomal proteins are one of the most frequent false positives detected in the yeast two hybrid system. Further verifications are required to either prove rL17 is a false positive or a true positive ligand involved in the 5 S RNA transport as proposed above or some other nuclear process.

DHFR also has been identified as one of the most frequent false positives. With no cytoplasmic function(s) of galectin-3 determined yet, the interaction between DHFR and galetin-3 might present somewhat an intriging aspect of

galectin-3's roles in the cytoplasm if such interaction can be proven.

Future plans

At the present time, the biological significance of interactions between galectin-3 and DHFR and rL17 are not understood in the context of splicing activity. Neither proteins have been shown to be splicing factors. It is possible, however, that these interaction might be important for other biological roles. Further experiments are required to postulate a biological role of these interactions. In the mean time, it was our main focus to identify splicing factors as ligands for galectin-3 since 1) we intend to test the model that galectin-3 associates with splicing factors as a part of spliceosome, and 2) identification of splicing factors as ligands could provide insight to the understanding of mechanism of interaction between galectin-3 and its ligands. For example, a mutant in which the carbohydrate binding activity is abolished can be tested to determine whether the interaction between galectin-3 and its splicing ligands is still observed with some information about the binding affinity that can be derived from a quantitative assay. We could demonstrate whether the CRD of galectin-3 is responsible for its ligands binding. Therefore, it is my intention to isolate more His / LacZ positives and identify them. At the same time, the screening with recombinant human galectin-1 as a bait could solve the autonomous activation of reporter

genes and identification of galectin-1's ligands leads to further speculation whether such ligand(s) interacts with galectin-3 also. Use of recombinant human galectin-3 as a bait as opposed to murine galectin-3 in screening HeLa cDNA library could also reduce the number of non-specific interactions if a host species specificity exists.

References

1. Barondes, S.H. et al. (1994) Cell **76**, 597-598
2. Hirabayashi, J., and Kasai, K.I. (1993) Glyco biol. **3**, 297-304
3. Wells, V., and Mallucci, L., (1991) Cell **64**, 91-97
4. Zhou, Q., and Cummings, R.D. (1993) Arch. Biochem. Biophys. **300**, 6-17
5. Perillo, N.L., and Baum, L.G. (1995) Nature **378**, 736-739
6. Raz, A. et al. (1990) Int. J. Cancer **46**, 871-877
7. Cherayil, B.J., et al. (1990) Proc. Natl. Acad. Sci. USA **87**, 7324-7328
8. Current special issue of TIGG (1997) Jan.
9. Barondes, S.H. (1994) J. Biol. Chem. **269**, 20807-20810
10. Liang, J.G., and Wang, J.L. (1988) Biochemistry **27**, 5329-5334
11. Kasai, K.I., Hirabayashi, J.,(1996) J. Biochem. **119**, 1-8
12. Sato, S., and Hughes, R.C., (1994) J. Biol. Chem. **269**, 4424-4430
13. Harrison, F.L., and Wilson, T.J., (1992) J. Cell.Sci. **101**, 635-646

14.Cherayil, B.J., et al. (1989) J. Exp. Med. **170**, 1959-1972

15.Runyan, R.B., et al. (1988) J. Cell. Biol. **107**, 1863-1871

16.Wang, J.L., Werner, E.A., Patterson, R.J. (1992) Biochem. Soc. Trans. **20**, 269-274

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

18.Harrison, F.L., (1990) J. Cell. Sci. **100**, 9-14

19.Sanford, G.L., and Harris, H.S. (1990) FASEB J. **4**, 2912-18

20.Cooper, D.N.W., Massa, S.M., and Barondes, S.H. (1991) J. Cell. Biol. **115**, 1437-1448

21.Poirier, F. et al. (1992) Developement **115**, 143-155

22.Perillo, N.L., and Baum L.G. (1995) Nature **378**, 736-739

23.Yaron, R.H., et al. (1995) J. Biol. Chem. **270**, 3447-3453

24.Oda, Y., et al. (1993) J. Biol. Chem. **268**, 5929-5939

25.Sato, S., and Hughes, R.C. (1992) J. Biol. Chem. **267**, 6983-6990

26.Lamm, G.M., and Lamond, A., (1993) Biochimica et. Biophysica. Acta. **1173**, 247-265

27.Alblandt, K., et al (1987) Proc. Natl. Acad. Sci. USA **84**, 6859-6863

28.Anne Mey, J.P., and Revillard (1996) J. Immunol. **156**, 1572-1577

29.Yang, R.Y., and Liu, T.F., (1996) Proc. Natl. Acad. Sci.

USA **93**, 6737-6742

30.Hebert, E., and Monsigny, M., (1994) Biol. Cell **81**, 73-76

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

32a.Moutsatsos, I.K. et al. (1987) Proc. Natl. Acad. Sci. USA **84**, 6452-6456

32b.Cowles, E.A. et al (1990) J. Biol. Chem. **265**, 17706-17712

33a.Sharp, P. (1994) Cell **77**, 805-815

33b.Krainer, A.R. et al. (1984) Cell **36**, 993-1005

33c.Sawa, H., and Abelson, J., (1992) Proc. Natl. Acad. Sci. USA **89**, 1126-1127

33d.Parker, et al. (1987) Cell **49**, 229-239

34a.Wasserman, D.A., and Steitz, J.A. (1992) Science **257**, 1918-1925

34b.Stakinis, D., and Reed, R. (1994) Mol. and Cell. Biol. **14**, 2994-3005

35.Kruger, K., et al., (1982) Cell **31**, 147-157

36.Konarska, M.M., and Sharp, P.A. (1986) Cell **46**, 845-855

37.Lamm, G.M. and Lamond, A. (1993) Biochimica et Biophysica Acta. **1173**, 247-265

38.Dagher, S.F., and Patterson, R.J. (1995) Proc. Natl. Acad. Sci. USA **92**, 1213-1217

38a.Lamond, A., (1993) Bioessay **15**, 595-602

39.Keegan, L., and Ptashne, M (1986) Science **231**, 699-704

40.Brent, R., and Ptashne, M. (1985) Cell **43**, 729-736

41a.Ma, J., and Ptashne, M. (1987) Cell **51**, 113-119

41b.Sadowski, I., at al.(1987) Nature **335**, 564-564

42.Ma, J., and Ptashne, M. (1988) Cell **55**, 443-446

43.Fields, S. and Song, O. (1989) Nature **340**, 245-246

44.Gyuris, J., and Brent, R., (1993) Cell, **75**, 791-803

45.Harper, J.W., et al. (1993) Cell **75**, 805-816

46.Gietz, D. et al. (1992) Nucleic Acids Res. **20**, 1425-1435

47.Clonetech User Manual page 34 (1996)

48.Kaiser, P. and Auer, B. (1993) BioTechniques **14**, 55245.

49.Wang, L., and Raz, A. (1995) Biochem. Biophys. Res. Commun. **217**, 292-303

50a.Hanover, J.A. (1987) J. Biol. Chem. **262**, 9887-9894

50b.Jackson, S.P. and Tjian, R. (1988) Cell **57**, 125-133

51.Oldenburg, K.R. et al. (1992) Proc. Natl. Acad. Sci. **89**, 5393-5397

52.Smalheiser, N.R. (1996) Mol. Biol. Cell. **7**, 1003-1014

53.Whittaker, E. and Beggs, J.D. (1991) Nucleic Acid Res. **19**, 5483-5489

54.Wyatt, J.R. and Steitz, J.A. (1992) Genes Dev. **6**, 2542-2553

55.Miller, M.W., and Hanover, J.A. (1994) J. Biol. Chem. **269**, 9289-9297

56.Van Aelst et al. (1993) Proc. Natl. Acad. Sci. USA **90**, 6213-6217

57.Cho, M. and Cummings, R.D., (1996) Biochemistry **35**, 13081-13088

58.McCraken, S., et al. (1997) Science **385**, 357-361

59.Bregman, D.B., et al. (1995) J. Cell. Biol. **129**, 287-298

60.Kim, E.K., et al (1997) J. Cell. Biol. **136**, 19-28

61.Michael, W.M., and Dreyfuss, G. (1996) J. Biol. Chem. **271**, 11571-11574

Appendix

Introduction

These data were generated from my early works before I was solely engaged in the yeast two hybrid system to identify the galectin-3 ligand(s). Although these data were not confirmed by repeat experiments and further supporting evidence, they provide some valuable insights about the current working model of our investigating topic. Some data briefly describe *in vitro* splicing assay that we use as a model system to understand the role of galectin-3 in splicing. There are also some data that present compelling evidence that galectin-3 physically associates with the spliceosome. Further pursuit of these experiments will contribute to the confirmation of our current working model.

Appendix 1 : Time Course of the Splicing Reaction

Materials and Methods and Results

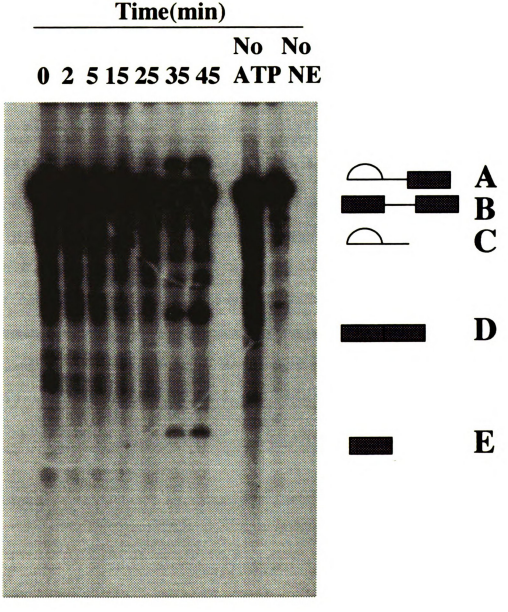
The MINX pre-mRNA substrate for *in vitro* splicing was radioactively transcribed with [^{32}P]GTP. The monomethyl cap was added during SP6 polymerase (New England Biolabs) transcription (1). Splicing reaction mixture (10 µl) contained NE prepared by the Dignam method (2), 25 mM $MgCl_2$, 1.5 mM ATP, 20 mM Creatine Phosphate, 0.5 mM Dithiothreitol, and 20 units of RNAsin (Clonetech). Splicing reactions were carried out at 30°C for 45 minutes and worked up as described (3). The extracted RNAs were subjected to electrophoresis through 13 % polyacrylamide (bisacrylamide / acrylamide, 1.9 : 50 (wt / wt) / 8.3 M urea gels followed by autoradiography.

For the kinetic experiment, the individual splicing reactions were removed from 30 °C water bath at appropriate time points and kept on ice until the completion of incubation of the last sample and further work-up. As negative controls, ATP was omitted from the splicing mixtures and no NE was added to the splicing mixture. Mature products began to form after incubation for 35 minutes. First splicing intermediate (laria-exon 2, A) begins to appear about 25 minutes.

Figure of Appendix 1

This figure shows time-dependent formation of splicing products. Splicing mixtures were incubated for indicated periods of time and the products of splicing reactions were analyzed by electrophoresis through a 13 % polyacrylamide /urea gel and autoradiography. The positions of migration of the pre-mRNA substrate (B), the splicing intermediates (lariat-exon 2, A; lariat, C; exon 1, E), and mature product (exon 1-exon 2, D) are indicated on the right. Two negative controls show that splicing is an ATP dependent process.

Appendix 1
Time course splicing assay



Appendix 2 : Time Course of Splicing Complexes Formation Detected by Native Gel Electrophoresis

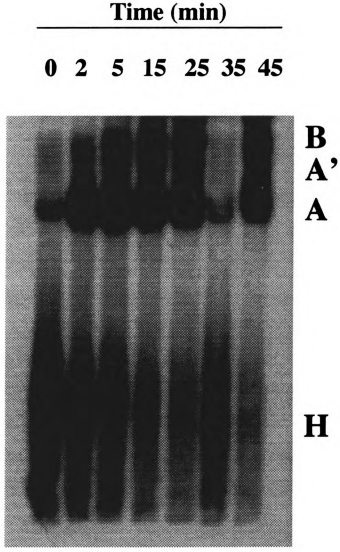
Materials and Methods and Results

The formation of splicing complexes was monitored by native gel electrophoresis. Heparin was added (final conc. 0.6 mg / ml) to each splicing reaction (10 μl) and the mixtures were incubated at 30°C for 15 minutes. Each sample was loaded onto a pre-run polyacrylamide gel (bisacrylamide / acrylamide 1:80 (wt / wt)) with 1 μl of glycerol containing 20 % each of bromphenol blue, xylene, cyanole, and phenol red, and electrophoresis was carried out in 0.5 M Tris base / 0.5 M glycine, pH 8.8, at 4 °C and 25 V / cm for 90 minutes. The migration of splicing complexes was determined by autoradiography. As described above, each splicing sample was removed from 30 °C water bath at indicated time points and subjected to native gel electrophoresis.

Figure of Appendix 2

This figure shows time-dependent formation of spliceosome complexes. Splicing mixtures were incubated for the time periods indicated above and the whole splicing mixtures were analyzed by native gel electrophoresis. These data show that H complex, also known as pre-commitment complex, forms quickly and before the pre-mRNA associates with spliceosomes. Upon incubation, H complexes are chased to A complexes, and subsequently, to A', B which are active splicing complexes. As splicing reactions continue, the formation of H complexes decreases.

Appendix 2
Detection of splicing complexes
using native gel electrophoresis



Appendix 3 : Lactose Depletion of the Galectins During the Splicing Reaction

Materials and Methods and Results

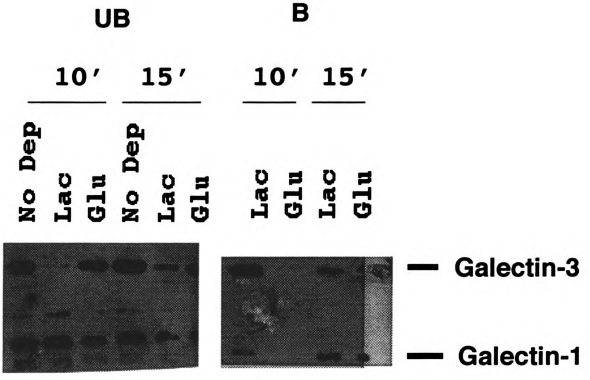
The splicing mixture was subjected to binding to α-Lac insolubilized on 6 % beaded agarose (LAC-A) (Sigma) using a ratio of 3 volume Lac-A to 1 volume splicing mixture. Glucose agarose was used as a negative control for depletion. The unbound fractions from each affinity matrix were mixed with SDS-PAGE sample buffer, boiled and loaded onto SDS-PAGE gels. Bound fractions were eluted from beads in SDS-sample buffer. SDS / polyacrylamide gel electrophoresis was carried out in 12 % acrylamide gels. The separated proteins were transferred to polyvinylidene difluoride (PVDF) membranes (Bio-Rad) in 25 mM Tris / 193 mM glycine / 20 % methanol. After the protein transfer, the membranes were blocked with 5 % nonfat dry milk dissolved in phosphate-buffered saline (PBS; 10 mM sodium phosphate / 150 mM NaCl, pH 7.2) containing 0.05 % Tween 20 (PBS-T). Galectin-1 and galectin-1 were detected on the membranes by using the rat monoclonal antibody anti-MaC 2 (galectin-3) and the rabbit polyclonal antibody anti-L14

(galectin-1) respectively. Incubation with primary antibody (freshly diluted 1 : 5000 in 5 % nonfat dry milk in PBS-T) was carried out at room temperature for 2 hours, followed by five washes in PBS-T for 15 minutes each. The membranes were incubated with horseradish peroxidase-conjugated goat anti-rat or anti-rabbit antibodies (Pierce) diluted 1:3000 in PBS-T for 30 minutes at room temperature and washed five times for 15 minutes each in PBS-T. Membranes are exposed to X-ray film after incubating the membranes with the enhanced chemiluminescence reagents from Amersham according to the manufacturer's protocol.

Figure of Appendix 3

A comparison was made of the levels of galectin-1 and galectin-3 in nuclear extracts with those in the unbound (UB) and bound (B) fractions of nuclear extracts subjected to adsortion on Lactose-Agarose or Glucose-Agarose during splicing condition. While galectin-3 is mostly adsorbed to Lac-A and found in the bound fraction, galectin-1 is detected in unbound fraction. These data suggest that the CRD of galectin-3 is accessible to Lac-A but the CRD of galectin-1 is not during the splicing condition. It also might imply galectin-1 associates with splicing complexes via its carbohydrate recognition domain. For galectin-3, it is unknown whether the N-terminal half or a region other than CRD in C-terminal half is responsible for the association with splicing complexes.

Appendix 3
Lactose / Glucose Depletion of HeLa Nuclear Extract During pre-mRNA Splicing



Appendix 4 : Association of ATP-dependent Splicing Complexes with Galectin-3 Detected by Native Gel Electrophoresis

Materials and Methods and Results

The splicing mixture was subjected to binding to α-Lac insolubilized on 6 % beaded agarose (LAC-A) (Sigma) during normal splicing conditions using a ratio of 3 volume Lac-A to 1 volume splicing mixture. Cellobiose-agarose was used as a negative control for depletion. Each splicing mixture, incubated for the indicated times followed by subsequent depletion, was loaded onto a pre-run polyacrylamide gel (bisacrylamide / acrylamide 1:80 (wt / wt)) with 1 μl of glycerol containing 20 % each of bromphenol blue, xylene, cyanole, and phenol red, and electrophoresis was carried out in 0.5 M Tris base / 0.5 M glycine, pH 8.8, at 4 ℃ and 25 V / cm for 90 minutes. Splicing complexes are detected by autoradiography.

Figure of Appendix 4

Splicing reactions were removed after incubating for 5 minutes and poured over Lactose-A or Cellobiose-A. The unbound fractions were further analyzed by native gel eletrophoresis. These data show that splicing complexes A and B were removed by the Lac-A column, but not by the Cello-A. These data imply that galectin-3 associates with ATP-dependent splicing complexes and the splicing complexes can be depleted with galectin-3 by Lac-A chromatography. This strongly suggests that galectin-3 associates with active spliceosome during splicing and the binding between galectin-3 and spliceosome occurs via a region other than carbohydrate recognition domain.

Appendix 4
Galectin-3 Coprecipitates With ATP-dependent Splicing Complexes by Lactose-Sephrose column