Glycoconjugate Journal 19, 451–458, 2004

The speciation of conger eel galectins by rapid adaptive evolution

Tomohisa Ogawa[1], Tsuyoshi Shirai[2], Clara Shionyu-Mitsuyama[2], Takashi Yamane[3], Hisao Kamiya[4] and Koji Muramoto[1]

[1]*Department of Biomolecular Science, Graduate School of Life Sciences, Tohoku University, Sendai 981-8555,* [2]*Department of Computational Biology, Biomolecular Engineering Research Institute, Osaka 565-0874,* [3]*Department of Biotechnology and Biomaterial Chemistry, Graduate School of Engineering, Nagoya University, Nagoya 464-8603,* [4]*School of Fisheries Sciences, Kitasato University, Iwate 022-0101, Japan*

Many cases of accelerated evolution driven by positive Darwinian selection are identified in the genes of venomous and reproductive proteins. This evolutional phenomenon might have important consequences in their gene-products' functions, such as multiple specific toxins for quick immobilization of the prey and the establishment of barriers to fertilization that might lead to speciation, and in the molecular evolution of novel genes. Recently, we analyzed the molecular evolution of two galectins isolated from the skin mucus of conger eel (*Conger myriaster*), named congerins I and II, by cDNA cloning and X-ray structural analysis, and we found that they have evolved in the rapid adaptive manner to emergence of a new structure including strand-swapping and a unique new ligand-binding site. In this review article we summarize and discuss the molecular evolution, especially the rapid adaptive evolution, and the structure-function relationships of conger eel galectins.
Published in 2004.

Keywords: **adaptive evolution, conger eel, domain swapping, galectin, X-ray crystal structure**

Abbreviations: **CRD(s), carbohydrate recognition domain(s); LPS, lipopolysaccharide; MES, 2-*N*-morpholinoethane sulfonic acid; UTR(s), untranslated region(s).**

Introduction

The diversity of life on the earth is the result of perpetual evolutionary processes; evolution is the fundamental development strategy of life. Today, we are gaining insight into these processes and events *via* analysis of gene and protein sequences; some of the mechanisms for evolution of new genes have been clarified, for example, gene duplications, exon shuffling and alternative recombination.

Recently, rapid adaptive evolution, which is characterized by the higher mutation rate of non-synonymous nucleotides (causing amino acid change) to synonymous ones (not causing amino acid change) has been identified in several gene families. For example, animal toxins comprise a diverse array of proteins that show similar molecular architectures but different biological and pharmacological functions. The molecular evolution of animal toxins has been studied in some multigene families from invertebrates and vertebrates. In invertebrates, rapid evolution of conotoxins [1–4] and scorpion toxins [5] has been observed. In venomous vertebrates, snake venom proteinatious toxins, which contain several types of toxin scaffold and act specifically on target molecules in different physiological systems of prey animals, have evolved in a rapid adaptive manner. Comparison and molecular evolutionary analysis of phospholipase A_2 multigene families, which possess diverse physiological functions such as neurotoxicity, myonecrotic activity, cardiotoxicity, anticoagulant activity, hypotensive activity and edema-inducing activity, show that the accelerated evolution occurred in the coding region except for signal peptide [6–12]. The same evolutionary phenomena, accelerated evolution, have also been observed in other snake venom multigene families such as serine proteases [13], metalloproteases, three finger type neurotoxins [14,15] and C-type lectin like proteins [16–18] and also in the reproductive proteins such as female zona pellucida glycoproteins and acrosin [19]. Despite

To whom correspondence should be addressed: Tomohisa Ogawa, Ph.D., Department of Biomolecular Science, Graduate School of Life Sciences, Tohoku University, 1-1 Tsutsumidoriamamiyamachi, Aoba-ku, Sendai 981-8555, Japan. Tel.: +81-22-717-8808; Fax: +81-22-717-8807; E-mail: ogawa@biochem.tohoku.ac.jp

the many examples of adaptive evolution, the molecular mechanisms and selective pressures causing them still remain unclear. More recently, we found that conger eel galectins, named congerins I and II, also evolved in a rapid adaptive manner to generate a new structure and unique carbohydrate binding activity. In this review, we focus on the rapid adaptive evolution of conger eel galectins.

Galectins are a family of carbohydrate-binding proteins defined by their affinity for ß-galactoside sugars and sharing a conserved sequence motif within their carbohydrate recognition domain (CRD). To date, the primary structures of more than 130 galectins were determined, and there are now at least 14 subfamilies (galectin-1 to 14) in mammals. They are structurally classified into three types: proto-, chimera-, and tandem-repeat types [20]. Proto-type galectin forms a non-covalent dimer consisting of two identical CRDs. On the other hand, the chimera-type has two distinct domains consisting of an N-terminal collagenous domain and C-terminal galectin CRD. The tandem repeat-type has two homologous but different sugar specific CRDs on a single polypeptide chain. Although the fundamental galectin function is the specific recognition of glycoconjugates at the molecular level, galectins have been proposed to participate in diverse physiological functions such as development, differentiation, morphogenesis, immunity, apoptosis, metastasis of malignant cells, etc. [21]. Recently, it was reported that oxidized galectin-1, which had no carbohydrate binding activity, acted as a biomodulator for nerve cell growth [22]. The multiple and diverse physiological activities and structures of galectins make their structure-activity relationships and their evolutionary histories of great interest.

Molecular properties of conger eel galectins

Two galectins named congerins I and II were isolated from the skin mucus of conger eel (*Conger myriaster*) [23]. Congerins I and II are dimers composed of two identical subunits of 136 and 135 amino acid residues, respectively. They belong to proto-type galectins [24,25], and show several common characteristics among all known galectins: a blocked (acetylated) N-terminus and lack of a secretion signal peptide region, and attached carbohydrate. However, congerins I and II have no cysteine residue that is related to oxidizing inactivation found in some galectins of higher vertebrates. Although congerins I and II show 48% identity in amino acid sequence, they have different stabilities against pH and temperature [25]; congerin I was still stable at 50°C for at least 80 min in both structure and hemagglutinating activity, whereas congerin II completely lost its activity under the same condition. Thus, the thermostability of congerin I is higher than that of congerin II. Furthermore, binding experiments of congerins using the pyridylamino derivative of lactose showed that congerin I had much higher affinity for lactose than congerin II and they had different specificities for C_3-OH in the glucose moiety of lactose derivatives [25].

Skin mucus is an indispensable chemical and physical barrier in the self-defense system of fish against the invasion of pathogenic bacteria and parasites. In fact, the mucus contains many biologically active molecules such as hemolysin and lysozyme. Lectins and lectin-like molecules have been found in skin mucus of fish as reviewed by Ingram [26], and they may participate in innate or acquired immunity through their agglutinating activity. In the case for congerins I and II, they can recognize some marine bacteria such as *Vibrio anguillarum* [23]. Investigation of their localization in fish tissues by northern blot analysis [27] and immunohistochemical studies [28] suggested that they are expressed not only in skin but also in the upper digestive tract and gill filament. Congerins are produced and secreted into mucus by the club cells in the mucosal epithelium lining the skin and digestive tract [28]. These observations suggest that congerins participate in the self-defense system, including innate immunity, on the intra- and extra-body surface of conger eel.

cDNA cloning and evolutional analysis of congerins and other galectins

The cDNAs encoding congerins I and II were cloned and their nucleotide sequences were determined (Figure 1) [27]. These cDNAs contain the open reading frames of 411 and 408 nucleotides coding for 137 and 136 amino acid residues, respectively. The nucleotide sequences of these cDNAs revealed some interesting findings from the viewpoint of evolution. First, the 5′ and 3′ untranslated regions (UTRs) are more homologous than the protein-coding region, with nucleotide identities of 86% for 5′UTR, 73% for the protein-coding region, and 88% for 3′UTR. Usually, lower homology is observed in UTRs of cDNAs and genes. Second, base substitutions at nonsynonymous sites occurred more frequently than at synonymous sites, causing amino acid changes. To better understand the evolutionary phenomena of congerins, the number of nucleotide substitution per site (K_N) for the non-coding regions, and the numbers of nucleotide substitutions per synonymous site (K_S) (not causing amino acid changes) and per nonsynonymous site (K_A) (causing amino acid changes) for the protein-coding regions were computed. Usually, nonsynonymous substitutions in coding regions are restrained, by purifying selection, to maintain protein structure or function, whereas synonymous substitutions and nucleotide substitution for non-coding regions accumulate constantly by random genetic drift. Therefore, K_A/K_S ratios are normally smaller than 1.0 (~0.2 in various genes) and K_N/K_S ratios are around 1.0 [29]. Table 1 shows the K values and ratios for congerin I and II cDNAs. The K_N values for 5′ and 3′ UTRs are smaller than K_S value for coding region, that is, the K_N/K_S values are around 0.7. This indicates that the UTRs are unusually conserved as compared to the coding-region. Furthermore, the extremely high ratio (2.6) of K_A/K_S was observed for congerins I and II, suggesting that non-synonymous replacements causing the amino acid changes were favored in these proteins. In other words, congerins I and II have evolved *via* accelerated

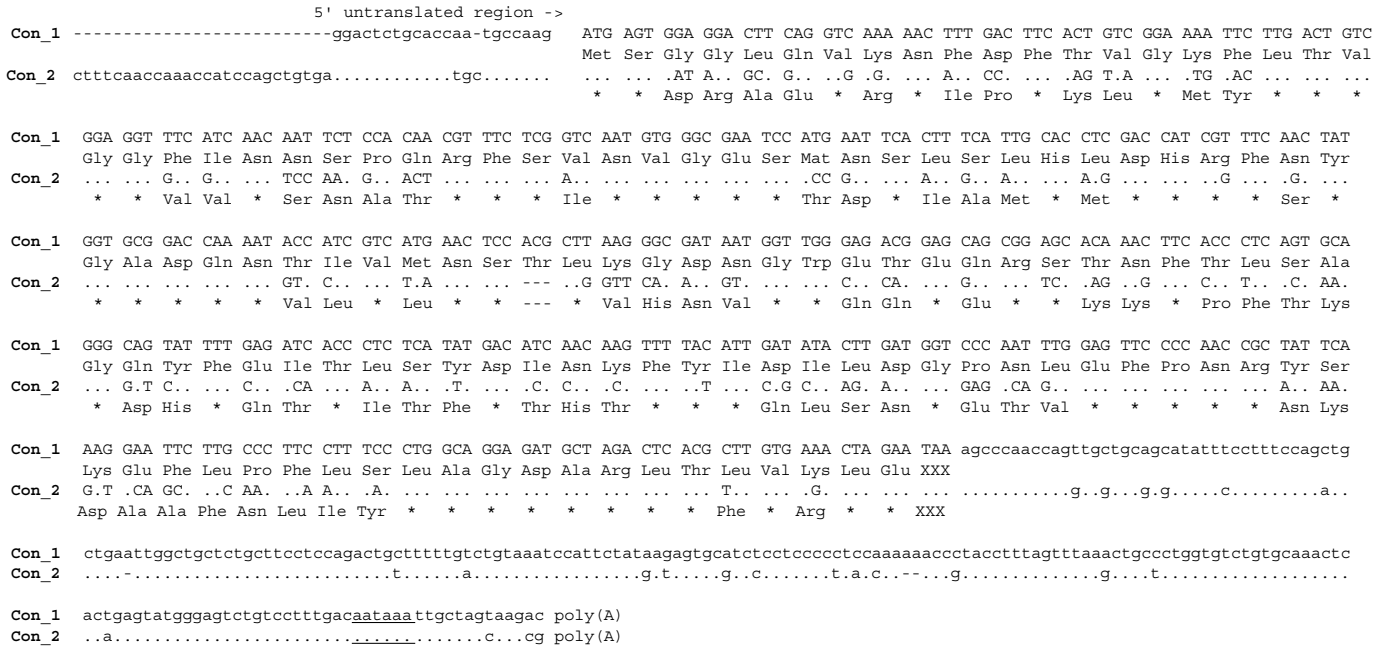

Figure 1. Alignment of cDNA sequences encoding Conger eel galectins, congerin I and congerin II and the deduced amino acid sequences. Con 1; congerin I, Con 2; congerin II. The 5′ and 3′ untranslated regions are shown in lower-case type. Dashes indicated gaps introduced to maximize sequence identity, and dots and asterisks represent the positions identical to Con 1 in the nucleotide and amino acid sequences, respectively. The poly adenylation signals are underlined.

amino acid substitutions under positive selection. Similar evolutionary behavior, has been observed only in several cDNAs and genes encoding animal toxins such as snake venom isozymes [6–18] and conus peptides [1–4] and encoding reproductive proteins [19].

Recently, a tandem-repeat type galectin has been identified with a cDNA library derived from rainbow trout (*Oncorhynchus mykiss*) head kidney stimulated by lipopolysaccharide (LPS) [30]. The deduced amino acid sequence shows two distinct N- and C-terminal domains composed of 148 and 130 amino acids, respectively, connected by a polypeptide linker of 63 amino acids. This tandem-repeat type galectin shows significant sequence identity of 40–55% with mammalian galectin 9, despite having only 19.3–25.3% (for N-terminal region) or 15.5–20.4% (for C-terminal region) identities to conger eel and electric eel galectins (Figure 2). This galectin was induced by stimulation with LPS, suggesting that it was also involved in the innate immune system. Furthermore, Japanese flounder (*Paralichthys olivaceus*) muscle galectin was cloned and sequenced [31]. Table 2 presents the pair wise comparison of the K_A and K_S valuesamong congerin I and II, *Oncorhynchus mykiss* galectin

Table 1. K_N/K_S and K_A/K_S values between congerins I and II cDNAs

	K_N	K_S	K_N/K_S
5′UTR	0.109	0.152	0.715
3′UTR	0.113	0.152	0.743
All noncoding regions	0.113	0.152	0.743
	K_A	K_S	K_A/K_S
Protein-coding region	0.391	0.152	2.57

Table 2. Comparison of K_A/K_S values of congerins and other galectins

Galectin pairs[a]	K_S	K_A	K_A/K_S
Con 1 vs. Con2	0.152	0.397	2.62
Con 1 vs. Po-gal	1.399	0.805	0.58
Con 1 vs. Om-galN	2.552	1.919	0.75
Con 1 vs. Om-galC	2.552	1.103	0.43
Con 1 vs. hGal-1	1.448	0.754	0.52
Con 2 vs. Po-gal	2.328	0.706	0.30
Con 2 vs. Om-galN	2.552	1.113	0.44
Con 2 vs. Om-galC	2.098	0.877	0.42
Con 2 vs. hGal-1	1.476	0.787	0.53
Po-gal vs. Om-galN	1.411	1.072	0.76
Po-gal vs. Om-galC	1.585	0.890	0.56
Po-gal vs. hGal-1	1.876	0.660	0.35
Om-galN vs. Om-galC	1.688	0.571	0.34
Om-galN vs. hGal-1	2.063	0.967	0.47
Om-galC vs. hGal-1	1.213	0.902	0.74

[a]Con1: congerin I, Con2: congerin II, el-eel: electric eel galectin, Po-gal: Japanese flounder (*Paralichthys olivaceus*) muscle galectin, Om-galN: rainbow trout (*Oncorhynchus mykiss*) galectin N-terminal domain, Om-galC: rainbow trout (*Oncorhynchus mykiss*) galectin C-terminal domain, hGal-1: human galectin-1.

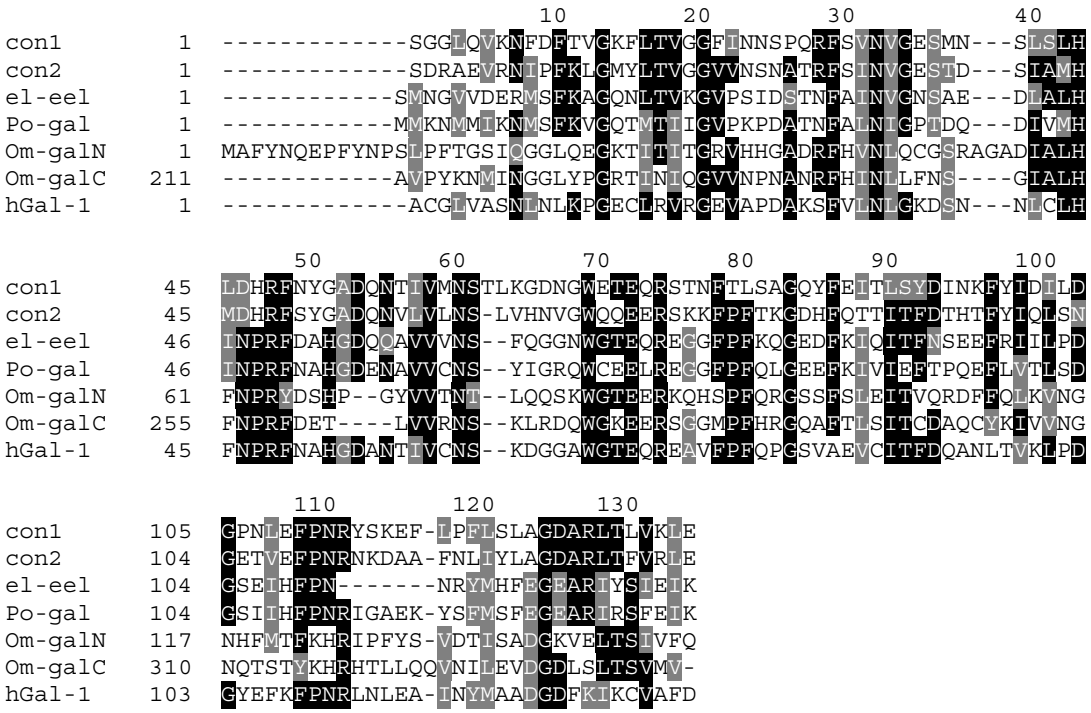

Figure 2. Aligned amino acid sequences of galectins from fish and human galectin-1. Alignment was achieved by clustal × program and represented by using BOX shade 3.21 (http://www.ch.embnet.org/software/BOX_form.html). Con1: congerin I, Con2: congerin II, el-eel: electric eel galectin, Po-gal: Japanese flounder (*Paralichthys olivaceus*) muscle galectin, Om-galN: rainbow trout (*Oncorhynchus mykiss*) galectin N-terminal domain, Om-galC: rainbow trout (*Oncorhynchus mykiss*) galectin C-terminal domain, hGal-1: human galectin-1. Residue number for Con1 is given above sequences. For Con2 it is the same up to residue 62, then one less.

(N- and C-terminal domains), *Paralichthys olivaceus* galectin, and human galectin-1. The K_A/K_S values are smaller than unity (0.30 to 0.76) except for the congerins I and II pair (2.6), indicating that only congerins I and II have evolved under positive selection (accelerated evolution). Other galectins are assumed to evolve under neutrality or purifying selection.

High-order structures of conger eel galectins

Three-dimensional domain swapping of congerin I as a stabilizing factor of dimer conformation

Schematic representation of the dimeric structures of congerins I and II, compared with that of bovine galectin-1 is shown in Figure 3. Both congerins I and II adopt the two-fold symmetric dimers with five-stranded and six stranded β-sheets called "jelly-roll" motif [32,33]. However, congerin I showed a remarkable different structure: a swap of strands between the two subunits in one of the intersubunit β sheets, causing a change in the folding topology. Contrary to the entirely anti-parallel β-sheets of other known galectins [34–41], the congerin I dimer has a partly parallel β-sheet by exchanging N-terminal strands between the subunits. Domain swapping has been hypothesized as a mechanism in quaternary structure formation in protein evolution [42] and reported in several proteins [43]. One of the important roles of the strand swapping seems to be increasing the stability of the quaternary structure. In congerin I, the strand swap increases the intersubunit contact surface area to nearly twice as large ($\sim$1280 A^2) as that of the other dimeric galectins including congerin II and bovine galectin-1 ($\sim$550 A^2). The numbers of intersubunit hydrogen bonds are also increased from 10 to 22. This suggests that the congerin I dimer is prevented from dissociation in both equilibrium and kinetic aspects. Congerin I maintained nearly whole activity after 80 min of incubation at 50°C, whereas congerin II completely lost activity under the same condition [25]. Replacement of the N-terminal strand of congerin I with that of congerin II or bovine galectin-1, by using protein engineering techniques, decreased its stability, while the replacement of both N- and C-terminal strands of congerin II with those of congerin I increased its stability, and the stability was almost equal to native congerin I (unpublished data). Thus, the strand-swap in congerin I explained the stability of congerin I. Stability of the dimeric structure is the most important factor for congerin I because its cross-linking activity in agglutinating cells depends on the structure. As the skin mucus is exposed to the surrounding outer environment, proteins in the skin mucus are required to maintain their activities under severe conditions such as higher temperature and hydraulic pressure. Thus, congerin I has evolved *via* accelerated amino acid substitution to stabilize the dimeric structure by strand swapping,

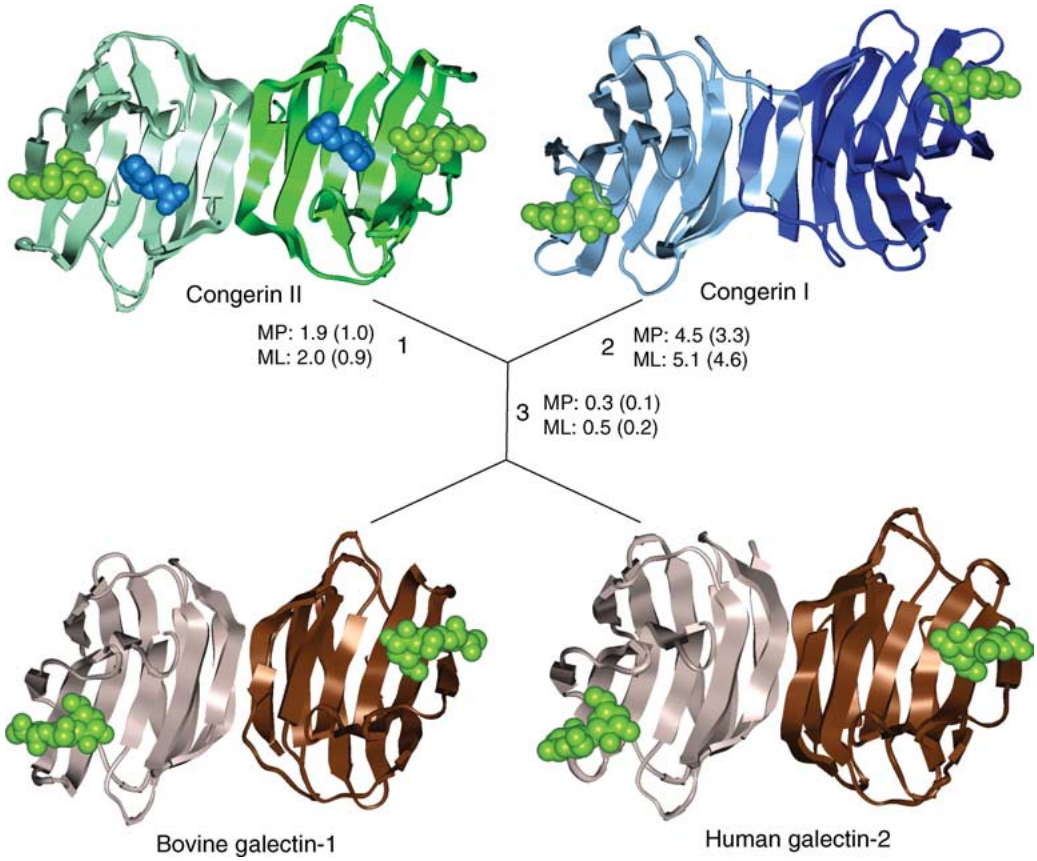

Figure 3. Schematic presentation of congerins I and II, bovine galectin-1, and human galectin-2 structures, and the molecular phylogeny of their genes. The subunits are differently colored in each model. The bound lactose molecules (white [online pdf version color: green]) and MES (grey [online pdf version color: blue]) are shown in space-filling model. The topology of the molecular phylogeny was inferred from nucleotide sequence of the genes by the maximum parsimony method. The numbers associated with branches are K_A/K_S values for the corresponding branches 1–3 estimated by the maximum parsimony (MP) and maximum likelihood (ML) methods. In parentheses are standard deviations of K_A/K_S value.

thereby promoting the emergence of a new structural class in the galectin family [32]. Recent studies on the crystal structural analysis of some snake C type lectin-like proteins such as factor XI/X binding protein [44], factor XI binding protein [45], flavocetin-A [46], botrocetin [47], and bitiscetin [48], revealed that they included domain swapping in the inter-subunit interface. These isoproteins also showed accelerated evolution to diversify their functions [16–18].

Additional carbohydrate binding site of congerin II acquired by adaptive evolution

Crystal structure analysis of congerin II at 1.45 Å resolution shows no strand swapping between the subunits. However, the congerin II structure suggests an extension of the carbohydrate binding cleft, which has not been found in congerin I. The crystal structure of congerin II revealed that a MES (2-*N*-morpholinoethane sulfonic acid) molecule which resembled sulfono-sugar in shape bound to the cleft near the known carbohydrate binding site, suggesting the presence of an additional carbohydrate binding cleft (Figure 4) [33]. The MES molecule fits into a depression, which is connected to the lactose-binding cleft. One hydrogen bond is formed between Tyr122 and the tertiary amine, and ionic interaction is formed between the sulfate group and Arg3.

From the structural analysis of galectins co-crystallized with *N*-acetyl-lactosamine or lactose, the highly conserved amino acid residues, namely His44, Arg48, Asn61, Glu73 and Arg75 (as numbered in congerin I, Figure 2), involved in the carbohydrate recognition through hydrogen bond interactions have been identified. Trp70, which is conserved among various galectins, plays a role in a stacking interaction with B face of the galactose residue (Figure 4). The number of protein-lactose hydrogen bonds in congerin II is smaller than that of congerin I, and rather close to those of the other galectins. The increase in the number of hydrogen-bonds improved the cell cross-linking activity of congerin I. The structure of congerin II showed no significant improvement in either dimer stability or ligand interaction. Congerin II looks like an orthodox galectin. However, some differences are observed in the carbohydrate-binding clefts of congerins compared with other galectins. The side chain of Arg29 in congerins I and II contributes to more extensive ligand

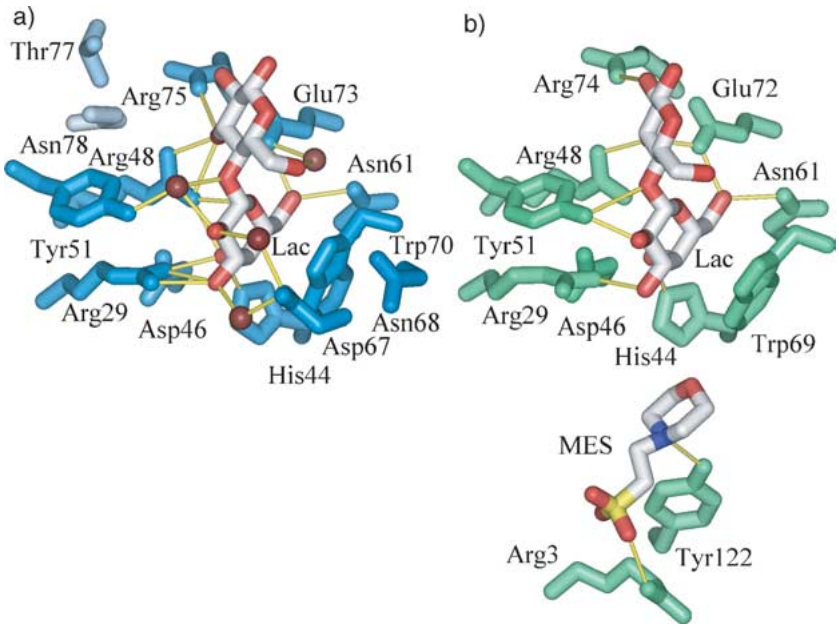

Figure 4. Structure of the carbohydrate-binding clefts of (a) congerin I and (b) congerin II. Bound lactose (Lac) and MES molecules and the side chains of the residues, that are mentioned in the text, are shown in stick model. Hydrogen-bonds between the ligand and the protein are show in a thin line [online pdf version color: yellow].

interaction. Although the arginine residue is conserved in some galectins such as human galectin-2 and galectin-3, its side chain is not used for direct interaction with ligands in these galectins. In congerins, Asp46 that is replaced by Asn in other galectins helps to hold the side chain of Arg29 in appropriate position for direct interaction. Although the conserved amino acid residues, namely Arg29, His44, Arg48, Tyr51, Asn61, Trp70, Glu73 and Arg75, which participate in direct binding with lactose in the CRD of congerins I and II can be completely superimposed in their tertiary structures, the peripheral residues of these CRDs are variable (Figure 4).

These variable residues in congerins I and II seem to be related to their carbohydrate binding specificities and the novel binding cleft in congerin II. For example, Asp67 and Asn68 in congerin I are found in a loop that makes up part of an inside wall of the binding cleft. These residues are substituted to Asn and Val in congerin II, respectively, resulting in another subsite of the binding-cleft at the end of galactose binding site. These amino acid changes, hence, may contribute to the difference in specificity between congerins I and II for glycoconjugate polymers with the trisaccharide segment of globotriaosyl ceramide (Gb3), which congerin II recognized with high affinity ($K_d = \sim 6.0 \times 10^{-6}$ M) while congerin I did not (unpublished data).

Congerin I shows no binding activity against C-3-modified methyl β-lactose while congerin II shows some activity [25]. Residues Thr77 and Asn78 located near Arg75 in congerin I, which interact with glucose O3 region of lactose, are substituted to two Lys residues in congerin II. Recently, the role of the Thr77 residue for binding specificities was studied by site-directed mutagenesis (data not shown). These residues might contribute to the binding specificities for some lactose derivatives. Congerin II seems to have broader specificities for carbohydrate recognition than congerin I.

The presence of multiple galectins that have been generated by accelerated evolution provides some advantages for the biodefense system, that is, the galectins having different specificities, could respond to a wide range of bacteria and parasites.

A model of congerin evolution

The ancestral congerin genes appear to have been subjected to purifying selection, like other galectins, until a gene duplication occurred (branch 3 in Figure 3, $K_A/K_S = 0.31$, SE $= 0.10$ by maximum parsimony method; $K_A/K_S = 0.49$, SE $= 0.16$ by maximum likelihood method) generating the most recent common ancestor. Then further evolution of both congerin genes seem to have been indifferent to the constraint on the ancestral gene, and instead evolved in a more rapid adaptive way. Why? It is possible that two consecutive gene duplications have occurred. One of the preceding duplicates became expressed in skin and stomach cells, and was recruited into a defense system. Alteration in expression pattern is frequently associated with a gene duplication, and is hypothesized to be a process to maintain both of the duplicates in genome [49], often with one carrying on the old function and the other taking on new functions. After the switch of expression pattern, the second duplication occurred and genes of congerin I and II were generated. Thus, the two congerin genes may have differentiated to adapt two related but different tasks. Congerin I was

enhanced in cross-linking activity, which was advantageous in skin mucus or stomach, because they are in direct contact with the environment such as higher temperature or higher hydraulic pressure. Congerin II did not respond to the selective pressure as congerin I did. Instead, as typical for a defense protein, it was targeted to a new carbohydrate molecule on some parasite. A gene duplication does not necessarily cause accelerated evolution. It is remarkable that both of the congerin genes differentiated by positive selection. Congerin (galectin) might be a protein with high plasticity and can be readily adapted to new purposes. In fact, galectins were recruited into a wide variety of biological activities including cell adhesion, recognition, differentiation, apoptosis or even RNA splicing. The high plasticity of congerins permitted the particular case that natural selection introduced a new protein fold and a new binding site for the two duplicate genes. However, galectins except for congerins have evolved neutrally or under purifying selection, and not in an accelerated manner under positive selection as shown in Table 2. Why do only conger eel galectins show the accelerated evolution? It is thought that galectins with diverse carbohydrate recognition specificity and stable structure are advantageous to act on various target bacteria and pathogens for survival of conger eels. However, the true mechanism and selective pressure for accelerated evolution is unknown. The unique futures of conger eel garectins as well as snake venom C-type lectin like proteins, which show accelerated evolution, prompt us to analyze the organization and structure of these genes. Genomics and proteomics analyses will allow the accumulation of much more data to understand the mechanism of accelerated evolution.

References

1 Duda TF Jr, Palumbi SR, Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod Conus, *Proc Natl Acad Sci USA* **96**, 6820–3 (1999).

2 Duda TF Jr, Palumbi SR, Evolutionary diversification of multigene families: Allelic selection of toxins in predatory cone snails, *Mol Biol Evol* **17**, 1286–93 (2000).

3 Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M, Mechanisms for evolving hypervariability: The case for conopeptides, *Mol Biol Evol* **18**, 120–31 (2001).

4 Olivera BM, Walker C, Cartier GE, Hooper D, Santos AD, Schoenfeld R, Shetty R, Watkins M, Bandyopadhyay P, Hillyard DR, Speciation of cone snails and interspecific hyperdivergence of their venom peptides: Potential evolutionary significance of introns, *Ann New York Acad Sci* 223–37 (1999).

5 Froy O, Sagiv T, Poreh M, Urbach D, Zilberberg N, Gurevitz M, Dynamic diversification from a putative common ancestor of scorpion toxins affecting sodium, potassium, and chloride channels, *J Mol Evol* **48**, 187–96 (1999).

6 Ogawa T, Oda N, Nakashima K, Sasaki H, Hattori M, Sakaki Y, Kihara H, Ohno M, Unusually high conservation of untranslated sequences in cDNAs for *Trimeresurus flavoviridis* phospholipase A_2 isozymes, *Proc Natl Acad Sci USA* **89**, 8557–61 (1992).

7 Nakashima K, Ogawa T, Oda N, Hattori M, Sakaki Y, Kihara H, Ohno M, Accelerated evolution of *Trimeresurus flavoviridis* venom gland phospholipase A_2 isozymes, *Proc Natl Acad Sci USA* **90**, 5964–8 (1993).

8 Nakashima K, Nobuhisa I, Deshimaru M, Nakai M, Ogawa T, Shimohigashi Y, Fukumaki Y, Hattori M, Sakaki Y, Hattori S, Ohno M, Accelerated evolution in the protein-coding regions is universal in crotalinae snake venom gland phospholipase A_2 isozyme genes, *Proc Natl Acad Sci* **92**, 5605–9 (1995).

9 Nobuhisa I, Nakashima K, Deshimaru M, Ogawa T, Shimohigashi Y, Fukumaki Y, Sakaki Y, Hattori S, Kihara H, Ohno M, Accelerated evolution of *Trimeresurus okinavensis* venom gland phospholipase A_2 isozyme-encoding genes, *Gene* **172**, 267–72 (1996).

10 John TR, Smith LA, Kaiser II, Genomic sequences encoding the acidic and basic subunits of Mojave toxin: Unusually high sequence identity of non-coding regions, *Gene* **139**, 229–34 (1994).

11 Kordis D, Gubensek F, Ammodytoxin C gene helps to elucidate the irregular structure of Crotalinae group II phospholipase A_2 genes, *Eur J Biochem* **240**, 83–90 (1996).

12 Kordis D, Bdolah A, Gubensek F, Positive Darwinian selection in vipera palaestinae phospholipase A_2 genes is unexpectedly limited to the third exon, *Biochem Biophys Res Commun* **251**, 613–9 (1997).

13 Deshimaru M, Ogawa T, Nakashima K, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S, Ohno M, Accelerated evolution of crotalinae snake venom gland serine proteases, *FEBS Lett* **397**, 83–8 (1996).

14 Ohno M, Menez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC, Tamiya T, Menez A, Molecular evolution of snake toxins: Is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? *Prog Nucleic Acid Res Mol Biol* **59**, 307–64 (1998).

15 Lachumanan R, Armugam A, Tan CH, Jeyaseelan K, Structure and organization of the cardiotoxin genes in Naja naja sputatrix, *FEBS Lett* **14**, 119–24 (1998).

16 Atoda H, Yoshihara E, Yamada M, Morita T, cDNA cloning of a heterogeneous two-chain anticoagulant protein IX-bp, *Thromb Res* **87**, 271–8 (1997).

17 Shin Y, Okuyama I, Hasegawa J, Morita T, Molecular cloning of glycoprotein Ib-binding protein, flavocetin-A, which inhibits platelet aggregation, *Thromb Res* **99**, 239–47 (2000).

18 Tani A, Ogawa T, Nose T, Nikandrov N, Deshimaru M, Chijiwa T, Chang C, Fukumaki Y, Ohno M, Characterization, primary structure and molecular evolution of anticoagulant protein from *Agkistrodon actus* venom, *Toxicon* **40**, 803–13 (2002).

19 Swanson WJ, Vacquier VD, The rapid evolution of reproductive proteins, *Nature Rev Genetics* **3**, 137–44 (2002).

20 Hirabayashi J, Saitoh M, Kasai K, Evidence that *Caenorhabditis elegans* 32-kDa beta-galactoside-binding protein is homologous to vertebrate beta-galactoside-binding lectins. cDNA cloning and deduced amino acid sequence, *J Biol Chem* **267**, 15485–90 (1992).

21 Kasai K, Hirabayshi J, Galectins: A family of animal lectins that decipher glycocodes, *J Biochem* **119**, 1–8 (1996).

22 Inagaki Y, Sohma Y, Horie H, Nozawa R, Kadoya T, Oxidized galectin-1 promotes axonal regeneration in peripheral nerves but does not possess lectin properties, *Eur J Biochem* **267**, 2955–64 (2000).

23 Kamiya H, Muramoto K, Goto R, Purification and properties of agglutinins from conger eel, *Conger Myriaster* (Brevoort), skin mucus, *Dev Comp Immuno* **12**, 309–18 (1988).

24 Muramoto K, Kamiya H, The amino-acid sequence of a lectin from conger eel, *Conger myriaster*, skin mucus, *Biochim Biophys Acta* **1116**, 129–36 (1992).

25 Muramoto K, Kagawa D, Sato T, Ogawa T, Nishida Y, Kamiya H, Functional and structural characterization of multiple galectins from the skin mucus of conger eel, *Conger myriaster*, *Comp Biochem Physiol B* **123**, 33–45 (1999).

26 Ingram GA, Substances involved in the natural resistance of fish to infection-A review, *J Fish Biol* **16**, 23–60 (1980).

27 Ogawa T, Ishii C, Kagawa D, Muramoto K, Kamiya H, Accelerated evolution in the protein-coding region of galectin cDNAs, congerin I and congerin II, from skin mucus of conger eel (*Conger myriaster*), *Biosci Biotechnol Biochem* **63**, 1203–8 (1999).

28 Nakamura O, Watanabe T, Kamiya H, Muramoto K, Galectin containing cells in the skin and mucosal tissues in Japanese conger eel, *Conger myriaster*: An immunohistochemical study, *Dev Comp Immuno* **25**, 431–7 (2001).

29 Nei M, *Molecular Evolutionary Genetics* (Columbia Univ. Press, Irvington-on-Hudson, N.Y., 1987).

30 Inagawa H, Kuroda A, Nishizawa T, Honda T, Ototake M, Yokomizo Y, Nakanishi T, Soma G, *Fish & Shellfish Immunol* **11**, 217–31 (2001).

31 Lee J, Jeon J, Song Y, unpublished data (Sequence data was available with accession number AF220550 from Genbank/EMBL/DDBJ data base) (2000).

32 Shirai T, Mitsuyama C, Niwa Y, Matsui Y, Hotta H, Yamane T, Kamiya H, Ishii C, Ogawa T, Muramoto K, High-resolution structure of conger eel galectin, congerin I, in lactose-liganded and ligand-free forms: Emergence of a new structure class by accelerated evolution, *Structure* **7**, 1223-33 (1999).

33 Shirai T, Matsui Y, Mitsuyama C, Yamane T, Kamiya H, Ishii C, Ogawa T, Muramoto K, Crystal structure of a conger eel galectin (congerin II) at 1.45 A resolution: Implication for the accelerated evolution of a new ligand-binding site following gene duplication, *J Mol Biol* **321**, 879–89 (2002).

34 Liao DI, Kapadia G, Ahmed H, Vasta GR, Herzberg O, Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein, *Proc Natl Acad Sci USA* **91**, 1428–32 (1994).

35 Bourne Y, Bolgiano B, Liao D, Strecjer G, Cantau P, Herzberg O, Feizi T, Cambillau C, Cross-linking of mammalian lectin (galectin-1) by complex biantennary saccharides,*Nature Struct Biol* **1**, 863–70 (1994).

36 Lobsanov YD, Gitt MA, Leffler H, Barondes SH, Rini JM, 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**, 27034–8 (1993).

37 Leonidas DD, Vatzaki EH, Vorum H, Celis JE, Madsen P, Acharya KR, Structural basis for the recognition of carbohydrates by human galectin-7, *Biochem* **37**, 13930–40 (1998).

38 Leonidas DD, Elbert BL, Zhou Z, Leffler H, Ackerman SJ, Acharya KR, Crystal structure of human Charcot-Leyden crystal protein, an eosinophil lysophospholipase, identifies it as a new member of the carbohydrate-binding family of galectins, *Structure* **3**, 1379–93 (1995).

39 Seetharaman J, Kanigsberg A, Slaaby R, Leffler H, Barondes SH, Rini JM, X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution, *J Biol Chem* **273**, 13047–52 (1998).

40 Varela PF, Solis D, Diaz-Maurin TG, Kaltner H, Gabius HJ, Romero A, The 2.15 Å crystal structure of CG-16, the developmentally regulated homodimeric chicken galectin, *J Mol Biol* **294**, 537–49 (1999).

41 Bianchet MA, Ahmed H, Vasta GR, Amzel LM, Soluble beta-galactosyl-binding lectin (galectin) from toad ovary: Crystallographic studies of two protein-sugar complexes, *Proteins* **40**, 378–88 (2000).

42 Bennett MJ, Schlunegger MP, Eisenberg D, 3D domain swapping: A mechanism for oligomer assembly, *Protein Sci* **4**, 2455–68 (1995).

43 Schlunegger MP, Benett MJ, Eisenberg D, Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly, *Adv Protein Chem* **50**, 61–122 (1997).

44 Mizuno H, Fujimoto Z, Koizumi M, Kano H, Atoda H, Morita T, Structure of coagulation factors IX/X-binding protein, a heterodimer of C-type lectin domains, *Nat Struct Biol* **4**, 438–41 (1997).

45 Mizuno H, Fujimoto Z, Koizumi M, Kano H, Atoda H, Morita T, Crystal structure of coagulation factor IX-binding protein from habu snake venom at 2.6 A: Implication of central loop swapping based on deletion in the linker region, *J Mol Biol* **289**, 103–12 (1999).

46 Fukuda K, Mizuno H, Atoda H, Morita T, Crystal structure of flavocetin-A, a platelet glycoprotein Ib-binding protein, reveals a novel cyclic tetramer of C-type lectin-like heterodimers, *Biochemistry* **39**, 1915–23 (2000).

47 Sen U, Vasudevan S, Subbarao G, McClintock RA, Celikel R, Ruggeri ZM, Varughese KI, Crystal structure of the von Willebrand factor modulator botrocetin, *Biochemistry* **40**, 345–52 (2001).

48 Hirotsu S, Mizuno H, Fukuda K, Qi MC, Matsui T, Hamako J, Morita T, Titani K, Crystal structure of bitiscetin, a von Willebrand factor-dependent platelet aggregation inducer, *Biochemistry* **40**, 13592–7 (2001).

49 Cooke J, Nowak MA, Boerlijst M, Maynard-Smith J, Evolutionary origins and maintenance of redundant gene expression during metazoan development, *Trends Genet* **13**, 360–4 (1997).