J. Biochem. 119, 1-8 (1996)

Galectins: A Family of Animal Lectins That Decipher Glycocodes

Ken-ichi Kasai and Jun Hirabayashi[1]

Faculty of Pharmaceutical Sciences, Teikyo University, Sagamiko, Kanagawa 199-01

Received for publication, September 25, 1995

Galectins, animal lectins exhibiting specificity for galactosides, are now known to be widely distributed from lower invertebrates, such as sponges and nematodes, to higher vertebrates. The origin of the family can be traced back to the Precambrian era. They are classified into proto-, chimera-, and tandem-repeat types on the basis of protein architecture. The molecular functions of these types should be different because they can cross-link pairs of biomolecules of different combinations. Their biological significance, however, is not yet fully understood because they are involved in too many phenomena, such as differentiation, morphogenesis, metastasis, *etc.*, and too many problems remain unsolved, such as those regarding their controversial cellular localization, mechanism of externalization, *etc.* Nevertheless, such difficulties seem to indicate their importance as household equipment and their common roles throughout the animal kingdom. They are likely to be responsible for recognizing the *N*-acetyllactosamine (LacNAc) structure, which is included in various glycoconjugates and considered to be an important glycocode, and then carry out appropriate tasks under given circumstances. Recently, crystallographic studies revealed that galectins and legume lectins such as concanavalin A have a common topology in spite of the absence of sequence homology. This suggests a possible relationship between animal and plant lectins, and the existence of a lectin super family. Studies on the galectin family are becoming increasingly important for glycobiology.

Key words: *N*-acetyllactosamine, animal lectin, *Caenorhabditis elegans*, galectin, glycocode.

Biological phenomena based on the specific recognition of sugar chains have attracted much less attention in comparison with those of nucleic acids and proteins. The biological significance of glycoconjugates is not yet fully understood, and too many questions remain unanswered. No big discovery, such as the double helix structure of DNA, which changed our concept of the life, has been made, nor have applications appeared such as gene engineering, which is greatly changing our way of life.

Under such circumstances, it is not surprising that galectins, animal lectins which do not induce biological responses as dramatic as those induced by plant lectins in *in vitro* experiments, have been less interesting for almost all researchers in biosciences. However, this does not mean they are unimportant, because the very diverse cellular functions attributed to galectins have made their biological significance rather ambiguous. Underestimation of galectins is a great mistake, because lines of experimental evidence have shown that they originated more than 800 millions years ago and are now widely distributed in the animal kingdom, from nematodes to vertebrates, and have conserved their basic molecular properties, such as specificity and protein architecture. These facts suggest that galectins should be important household equipment. Even if they do not play principal roles before the footlights, they are indispensable as all-around supporting actors.

Galectins are considered to decipher glycocodes: They recognize certain sugar structures, and then carry out appropriate tasks in given circumstances. Glycocodes are extremely different from the codes written in either nucleotides or amino acids, and thus, have been poorly understood. Therefore, the immaturity of glycoscience is one of the biggest reasons for our ignorance of galectins. We can conversely expect that studies on galectins will greatly promote glycobiology and consequently lead to the discovery of new principles of life.

This short article attempts to give readers a general view of the galectin family. Since the detailed description of each member has to be kept to be minimum in order to make the discussion clear, and to avoid giving readers too confused image, the readers are encouraged to refer also to more comprehensive reviews that have already appeared, if necessary (*1-5*).

1. Properties of galectins

1.1. Galectin is a new family name for a group of animal lectins. The term galectin might still be unknown to almost all readers. It is a new family name created for a group of animal lectins. Since the 1980s, lectins having common properties, that is, specific for β-galactosides and requiring no metal ion, have been found in various vertebrates (*e.g.*, electric eel, frog, chicken, mouse, human, *etc.*). They had been called by different names depending on the researchers (*e.g.*, electrolectin, soluble lectins, galaptins, S-Lac lectins, β-galactoside-binding lectins, S-type lectins,

[1] The work done in our laboratory was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, and grants from the Naito Foundation, the Kihara Memorial Yokohama Foundation for the Advancement of Life Sciences, and the Mitsubishi Foundation.

etc.). Since they turned out to have homologous sequences to each other, and the number was continuously growing, the necessity for a generic family name became very great in order to avoid inconvenience and confusion. Then, agreement was reached to use the new term, galectin, among concerned researchers (*6*).

1.2. The galectin family is growing. In the course of the rather short history of about two decades of animal lectin research, two groups of animal lectins became apparent, though not all animal lectins belong to these two groups. One distinguishing features was metal dependency. One requires the Ca ion for sugar binding while the other does not (*7*). The former is now called the C-type lectin family, and the latter the galectin family. Each family is also characterized by a specific amino acid sequence motif for sugar binding (CRD, standing for carbohydrate-binding domain).

Since the first discovery of a galectin in the electric organ of electric eel (*8*), galectins of small size have been found in various tissues of vertebrate species. They are non-covalent homodimers of a subunit of about 14 kDa, which comprises the minimum structural element (proto type) of the galectin family. Galectins of this type usually show hemagglutination activity. Our first determination of the complete primary structure of chicken 14-kDa galectin (*9, 10*) triggered structural studies on them, and the primary structures of various galectins [human (*11-13*), cow (*14*), rat (*15*), mouse (*16*), chicken 16 kDa (*17*), *etc.*] were reported successively. The production of recombinant galectins also became popular (*18*). Alignment of the revealed sequences indicated that the central parts are mostly well conserved and presumably correspond to the sugar-binding sites. Characteristic residues that compose the galectin CRD are shown in large capital letters in Fig. 1. This domain was shown to be entirely encoded by the third exon of the gene (*19*).

Then, the existence of a second type of galectin, a carbohydrate-binding protein of 3T3 cells of larger size (about 35 kDa, called CBP35), was revealed (*20*). The C-terminal half of this protein was found to contain the galectin CRD motif. The N-terminal half, however, is completely different. Therefore, this is a chimeric protein. Proteins reported as the IgE-binding protein of certain leucocytes and Mac-2 (a Macrophage-specific cell surface antigen) turned out to be the same protein as CBP35 (*21, 22*). Galectins of this type do not show hemagglutination activity. This is one of the reasons why they were not firstly reported as lectins.

Though galectins had only been found in vertebrate species for a long time, we found a galectin of 32 kDa in the nematode, *Caenorhabditis elegans* (*23*). This was the first demonstration of not only a nematode *lectin* but also a new type of galectin. The molecular architecture of the 32-kDa molecule was found to be unique because two CRD motifs are tandemly repeated in a single polypeptide, suggesting that it is a divalent monomer. This finding was soon followed by that of a similar type of galectin in mammals (*24*). Recently, we found *C. elegans* also has a proto-type galectin (*25*). Therefore, the nematode is likely to be equipped with a similar set of galectins to those in vertebrates, though a chimera-type galectin has not yet been found. Soon after the discovery of the 32-kDa galectin in *C. elegans*, marine sponges, more primitive multicellular animals, were reported to also have galectins (*26*). Therefore, the galectin family is now known to be distributed widely in the animal kingdom.

1.3. Common molecular properties of the galectin family. Galectins are water-soluble proteins. However, for the initial process of purification, it is necessary to add some hapten sugars such as lactose to the extraction buffer in order to dissociate them from insoluble substances of starting tissues. This does not mean they are integrated into lipid membranes, but they bind to sugar chains of insoluble glycoconjugates such as glycoproteins, glycolipids, and proteoglycans. They can be readily purified by affinity chromatography on adsorbents which immobilize galactoside derivatives.

Since the sugar-binding activity of galectins was often lost in the absence of a reducing agent such as β-mercaptoethanol, the presence of free SH group(s) was once thought to be essential. This is the reason why galectins were once inappropriately called S-type lectins. However, our site-directed mutagenesis experiments clearly showed that SH groups were not essential for sugar binding (*27*). This was also supported by the finding of the absence of a cysteine residue in *C. elegans* 32-kDa galectin (*23*). Oxidation of SH groups caused a global conformation change which resulted in loss of sugar binding ability (*28*). Therefore, the term

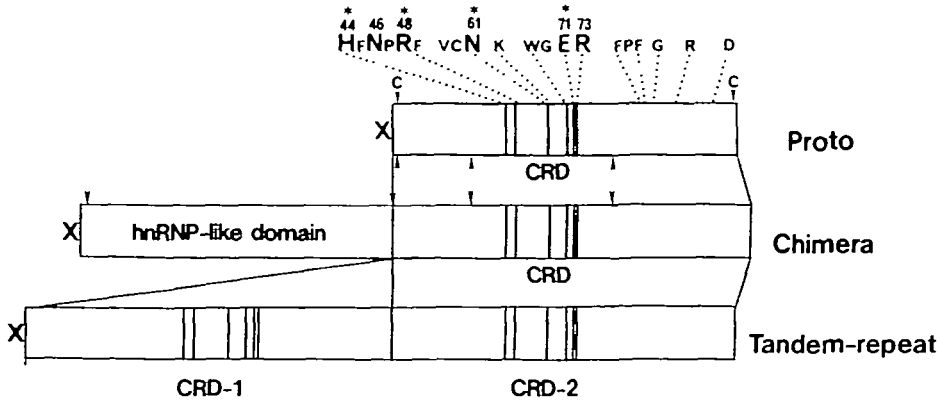

Fig. 1. **Schematic representation of the three types of galectins (proto-, chimera-, and tandem-repeat).** The proto-type is composed of only a single lectin domain. The chimera-type is composed of two parts: a C-terminal half containing the galectin CRD and an N-terminal half of unknown function. The tandem-repeat type is composed of two homologous CRD domains. Above the proto-type structure, conservative residues are shown in single letters. Large capital letters denote the most important residues for sugar binding. Asterisks denote residues mostly conserved in all galectins studied so far. c indicates cysteine residues, the oxidation of which results in loss of sugar-binding activity. Arrowheads indicate the positions of intron insertion in the case of mammalian galectin-1 (*19*). Residue numbers are those of human galectin-1 (*12*).

S-type lectin is now not recommendable because it tends to lead to misunderstanding.

All galectins studied so far have blocked N-termini, and an acetyl group was demonstrated to be the blocking group (*10, 12*). Galectins occasionally contain cysteine residues, but no disulfide bond is formed and all SH groups are in a free state. No galectins are glycosylated. Cloning of cDNAs revealed that galectins are synthesized without a signal sequence (*9*). These characteristics strongly suggest that galectins are designed as intracellular proteins.

1.4. Galectins are classified into proto-, chimera-, and tandem-repeat types according to their molecular architecture. An increasing number of galectins have been found in various animal species, especially in mammals. In order to distinguish this variety of mammalian galectins, it was proposed to give them sequential numbers, and so far the numbered mammalian galectins have reached as many as 8. However, it is difficult to have a general view because the numbers have no relation to molecular properties or biological functions. It seems helpful to group all galectins according to their molecular architecture into the following three types: proto-, chimera-, and tandem-repeat types (Table I).

The proto-type comprises small proteins of about 15,000 Da containing only one CRD domain, and the most common one is mammalian galectin-1. In addition, galectins-2, -5, and -7 have been found in mammals. Of them, galectins-1, -2, and -7 exist as non-covalent homodimers under physiological conditions, while galectin-5 was reported to exist as a monomer (*29-31*). In birds, two proto-type galectins have been found. The chimera-type, which has been found only in mammals, has a molecular weight of about twice that of the proto-type (30-35 kDa), and is composed of two different domains (*20*). The C-terminal half is the galectin domain containing one CRD. The N-terminal half exhibiting no homology to galectins but is related to components of the heteronuclear ribonucleoprotein complex (hnRNP). Characteristic repetitive sequences rich in proline and glycine are also found (*20-22, 32*). *C. elegans* 32-kDa galectin, and mammalian galectins-4, -6, and -8 belong to the tandem-repeat type. They contain two CRD domains in a single polypeptide chain (*32, 33*).

From a functional viewpoint, these three types behave differently. Proto-type galectins form divalent homodimers, which consequently will cross-link two glycoconjugates of a very similar nature. Tandem-repeat types are divalent, but the specificity and binding strength of the two binding sites are not necessarily the same (*24, 34*). Therefore, they can form bridges between glycoconjugates of different types. Chimera-types have one sugar-binding site and other unknown regions, which might interact with biomolecules other than sugars. Therefore, they can serve as cross-linkers between glycoconjugates and other biomolecules; in other words, as adaptor molecules between different kinds of biomolecules or as heterobifunctional cross-linker.

Numbering is not recommended for non-mammalian galectins because it is not yet easy to compare them to one of the mammalian galectins. For example, chicken has two proto-type galectins, which have apparent molecular weights of 14 and 16 kDa, and *C. elegans* has one proto-type (16 kDa) and one tandem-repeat type (32 kDa) galectin. They are distinguished by their apparent molecular weights.

1.5. Galectins are distributed widely in the animal kingdom. A systematic survey of galectins in animal species has not been conducted. Since galectins of invertebrate origin were not known for a long time, they were thought to have been created after vertebrates appeared on the earth in order to deal with some new biological processes required only for higher animals. However, the discovery of the *C. elegans* galectins completely changed this (*23*). This was not only demonstration of the first invertebrate galectin, but also confirmation of the fundamental importance of the galectin family throughout animals.

Two big branches of the phylogenic tree of metazoa, *i.e.*, protostomes and deuterostomes, diverged at the very beginning of evolution (Fig. 2). Nematodes are situated on the protostome branch while vertebrates are on the deuterostome one. The fact that both nematodes and vertebrates have galectins indicates the existence of a common ancestor before the divergence of these two branches; that is in the Precambrian era of more than 800 millions years ago. The discovery of sponge galectins pushed back the origin of the galectin family to a more ancient time, *i.e.*, nearly the start of multicellular animals. The fundamental roles of the galectin family does not seem to have changed significantly, because their basic structures have been conserved for more than a billion years. These findings also led us to expect to find galectins in all metazoan animals,

TABLE I. **Galectins classified according to molecular architecture.**

Type	Animal class	Species	Example
Proto	Mammals	Man, mouse, rat, cow, *etc.*	Galectins-1, -2, -5, -7
	Birds	Chicken, quail	14, 16 kDa
	Amphibia	Frog (*Xenopus, Rana*)	14, 16 kDa
	Fish	Electric eel, conger eel	14, 16 kDa
	Nematodes	*C. elegans*	16 kDa
	Sponges	*Geodia cydonium*	13-16 kDa
Chimera	Mammals	Man, mouse, rat	Galectin-3
Tandem-repeat	Mammals	Rat, mouse	Galectins-4, -6, -8
	Nematodes	*C. elegans*	32 kDa

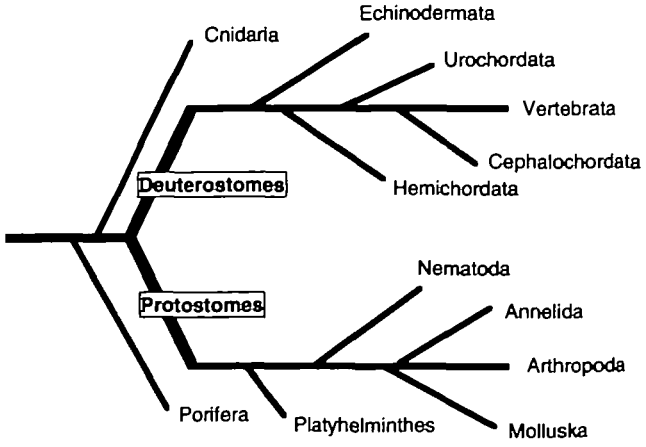

Fig. 2. **A phylogenic tree of metazoa.** Galectins have been found in porifera, nematoda, and vertebrata.

though at present no galectin has been found in other than vertebrates, nematodes, and sponges.

Since the highest deuterostomes, vertebrates, still require galectins, lower deuterostomes such as the chordata and echinodermata are expected to preserve galectins. Even if they have ceased to use galectins for some reasons, it is very possible that genetic information of galectins still remains as either silent genes or pseudogenes elsewhere. It also seems rather strange that no galectin has been found in the arthropoda, the biggest and mostly adapted group of protostomes. If they have abandoned galectins, many questions arise; why? when? and what are the alternatives? We have been trying continuously to find galectins in these animals, but have not yet succeeded. Insect galectins, if they exist, are attractive because they will enable us to examine insects of the Jurassic era trapped in amber.

2. Biological significance of the galectin family

As mentioned in the previous section, significant progress has been made on the molecular properties of the galectin family. However, when we wish to understand their biological meaning, we inevitably encounter a number of problems difficult to solve on the basis of the present knowledge of biological sciences. In this section, we will discuss these problems. Our efforts to provide appropriate solutions for all of these problems will make it possible to understand galectins and also unknown principles of life.

2.1. Biological functions of the galectin family are difficult to summarize. The molecular function of galectins is not so complicated; they bind sugar chains containing galactoside. Their biological roles, however, cannot be easily specified because they appear to correspond to too many diverse biological processes. It is not possible to assume a one-to-one relationship between a galectin and a biological function. However, if we ignore our preconception and take a different viewpoint, such a difficulty may become one of the clues for understanding the galectins.

In vertebrates, galectins were found in a variety of tissues and cells; *e.g.*, skin, muscle, brain, intestine, liver, kidney, placenta, cultured fibroblast, established cell lines such as HL60, and many tumor cells. A number of observations have indicated their possible involvement in a variety of important phenomena occurring in multicellular animals; *i.e.*, development, differentiation, morphogenesis, immunity, apoptosis, *etc.* Metastasis of malignant cells was also reported to be related to the presence of galectins on the cell surface (*35*). The biological functions of galectins vary from site to site, and from time to time. The too many cases in which the involvement of galectins is supposed make it almost impossible to define their principle mission. Although a mammalian species probably has at least 8 galectins, how many different biological roles each of them has is almost completely unknown.

As an example of this complexity, the behavior of galectins during the morphogenesis of chick embryonic skin is briefly described below. Since they are expressed significantly at a certain stage of tissue development, they have often been called developmentally regulated lectins. We studied extensively the expression and fate of two chicken isolectins (14 and 16 kDa), mainly by means of immunohistochemical procedure (*36-39*). Expression of these two isolectins was found to be regulated by a complicated program, and their roles seemed to differ from tissue to tissue. For example, 16-kDa galectin was localized mainly on the basement membrane at earlier stages, and then was abundantly found in connective tissue of the dermis. Sixteen-kilodalton galectin molecules seem to stabilize the structure of the basement membrane and connective tissue. On the other hand, 14-kDa galectin appeared at later stages very abundantly in the intercellular space between closely packed epidermal cells, which were fully differentiated and had begun to synthesize keratin. Since 14-kDa galectin molecules are found especially abundantly around desmosomes, they seem to stabilize cell-to-cell contact. These results suggest that galectins play a variety of roles depending on the time and location. It is interesting that 32-kDa galectin of *C. elegans* was also found abundantly in the cuticle (*34*). This suggests that the function of galectins as components of the outer barrier of animal body has been conserved in both vertebrates and nematodes, though the cuticle of *C. elegans* and chick skin are very different in both composition and structure.

From a mechanical viewpoint, various cellular processes, such as communication, adhesion, transportation, guidance, modulation, signal transduction, *etc.*, seem to be supported by the specific recognition of glycoconjugates by galectins. However, it is always difficult to obtain clear-cut experimental evidence that galectins are the major factor in spite of enough lines of circumstantial evidence.

Gene knock-out is a powerful and promising technique to prove the function of a certain protein. However, it was not necessarily easy to reach a definite conclusion. Poirier *et al.* succeeded in producing knock-out mice in which the gene for galectin-1 was inactivated. However, no apparent damage was observed and the mice were even able to produce offspring. The functions assigned to galectin-1 seemed to have been taken over by one of other galectins, such as galectin-3 (*40*). Therefore, knock-out of multiple genes for the galectin family will be needed to obtain conclusive evidence.

2.2. The galectin family is rather small and less diversified in comparison with C-type lectins. The galectin family is very different in many respects from another principal family of animal lectins, the C-type lectin family. (i) The galectin family is a relatively small family consisting of much less members than the C-type lectin family, to which more than 50 members are known to belong. (ii) For most C-type lectins, well-defined biological roles have been assigned. However, it is always difficult to assign a definite role to any galectin. (iii) The sugar-binding specificities of all galectins reported so far are very similar. No galectin specific for sugars other than galactosides has been found. On the other hand, the C-type lectin family is composed of lectins with a variety of specificities. What is the reason for such a monotonous specificity of the galectin family? (iv) Galectins are designed and biosynthesized as intracellular proteins, while C-type lectins are designed to function extracellularly. These differences suggest the existence of fundamental differences in their origins and biological roles. Consideration of the possible reason will facilitate understanding of the biological significance of the galectins.

A considerable number of C-type lectins were firstly regarded as functional proteins other than lectins. Usually, it was after determination of their primary structures that they were found to be C-type lectins. Therefore, their

biological functions are clear and they are often given names related to their functions, *e.g.*, asialoglycoprotein receptor, thrombomodulin, surfactant protein, *etc.* Each C-type lectin is specialized for one or limited tasks. On the other hand, galectins are much less specialized and thus cannot be given names related to functions.

Such a consideration leads to the assumption that galectins are in charge of general and household tasks, while C-type lectins have diverse specialized tasks. For appropriate operation of the living system, in addition to bureaucratic specialists, flexible generalists seem to be essential. This may explain why galectins are more conservative, in that their sugar-binding specificity has not been diversified and the number of family members has been suppressed. It is possible that a change in specificity has not been allowed during molecular evolution. On the other hand, C-type lectins are more radical: The family has expanded continuously in response to the increasing number of new tasks, by combining the C-type CRD with certain functional motifs of other proteins and also by changing their sugar-binding specificities.

2.3. Controversial tissue and cellular localization of galectins. It is difficult to generalize the location of galectins in the animal body because it depends on various factors such as time, tissue, and circumstance. In vertebrates, galectins appear in a variety of tissues and cells at programmed times. Galectins are usually not in a free state but bound noncovalently to galactoside-containing sugar chains of water-insoluble substances of tissues or cells. Lines of immunohistochemical evidence have shown a mysterious feature of the localization of galectins (*36-39*). They have been found at a variety of sites, both inside and outside cells. In the former case, both the cytoplasm and nucleus are sites of localization, though nothing is known about endogenous intracellular ligands. In the latter case, they are either attached to the cell surface or localized in the intercellular spaces between closely packed cells. They are also found in the extracellular matrix in the case of connective tissues.

Since almost all known glycoconjugates are localized extracellularly, it is not easy to understand why proteins supposed to interact with glycoconjugates are found intracellularly. However, from the viewpoint of protein structure, there are lines of evidence showing that galectins are designed as proteins supposed to play roles inside cells; *e.g.*, acetylated N-terminus (*10, 12*), cysteine residues fully in free states (*10, 12*), lack of signal sequences (*9*), and biosynthesis occurred on free ribosomes (*16*), *etc.* This suggests that their roles at the time of their first appearance should have been within cells, and that their extracellular functions were acquired in the course of the development of multicellular metazoan systems. Therefore, elucidation of their intracellular functions is very important. However, little research has been done from such a viewpoint except for that on the possible function of galectin-3 as one of the components of the splisosomes in the nucleus, which carries out splicing of precursors of mRNA (*41*).

Extracellular galectins do not seem mere cellular waste. They should be secreted on purpose. Though the process of externalization of galectin-1 from muscle cells was studied extensively (*42*), the detailed molecular mechanism, without signal sequences, remains to be solved. Another problem is that galectins are generally inactivated if certain SH groups are oxidized. If they remain within cells, such a risk is not high. However, when they are externalized, the risk of oxidative inactivation becomes high. Do galectins abundantly localized extracellularly retain sugar-binding activity? If they are already inactivated, what is the significance of their presence? A report has appeared suggesting that inactivated galectin-1 acquired TGF activity (*43*). Vertebrate galectins might have dual potential functions depending on the oxidative state, switching from one to another in different situations.

Recently, another protein family was found to be similar to the galectin family in some respects, though they are structurally unrelated. Annexins, which had been considered to be phospholipid-binding proteins, turned out also to have glycosaminoglycan-binding activity, and came to be regarded as a new lectin family (*44*). It is possible that they also serve as bifunctional cross-linkers between biomolecules. Their protein nature suggests that they are also designed as intracellular proteins like galectins; *i.e.*, lack of signal sequences, blocked N-terminus, lack of sugar chains, *etc.* Nevertheless, they are also found to be distributed extracellularly. It seems meaningful that these two sugar-binding protein families exhibit similar behavior.

2.4. *N*-Acetyllactosamine is the key glycocode for galectins and the specificity of galectins has been conserved during molecular evolution. Now, we consider endogenous ligands of galectins. As to extracellular ligands, there are too many candidates, such as glycoproteins, glycolipids, and proteoglycans. This makes the situation rather complicated, and confirmation of true ligands difficult. On the other hand, information on intracellular glycoconjugates is poor except for proteins having *O*-linked *N*-acetylglucosamine (*45*). Since no cytoplasmic or nuclear glycoconjugate having galactoside has been identified, nothing is known about intracellular ligands for galectins.

Though galectins bind galactose *in vitro*, their endogenous ligands are not free monosaccharides but sugar chains attached to glycoconjugates. The key structure recognized by galectins is the disaccharide unit, *N*-acetyllactosamine (Galβ1-4GlcNAc, LacNAc) (*46-49*). The LacNAc structure is included in a variety of glycoconjugates, and also serves as a backbone for various glycosignals such as blood group determinants and Lews antigens. Polymerized LacNAc structures, polylactosamine chains, are also often found in glycoconjugates, such as fibronectin, laminin, *etc.* Therefore, the LacNAc unit seems to be one of the important key words in glycobiology though its significance is poorly understood. Galectins are likely to decipher this key word.

The mechanism underlying the specific interaction of galectins and LacNAc (or lactose) has been elucidated by means of various experiments, such as binding experiments (*46-49*), site-directed mutagenesis (*27, 50, 51*), and X-ray crystallography (*52-54*). The essential residues revealed by the site-directed mutagenesis experiments (*27, 50*) (indicated by large capital letters in Fig. 1) were found to form hydrogen bonds with principal hydroxyl groups in lactose; 4-OH of Gal with His45, Asn47, and Arg49, 6-OH of Gal with Asn58 and Glu68, and 3-OH of Glc with Arg49, Glu68, and Arg70.

Though the affinities of galectins for various galactoside-containing oligosaccharides slightly differ for each other,

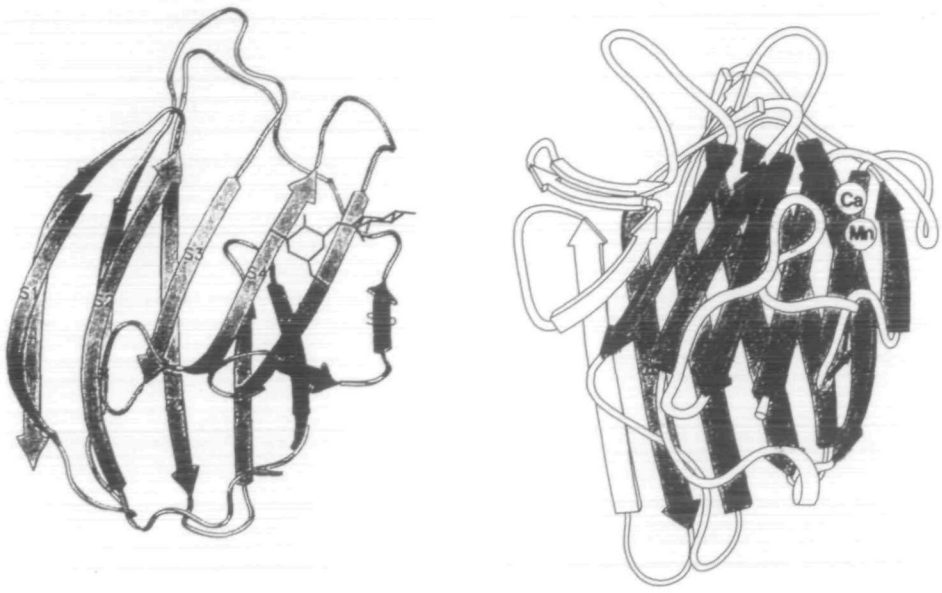

Fig 3. **Comparison of the topologies of human galectin-2 (left, from Ref. *38* with permission) and concanavalin-A (right, modified from a figure drawn by Richardson in *Advances in Protein Chemistry*, Vol. 34, 1981).** β-Strands in concanavalin A corresponding to those in galectin-2 are shaded. A bound lactose molecule is shown in galectin-2, and the positions of metal ions (Ca and Mn) are indicated in concanavalin A.

the fundamental specificity for LacNAc seems to have been preserved since the appearance of the ancestor galectin. This suggests that the key word, LacNAc, was also created at the very beginning of life on earth, and that the deciphering or decoding of this glycocode by the galectin family has been continuously one of the most important processes in living organisms.

The interaction of galectins with poly LacNAc was firstly shown by our affinity labeling experiments (*55*), and now is attracting the attention of many researchers. It is well known that the arrangement of the LacNAc unit changes in the course of mammalian development. For example, poly LacNAc chains on the fetal erythrocyte surface (i antigen) are linear while those on the adult erythrocyte surface (I antigen) are branched. The biological meaning of such a change remains unexplained, but it undoubtedly alters the nature of the interaction between galectins and the LacNAc unit.

3. Are galectins relatives of legume lectins?

Recent crystallographic analysis revealed that galectins have the so-called jellyroll topology, *i.e.*, a sandwich composed of two β-sheets (Fig. 3) (*52-54*). This is very similar to that of legume lectins such as concanavalin A (Con A) in spite of the absence of similarity in the primary structures (Fig. 3). The spatial arrangement of subunits and also the relative positions of the two sugar-binding sites in the dimeric molecule are also close. Some crystallographers have supposed that the number of topologies of existing proteins is not infinite but less than one thousand. If this is the case, the topological resemblance between the sugar-binding proteins of legumes and animals might be the result of convergent molecular evolution, though this is not fully persuasive. It seems more plausible that an ancestor protein related to sugar molecules somehow, having the jellyroll topology, existed on the primitive earth before the divergence of plants and animals. Even after the divergence, this topology was conserved, and molecular evolution gave rise to legume lectins in plants and galectins in animals, though the accumulation of point mutations makes it impossible to find sequence similarities today.

The search for proteins linking legume lectins and galectins will need a new strategy because the homology in primary structure is too low, and the only clue is their topology. However, recent observations, shown below, are signs that the galectin family is now growing into the galectin super family. The Charcot-Leyden crystal protein, which exhibits phospholipase activity but no sugar-binding activity, was found to exhibit sequence homology to galectins (*56*). Crystallography revealed that pentraxin, a serum amyloid P-component, has a similar topology to galectins (*57*). Very recently, galectins were found in some species of fungi (Cooper, D., personal communication).

4. Seven wonders of the galectin family

As discussed so far, we are confronted with a number of problems, that are difficult to solve at present, in the course of galectin research, which are again listed below. All of them are worth attention because they certainly hide important clues for the understanding of the biological meaning of sugars.

1) What is the principal mission of galectins?
2) Why are they found extracellularly in spite of that they were designed as intracellular proteins?
3) How are they externalized from cells?
4) Why are all galectins galactoside-specific?
5) What are their endogenous extracellular and intracellular ligands?
6) Why are they localized at sites where the risk of oxidative inactivation is high?
7) Why do galectins and plant lectins have a common topology?

Concluding remarks: Glycocode and galectins

The recognition of sugar chains is completely different from that of nucleic acids and proteins, and likely to be related to unknown principles. It also seems to have greater significance for multicellular organisms than unicellular organisms. At such a level, the situation in man cannot be explained on the basis of knowledge obtained for *Escheri-*

chia coli. As informational biomolecules, sugar chains are peculiar. For example, why are sugar chains allowed to escape from direct control of genes and be heterogeneous? Why do glycoproteins accept sugar chains heterogeneous in both size and structure, in spite of that the protein parts are synthesized with the greatest care and strictness?

When unicellular organisms appeared on the earth, nucleic acids and proteins should already have had fundamental structures and roles. On the other hand, sugar chains, which would have been present within cells, should have been less important as informational molecules. When multicellular organisms were generated, communication between component cells should have become much more important. The unique nature of sugar chains may have had certain advantages for such new purposes.

What is the uniqueness of sugar chains as informational molecules? Since the structures of sugar chains are not directly controlled by genes, the properties of glycocodes are almost opposite to the codes written with nucleotides and amino acids. For example, rough *vs.* precise, fussy *vs.* distinct, analogue *vs.* digital, heterogeneous *vs.* homogeneous, guerrilla *vs.* regular army, ambiguous *vs.* clear, and harmonious *vs.* melodious. It can be compared to the contrast between Japanese culture, which is characterized by ambiguity and compromise, and European culture, which always demands clear-cut answers. The way glycocodes are composed is also different; it is based on the principle of the generation of Chinese characters. In sugar chains, monosaccharides are arranged two-dimensionally, *e.g.*, by various linkages such as 1-3, 1-4, 1-6, *etc.*, while nucleic acids and polypeptides are composed of monomers connected uni-dimensionally as in words written with the alphabet.

Why did living organisms adopt such a fuzzy informational system and still use it? If this system was ineffective, it should have been discarded. It should have been impossible for life on earth to support such remarkable evolution by means of only precise but stiff informational systems conducted by nucleic acids and proteins, and another fuzzy but extremely flexible information system involving sugars should have been essential. Elucidation of the biological meaning of the interaction between the glycocode, LacNAc, and household equipment galectins, which have lasted more than a billion years, should make a great contribution to glycobiology. *C. elegans* should be one of the most effective experimental tools in this field because it enables us to analyze an animal from a single fertilized egg to a multicellular adult using a whole body.

REFERENCES

1. Kasai, K. (1990) Biochemical properties of vertebrate 14K beta-galactoside-binding lectins, in *Advances in Lectin Research* (Franz, H., Kasai, K., Kocourek, J., Olsnes, S., and Shannon, L.M., eds.) Vol. 3, pp. 10-35, VEB Verlag Volk und Gesundheit, Berlin
2. Hirabayashi, J. (1993) A general comparison of two major families of animal lectins. *Trends Glycobiol. Glycotechnol.* **5**, 251-270
3. Hirabayashi, J. and Kasai, K. (1993) The family of metazoan metal-independent β-galactoside-binding lectins; structure, function and molecular evolution. *Glycobiology* **3**, 297-304
4. Caron, M., Bladier, D., and Joubert, R. (1990) Soluble galactoside-binding vertebrate lectins: A protein family with common properties. *Int. J. Biochem.* **22**, 1379-1385
5. Barondes, S.H., Cooper, D.N.W., Gitt, M.A., and Leffler, H. (1994) Galectins. Structure and function of a large family of animal lectins. *J. Biol. Chem.* **269**, 20807-20810
6. Barondes, S.H., Castronovo, V., Cooper, D.W., Cummings, R.D., Drickamer, K., Feizi, T., Gitt, M.A., Hirabayashi, J., Hughes, C., Kasai, K., Leffler, H., Liu, F.-T., Lotan, R., Mercurio, A.M., Monsigny, M., Pillai, S., Poirier, F., Raz, A., Rigby, P.W., Rini, J.M., and Wang, J.L., (1994) Galectins: A family of anima β-galactoside-binding lectins. *Cell* **76**, 597-598
7. Drickamer, K. (1988) Two distinct classes of carbohydrate-recognition domain in animal lectins. *J. Biol. Chem.* **263**, 9557-9560
8. Levi, G. and Teichberg, V.I. (1981) Isolation and physicochemical characterization of electrolectin, a β-galactoside binding lectin from the electric organ of *Electrophorus electricus*. *J. Biol. Chem.* **256**, 5735-5740
9. Ohyama, Y., Hirabayashi, J., Oda, Y., Oono, S., Kawasaki, H., Suzuki, K., and Kasai, K. (1986) Nucleotide sequence of chick 14K β-galactoside-binding lectin mRNA. *Biochem. Biophys. Res. Commun.* **134**, 51-56
10. Hirabayashi, J., Kawasaki, H., Suzuki, K., and Kasai, K. (1987) Complete amino acid sequence of 14 kDa β-galactoside-binding lectin of chick embryo. *J. Biochem.* **101**, 775-787
11. Hirabayashi, J. and Kasai, K. (1984) Human placenta β-galactoside-binding lectin. Purification and some properties. *Biochem. Biophys. Res. Commun.* **122**, 938-944
12. Hirabayashi, J. and Kasai, K. (1988) Complete amino acid sequence of a β-galactoside-binding lectin from human placenta. *J. Biochem.* **104**, 1-4
13. Hirabayashi, J. and Kasai, K. (1989) Cloning and nucleotide sequence of a full-length cDNA for human 14-kDa β-galactoside-binding lectin. *Biochim. Biophys. Acta* **1008**, 85-91
14. Abbott, W.M., Mellor, A., Edwards, Y., and Feizi, T. (1989) Soluble bovine galactoside-binding lectin cDNA reveals the complete amino acid sequence and an antigenic relationship with the major encephalitogenic domain of myelin basic protein. *Biochem. J.* **259**, 283-290
15. Clerch, L.B., Whitney, P., Hass, M., Brew, K., Miller, T., Werner, R., and Massaro, D. (1988) Sequence of a full-length cDNA for rat lung β-galactoside-binding protein: Primary and secondary structure of the lectin. *Biochemistry* **27**, 692-699
16. Wilson, T.J.G., Firth, M.N., Powell, J.T., and Harrison, F.L. (1989) The sequence of the mouse 14 kDa β-galactoside-binding lectin and evidence for its synthesis on free cytoplasmic ribosomes. *Biochem. J.* **261**, 847-852
17. Sakakura, Y., Hirabayashi, J., Oda, Y., Ohyama, Y., and Kasai, K. (1990) Structure of chicken 16-kDa β-galactoside-binding lectin: Complete amino acid sequence, cloning of cDNA, and production. *J. Biol. Chem.* **265**, 21573-21579
18. Hirabayashi, J., Ayaki, H., Soma, G., and Kasai, K. (1989) Production and purification of a recombinant human 14-kDa β-galactoside-binding lectin. *FEBS Lett.* **250**, 161-165
19. Ohyama, Y. and Kasai, K. (1988) Isolation and characterization of the chick 14K β-galactoside-binding lectin gene. *J. Biochem.* **104**, 173-177
20. Jia, S. and Wang, J.L. (1988) Carbohydrate binding protein 35: Complementary DNA sequence reveals homology with proteins of the heterogeneous nuclear RNP. *J. Biol. Chem.* **263**, 6009-6011
21. Albrandt, K., Orida, N.K., and Liu, F.-T. (1987) An IgE-binding protein with a distinctive repetitive sequence and homology with an IgG receptor. *Proc. Natl. Acad. Sci. USA* **84**, 6859-6863
22. Cherayil, B.J., Weiner, S.J., and Pillai, S. (1989) The Mac-2 antigen is a galactose-specific lectin that binds IgE. *J. Exp. Med.* **170**, 1959-1972
23. Hirabayashi, J., Satoh, M., and Kasai, K. (1992) Evidence that *Caenorhabditis elegans* 32-kDa β-galactoside-binding protein is homologous to vertebrate β-galactoside-binding lectins. *J. Biol. Chem.* **267**, 15485-15490
24. Oda, Y., Herrmann, J., Gitt, M.A., Turck, C.W., Burlingame, A.L., Barondes, S.H., and Leffler, H. (1993) Soluble lactose-

binding lectin from rat intestine with two different carbohydrate-binding domains in the peptide chain. *J. Biol. Chem.* **268**, 5929-5939
25. Hirabayashi, J., Ubukata, M., and Kasai, K. Purification and molecular characterization of a novel 16-kDa galectin from the nematode *Caenorhabditis elegans*. *J. Biol Chem.* (in press)
26. Pleifer, K., Hassemann, M., Gamulin, V., Bretting, H., Fahrenholz, F., and Muller, W.E.G. (1993) S-type lectins occur also in invertebrates: High conservation of the carbohydrate recognition domain in the lectin genes from the marine sponge *Geodia cydonium*. *Glycobiology* **3**, 179-184
27. Hirabayashi, J. and Kasai, K. (1991) Effect of amino acid substitution by site-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa β-galactoside-binding lectin. *J. Biol. Chem.* **266**, 23648-23653
28. Tracey, B.M., Feizi, T., Abbott, W.M., Carruthers, R.A., Green, B.N., and Lawson, A.M. (1992) Subunit molecular mass assignment of 14,654 Da to the soluble β-galactoside-binding lectin from bovine heart muscle and demonstration of intramolecular disulfide bonding associated with oxidative inactivation. *J. Biol. Chem.* **267**, 10342-10347
29. Gitt, M.A., Massa, S.M., Leffler, H., and Barondes, S.H. (1992) Isolation and expression of a gene encoding L-14-II, a new human soluble lactose-binding lectin. *J. Biol. Chem.* **267**, 10601-10606
30. Gitt, A.M., Wiser, M.F., Leffler, H., Herrmann, J., Xia, Y.-R., Massa, S.M., Cooper, D.N.W., Lusis, A.J., and Barondes, S.H. (1995) Sequence and mapping of galectin-5, a β-galactoside-binding lectin, found in rat erythrocytes. *J. Biol. Chem.* **270**, 5032-5038
31. Magnaldo, T., Bernerd, F., and Darmon, M. (1995) Galectin-7, a human 14-kDa S-lectin, specifically expressed in keratinocytes and sensitive to retinoic acid. *Dev. Biol.* **168**, 259-271
32. Oda, Y., Leffler, H., Sakakura, Y., Kasai, K., and Barondes, S.H. (1991) Human breast carcinoma cDNA encoding a galactose-binding lectin homologous to mouse Mac-2 antigen. *Gene* **99**, 279-283
33. Hadari, Y.R., Paz, K., Dekel, R., Mestrovic, T., Accili, D., and Zick, Y. (1995) Galectin-8. A new rat lectin, related to galectin-4. *J. Biol. Chem.* **270**, 3447-3454
34. Arata, Y., Hirabayashi, J., Akimoto, Y., Hirano, H., and Kasai, K. An immunohistochemical study of the 32-kDa galectin (β-galactoside-binding lectin) in the nematode *Caenorhabditis elegans*. *Histochem. J.* (in press)
35. Raz, A. and Lotan, R. (1987) Endogenous galactoside-binding lectins: a new class of functional tumor cell surface molecules related to metastasis. *Cancer Metast. Rev.* **6**, 433-452
36. Oda, Y., Ohyama, Y., Obinata, A., Endo, H., and Kasai, K. (1989) Endogenous β-galactoside-binding lectin expression is suppressed in retinol-induced mucous metaplasia of chick embryonic epidermis. *Exp. Cell Res.* **182**, 33-43
37. Akimoto, Y., Kawakami, H., Oda, Y., Obinata, A., Endo, H., Kasai, K., and Hirano, H. (1992) Changes in expression of the endogeneous β-galactoside-binding lectin of chick embryonic skin during epidermal differentiation. *Exp. Cell Res.* **199**, 3-18
38. Akimoto, Y., Obinata, A., Hirabayashi, J., Sakakura, Y., Endo, H., Kasai, K., and Hirano, H. (1993) Secretion of endogenous 16-kDa β-galactoside-binding lectin from vitamin A-pretreated chick embryonic cultured skin. *Exp. Cell Res.* **205**, 251-260
39. Akimoto, Y., Obinata, A., Hirabayashi, J., Sakakura, Y., Endo, H., Kasai, K., and Hirano, H. (1995) Changes in expression of two endogenous β-galactoside-binding isolectins in the dermis of chick embryonic skin during development *in ovo* and *in vitro*. *Cell Tissue Res.* **279**, 3-12
40. Poirrier, F. and Robertson, E.J. (1993) Normal development of mice carrying a null mutation in the gene encoding the L14 S-type lectin. *Development* **119**, 1229-1236
41. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995) Identification of galectin-3 as a factor in pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* **92**, 1213-1217
42. Cooper, D.N.W. and Barondes, S.H. (1990) Evidence for export of a muscle lectin from cytosol to extracellular matrix and for novel secretory mechanism. *J. Cell Biol.* **110**, 1681-1691
43. Yamaoka, K., Ohno, S., Kawasaki, H., and Suzuki, K. (1991) Overexpression of a β-galactoside binding protein causes transformation of BALB3T3 fibroblast cells. *Biochem. Biophys. Res. Commun.* **179**, 272-279
44. Kojima, K., Ogawa, H., Seno, N., Yamamoto, K., Irimura, T., Osawa, T., and Matsumoto, I. (1992) Carbohydrate-binding proteins in bovine kidney have consensus amino-acid sequences of annexin family proteins. *J. Biol. Chem.*, **267**, 20536-20539
45. Hart, G.W., Haltiwanger, R.S., Holt, G.D., and Kelly, W.G. (1989) Glycosylation in the nucleus and cytoplasm. *Annu. Rev. Biochem.* **58**, 841-874
46. Abbott, W.M., Hounsell, E.F., and Feizi, T. (1988) Further studies of oligosaccharide recognition by the soluble 13-kDa lectin of bovine heart muscle. *Biochem. J.* **252**, 283-287
47. Leffler, H. and Barondes, S.H. (1986) Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian β-galactosides. Sugar binding specificity. *J. Biol. Chem.* **261**, 10119-10126
48. Lee, R.T., Ichikawa, Y., Allen, H.J., and Lee, C.Y. (1990) Binding characteristics of galactoside-binding lectin (galaptin) from human spleen. *J. Biol. Chem.* **265**, 7868-7871
49. Kakita, H., Nakamura, K., Kato, Y., Oda, Y., Shimura, K., and Kasai, K. (1991) High-performance affinity chromatography of a chick lectin on an adsorbent based on hydrophilic polymer gel. *J. Chromatogr.* **543**, 315-326
50. Hirabayashi, J. and Kasai, K. (1994) Further evidence by site-directed mutagenesis that conserved hydrophilic residues form a carbohydrate-binding site of human galectin-1. *Glycoconjugate J.* **11**, 437-442
51. Abbott, M.W. and Feizi, T. (1991) Soluble 14-kDa β-galactoside-specific bovine lectin: Evidence from mutagenesis and proteolysis that almost the complete polypeptide chain is necessary for integrity of the carbohydrate recognition domain. *J. Biol. Chem.* **266**, 5552-5557
52. Lobsanof, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., and Rini, J.M. (1993) X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9 Å resolution. *J. Biol. Chem.* **268**, 27034-27038
53. Bourne, Y., Bolgiano, B., Liao, D.-I., Strecker, G., Cantau, P., Herzberg, O., Feizi, T., and Cambillau, C. (1994) *Struct. Biol.* **1**, 863-870
54. Liao, D.-I., Kapadia, G., Ahmed, H., Vasta, G.R., and Herzberg, O. (1994) Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. *Proc. Natl. Acad. Sci. USA* **91**, 1428-1432
55. Oda, Y. and Kasai, K. (1984) Photochemical cross-linking of β-galactoside-binding lectin to polylactosamino-proteoglycan of chick embryo skin. *Biochem. Biophys. Res. Commun.* **123**, 1215-1220
56. Ackerman, S.J., Corrette, S.E., Rosenberg, H.F., Bennett, J.C., Mastrianni, D.M., Nicholson-Weller, A., Weller, P.F., Chin, D.T., and Tenen, D.G. (1993) Molecular cloning and characterization of human eosinophil Charcot-Leyden crystal protein (lysophospholipase). *J. Immunol.* **150**, 456-468
57. Emsley, J., White, H.E., O'Hara, B.P., Oliva, G., Srinivasan, N., Tickle, I.J., Blundell, T.L., Pepys, M.B., and Wood, S.P. (1994) Structure of pentameric human serum amyloid P component. *Nature* **367**, 338-345