Reconstruction of a Probable Ancestral Form of Conger Eel Galectins Revealed Their Rapid Adaptive Evolution Process for Specific Carbohydrate Recognition

Ayumu Konno,* Tomohisa Ogawa,* Tsuyoshi Shirai,† and Koji Muramoto*

*Department of Biomolecular Science, Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi, Japan; and †Nagahama Institute of Bio-Science and Technology, and Japan Science and Technology Agency, BioInformatic Research Division, Nagahama, Shiga, Japan

Recently, many cases of rapid adaptive evolution, which is characterized by the higher substitution rates of nonsynonymous substitutions to synonymous ones, have been identified in various genes of venomous, biodefense, and reproductive proteins (Ogawa 2006; Yang and Bielawski 2000; Ohno et al. 1998). For example, animal toxins such as snake toxins, scorpion toxins, and conotoxins involve a diverse array of peptides and proteins that show similar molecular scaffold but different biological functions (Ohno et al. 1998; Froy et al. 1999; Duda and Palumbi 1999). This evolutionary phenomenon must have important consequences on the functions of their gene products, such as multiple specific toxins for quick immobilization of the prey and the establishment of barriers in fertilization that might lead to specialization and on the generation of novel genes.

Galectins are a family of galactoside-specific animal lectins that are responsible for a variety of biological activities, including cell adhesion, differentiation, morphogenesis, apoptosis, and metastasis of malignant cells, and they are classified into 3 types: prototype, tandem repeat, and chimera (Kasai and Hirabayashi 1996). Two prototype galectins, namely, congerins I and II (ConI and ConII), were distributed not only in the epidermis but also in the epithelia of the upper digestive tract and gills, and they exhibited agglutinating activity against the marine pathogen Vibrio anguillarum as well as opsonic effect (Kamiya et al. 1988; Muramoto and Kamiya 1992; Muramoto et al. 1999, 2004). These galectins are produced by club cells that distribute not only in the epidermis but also in the epithelium of the upper digestive tract and gills, and they exhibited agglutinating activity against the marine pathogen Vibrio anguillarum as well as opsonic effect (Kamiya et al. 1988; Nakamura et al. 2001, 2006). These observations suggest that congerins contribute to the biodefense of the external and internal body surface of conger eel. We previously demonstrated that congerins have evolved under significant selection pressure, that is, accelerated evolution with the emergence of a new strand-swap structure between the subunits in ConI and a unique new ligand-binding site in ConII (fig. 1A; Ogawa et al. 1999, 2004; Shirai et al. 1999, 2002). The strand swapping is a motility in quaternary structure formation in protein evolution; a protein becomes multimeric by donating a part (β strand) of the molecule to a cognate molecule and then accepting the corresponding portion from the cognate. The strand-swap structure in ConI (fig. 1A) increases the intersubunit contact surface area and the number of intersubunit hydrogen bonds. ConI increased the dimeric stability that required its cross-linking activity in agglutinating cells or bacterial pathogens under severe conditions. Thus, the selection pressure for the ConI gene was estimated to be the enhancement of its dimeric stability and its cross-linking activity (Shirai et al. 1999). On the other hand, the pressure for the ConII gene was estimated to be the adaptation to new ligands, which resulted in the modified binding site in ConII (Shirai et al. 2002). However, the selection pressures implicated in the differences between the congerin genes remain unknown. In order to understand the mechanisms of rapid adaptive evolution, it is essential to identify the relationships between the specific adaptive substitutions and the resulting functions.

In the present study, to understand and reveal the mechanism of the rapid adaptive evolution process of congerins, we designed and prepared the probable ancestral congerin (Con-anc) mutant that corresponded to the putative amino acid sequence at the divergence of ConI and ConII in the phylogenetic tree. Ancestral proteins can provide a useful model system for analyzing functional diversification and evolutionary process under adaptive natural selection pressure.

Materials and Methods
Design and Construction of DNA Encoding the Con-anc

The hypothetical Con-anc corresponding to the putative amino acid sequence at the divergence of ConI and
ConII in the phylogenetic tree (branch 3 in fig. 1) was predicted by using the molecular phylogeny tree of galectins. This phylogeny tree was constructed from nucleotide sequences of galectins by the maximum parsimony (MP) method using the ODEN program package (Ina 1994) available at the National Institute of Genetics, Japan. The percent reproduction of the branches 3, 5, and 7 in figure 1 are 100%, 82%, and 100%, respectively, in 1,000 bootstrap reconstructions of the tree. When the deduced ancestral sequence was used for tree construction along with the other galectins, it branched out from the node between ConI and ConII with zero distance according to expectation (tree not shown). The ancestral sequence was also confirmed by maximum likelihood (ML) method (Felsenstein 1981) using PAML program (Yang 1997). Galectin sequences and the 3D structural data were retrieved from DDBJ, Swiss-Prot, and PDB databases (Abbott et al. 1989; Hirabayashi et al. 1990; Gitt et al. 1992; Lobsonov et al. 1993; Liao et al. 1994; Ogawa et al. 1999; Shirai et al. 1999, 2002; Varela et al. 1999; Bianchet et al. 2000).

The accession numbers of the respective nucleotide sequences, amino acid sequences, and structural coordinates are as follows: ConII, ab010277, leg2_conmy (q9yic2), lis3; ConI, ab010276, leg1_conmy (p26788), lic1; bovine galectin-1, x14330, leg1_bovin (p11116), lisl; human galectin (hGal)-2, m87842, leg2_human (p05162), hlc; hGal-1, x14829, leg1_human (p09382), not available (n. a.); chicken galectin-1, m57240, leg6_chick (p23668), hmqj; toad galectin-1, n.a., leg1_bufar (p56217), lat8; and Xenopus galectin-1, af170341, q98ud4, n.a.

Construction of Con-anc Expression Plasmids

The cDNA encoding Con-anc was constructed by polymerase chain reaction (PCR), and 4 fragments, namely,
Con-ancN, Con-ancMa, Con-ancMb, and Con-ancC, were ligated at suitable restriction sites (supplementary fig. S2, Supplementary Material online). Synthetic oligonucleotides—anN100S, ancN20S, ancN20A, ancM45S, ancM54A, ancM90S, ancMb21S, ancMb21A, ancC100S, ancC21S, and ancC48AS—were purchased from Kurabo Co. (Osaka, Japan). Three cDNA fragments, namely, Con-ancN, Con-ancMb, and Con-ancC, were prepared by PCR using ancN100S, ancM90S, and ancC100S as templates and ancN25S/ancN20A, ancMb21S/ancMb21A, and ancC21S/ancC48AS as primer sets. The Con-ancMb fragment was prepared by PCR using the ConI cDNA as the template and ancMa45S and ancMa54A as primers. The 2 sets of PCR products—Con-ancN and Con-ancMa, and Con-ancMb and Con-ancC—were ligated to each other after digestion with AcI and SacI, respectively. Next, a second PCR was carried out using the ligation products Con-ancN/Con-ancMa and Con-ancMb/Con-ancC as templates and ancN25S/ancMa54A and ancMb21S/ancC48AS as external primers. Subsequently, these fragments were ligated at the PvuI site and used for the third PCR as the template. Finally, the PCR product encoding the complete sequence of Con-anc was ligated into the pTV-118N vector at the NeoI and PstI sites after digestion with BspHI and PstI, thereby resulting in pTV-Con-anc. The nucleotide sequence of pTV-Con-anc was confirmed by DNA sequencing.

Site-Directed and Cassette Mutagenesis of Con-anc and ConII

Con-anc and ConII mutants, namely Con-anc (G3R), Con-anc (L41I), Con-anc (S123Y), Con-anc (G3R/S123Y), ConII (R3G), ConII (I41L), ConII (Y123S), and ConII (R3G/Y123S), were generated using PCR-based site-directed mutagenesis. PCRs were performed using KOD plus DNA polymerase (Toyobo Co., Osaka, Japan) containing modified oligonucleotides not shown), and N-terminal amino acid sequencing using a gas-phase protein sequencer (Model PSQ-1, Shimadzu, Kyoto, Japan).

Structural Stability Measurements

The thermostability of rCon-anc was evaluated by the residual hemagglutination activities after incubating at various temperatures ranging from 38 to 62 °C (fig. 3A). After cooling on ice for 30 min, the residual hemagglutination activities were examined using 2% rabbit erythrocytes.

The urea-induced denaturation of rCon-anc was studied by circular dichroism (CD) measurements at 25 °C within the urea concentration range from 0 to 8 M (fig. 3B). Far-UV CD spectra of purified rCon-anc, rConI, and rConII were measured by means of a JASCO J 720 spectropolarimeter with a quartz cell (a path length of 0.1 cm). Sample proteins were dissolved in 50 mM Tris–HCl buffer (pH 7.5) containing 0–8 M urea at a protein concentration of 0.1 mg/ml, respectively. Each measurement was background corrected from the average of 5 repeated scans at 25 °C. Profiles are normalized to mean residue ellipticity for comparison. Fractions denatured (%) of rCon-anc, rConI, and rConII were estimated by CD at 222 nm.

Frontal Affinity Chromatography

The sugar-binding specificities of Con-anc were analyzed by frontal affinity chromatography (FAC) according to the reported method (Hirabayashi et al. 2002, 2003). The purified Con-anc was immobilized onto HiTrap NHS-activated columns and packed into a stainless column (4.0 × 10 mm). In this chromatography, ethylenediaminetetraacetic acid (EDTA)–phosphate-buffered saline (1 mM EDTA, 20 mM Na phosphate [pH 7.2], and 150 mM NaCl)
was used as the elution buffer. Various pyridylaminated (PA) oligosaccharides (final concentration 10 nM) (fig. 4A) were applied to the column through a 2-ml sample loop connected to an injector at a flow rate of 0.25 ml/min at 25 °C, and the eluted PA oligosaccharides were monitored by a fluorescence detector (Excimer: 320 nm, Emission: 400 nm). The $B$ values were determined using various concentrations of nonlabeled lacto-N-tetraose in the presence of 10 nM PA-labeled lacto-N-tetraose. We used PA rhamnose as the control. ConI and ConII were also analyzed by the same method.

Surface Plasmon Resonance Analysis

All surface plasmon resonance (SPR) experiments were performed on a BIAcore X biosensor (Biacore AB, Uppsala, Sweden) at 25 °C in the HEPES buffered saline-EDTA-P20 (HBS–EP) buffer (pH 7.4). Lyso-GM3 (Takara Bio, Otsu, Japan) was immobilized on the sensor chip CM5 (Biacore AB). Purified congerins, Con-anc, and their mutants were injected as analytes at the concentrations of 1.0, 0.67, 0.50, 0.33, 0.25, 0.17, 0.13, and 0.083 mg/ml over the GM3-immobilized surface at a flow rate of 20 μl/min. Sensorgrams were analyzed using the BIAevaluation 4.1 software. The affinity constant $K_A$ was calculated from the plots of the steady-state binding level $R_{eq}$ to the analyte concentrations.

Molecular Modeling

The 3D structure of rCon-anc was predicted using the MOE package (Chemical Computing Group Inc., Montreal, Quebec, Canada) with the X-ray crystal structure of ConI (PDB code 1C1L) (Shirai et al. 1999), which shares the highest sequence homology to Con-anc. The modeled structure of Con-anc was superimposed with ligand-free ConI (PDB code 1C1F), lactose-liganded ConII, and 2-morpholinoethanesulfonic acid (MES) (PDB code 1IS3) (Shirai et al. 2002).

Results

Design and Preparation of the Con-anc

The hypothetical Con-anc corresponding to the putative amino acid sequence at the divergence of ConI and ConII in the phylogenetic tree (the node indicated as Con-anc in fig. 1A) was determined by constructing the molecular phylogenetic tree of galectins using MP or ML methods. The ancestral sequences deduced by MP and ML methods were identical except for residue 115 (Lys or Arg, respectively). The lysine residue was adopted as Con-anc sequence at 115th position because the corresponding residues in ConI and ConII are Ser and Lys, respectively. Figure 2A showed the deduced amino acid sequence of Con-anc, which shared 76%, 61%, and 40% identities with ConI, ConII, and hGal-1, respectively. To assess the accuracy of the putative ancestral sequence, the distribution of posterior probabilities for Con-anc in ML was analyzed by PAML (fig. 2B). The posterior probabilities for approximately 50% sites of Con-anc were more than 0.9 and those for 90% sites were >0.5. The key amino acid residues involved in carbohydrate recognition (His44, Arg48, Asn61, Trp70, and Glu73) of galectins are completely conserved in Con-anc. Furthermore, similar to ConI and ConII, Con-anc has no cysteine residues; this is related to oxidizing inactivation commonly found in other mammalian galectins. The $K_A/S$ ratios on branch 1 (gene duplication to ConI: $K_A/S = 4.50$, standard error [SE] = 3.31 by MP; $K_A/S = 5.14$, SE = 4.57 by ML) and branch 2 (gene duplication to ConII: $K_A/S = 1.92$, SE = 0.95 by MP; $K_A/S = 1.98$, SE = 0.90 by ML) are larger than unity (fig. 1B); this indicates that both ConI and ConII have evolved under positive selection pressure from their Con-anc gene after the gene duplication event. On the other hand, the $K_A/S$ ratios on branches among other galectins, including that right before the gene duplication (branch 3 in fig. 1A), were below unity (fig. 1B).

The rCon-anc protein was successfully expressed as the soluble form with hemagglutinating activity. The yield of purified rCon-anc was ~4.0 mg from 11 of 2 × YT medium after 2-step column chromatographies (see supplementary fig. S1A, Supplementary Material online). The rCon-anc protein gave a single band at 15 kDa on SDS–PAGE (supplementary fig. S1B, Supplementary Material online). The molecular mass of rCon-anc (15,251.49 Da) analyzed by MALDI-TOF/MS was in good agreement with its theoretical value (15,234.72 Da) calculated from the amino acid sequence. Furthermore, rCon-anc was confirmed by N-terminal amino acid sequencing.

Structural Stability of rCon-anc

The thermostability of rCon-anc was estimated by analyzing its residual hemagglutinating activity after incubation at various temperatures, and the activity was compared with those of rConI and rConII (fig. 3A). After 30 min of incubation at 50 °C, the residual activities of rCon-anc and rConII decreased by 13%, whereas that of rConI remained at 50%. Although rCon-anc and rConII showed no or very weak (3%) activity at 52 °C, rConI kept 25% activity. The half-activity retention temperatures, wherein half of the hemagglutinating activity was retained after incubation for 30 min, were estimated to be 46.5 °C for rCon-anc and rConII and 51 °C for ConI. Therefore, the thermostability level of rCon-anc was the same as that of rConII and ~5 °C lower than that of rConI.

To examine the effect of a denaturing agent on the protein conformation, the CD spectroscopy was used to monitor the urea-induced unfolding reaction at a series of the urea concentrations. The CD spectra of rCon-anc in 50 mM Tris–HCl buffer (pH 7.5) showed very similar profiles with ConI and ConII, indicating that probable ancestral form, rCon-anc, is also adopted similar scaffold including β structure with current congerins (data not shown). Figure 3B showed the denatured fraction (%) of proteins as a function of urea concentration in the range from 0 to 8 M. EC50 values of urea for protein denaturing were estimated to be 3.58, 5.04, and 3.50 M for Con-anc, ConI, and ConII, respectively. These results indicate that the stability of rCon-anc was similar to that of rConII and lower than that of rConI.
Sugar-Binding Specificity of rCon-anc

The sugar-binding specificities of rCon-anc, ConI, and ConII to 34 types of PA oligosaccharides, including N-linked type and glycolipid type sugars (fig. 4A), were analyzed using FAC system. Figure 4 shows the comparison of the $K_A$ values for Con-anc, ConI, and ConII. In general, rCon-anc showed a sugar-binding specificity similar to that of ConII but had a lower affinity than ConI for almost all PA oligosaccharides (fig. 4B). However, with regard to GM3- and GD1a-binding ability, the binding specificities of rCon-anc and ConII were remarkably different; that is, rCon-anc had lower affinity to GM3(NeuAc), GM3(NeuGc), and GD1a when compared with that of ConII (fig. 4C). Con-anc showed only 20% binding activity of ConII to GM3 and GD1a, whereas it showed 85% activity to other oligosaccharides. These results suggest that ConII acquired the ability to bind to the α2,3-sialyl galactose moieties such as GM3 and GD1a from Con-anc during the accelerated evolutionary event.

To investigate the difference in binding activities among rCon-anc, rConI, and rConII to GM3, we employed SPR analysis using the GM3-immobilized surface (fig. 5). SPR is a simple and direct sensing technique that measures the refractive index ($\eta$) changes near the thin metal film surface that occurs during complex formation or dissociation in real time without labeling requirements (Markey 1999). The SPR analysis revealed that rConI exhibited no binding activity toward GM3 and that Con-anc ($K_A = 1.5 \times 10^4$ M) had slightly lower affinity than ConII ($K_A = 2.2 \times 10^4$ M). Furthermore, to elucidate the structural basis of ConII for recognition of α2,3-sialyl galactose moieties of GM3, comparative mutagenesis analysis of ConII and Con-anc was conducted; ConII and Con-anc mutants in which amino acid residues at positions 3, 41, and 123 involved in the ability of ConII to bind to the α2,3-sialyl galactose moieties in GM3 and GD1a (fig. 4C) were intersubstituted with each other. A comparative mutagenesis study of ConII indicated that the replacement of Arg3 and Tyr123 residues of ConII with Gly and Ser, respectively, decreased its binding activity to GM3, whereas the replacement of Ile41 with Leu showed no apparent effect on the GM3-binding activity (fig. 5).
Effect of L5 Loop Structure of ConI on Sugar-Binding Ability: Properties of Con-anc-L5

Recombinant Con-anc showed similar properties entirely with ConII rather than ConI in terms of the thermostability and sugar specificity although Con-anc is highly homologous to ConI with 70% identity, indicating that only 31 amino acid residues, which differ between ConI and Con-anc, are involved in the characteristic of ConI (figs 2 and 6A). Here, to identify the structural elements for the strong binding ability of ConI, Con-anc mutant named Con-anc-L5, of which amino acid residues at L5 loop were substituted from M63VNS66 of Con-anc sequence into the corresponding segment of ConI, T63LKGDN68, was constructed. Because the L5 loop is included in the carbohydrate-binding site, of which amino acid residue, D67, bind directly to lactose in Con1, and is structurally different between Con1 and Con2, the L5 loop mutant of Con-anc was prepared. The sugar-binding specificity of Con-anc-L5 was also similar to that of ConII except for the specific oligosaccharides including lacto-N-biosyl (Galβ1-3GlcNAc) or lacto-N-neobiosyl (Galβ1-4GlcNAc) (#41–#48 in fig. 5C). The binding ability of Con-anc-L5 was apparently increased to the lacto-N-biosyl (Galβ1-3GlcNAc) and lacto-N-neobiosyl oligosaccharides, especially, lacto-N-fucopentaose II (LNFP-II), lacto-N-difucohexaose (LNDFH), and A-hepta saccharide (fig. 4C). The thermostability of rCon-anc-L5 was increased as compared with rCon-anc, but still less than that of ConI (fig. 3A).

Discussion

One of the important questions in molecular evolution is how proteins undergo new adaptations to recognize the target molecule and acquire novel functions (functional divergence) in the complex biological networks. Gene duplication is thought to be a major source of new functional genes (Tailor and Raes 2004). Furthermore, rapid adaptive evolution, which is characterized by the higher substitution rates of nonsynonymous to synonymous substitutions, is frequently inferred for duplicated genes. This is because one copy of a duplicate set of genes is often independent of the purifying selection pressures, thereby allowing it to freely or more rapidly mutate the gene to acquire novel functions. Recently, many cases of gene duplication with subsequent rapid adaptive evolution have been identified in various multigene families of biological offense (venomous) and defense systems as well as in reproductive proteins (Ohno et al. 1998; Yang and Bielawski 2000; Ogawa 2006).

In order to understand the molecular evolution, it is important to identify the molecular changes (substitutions) that are responsible for various phenotypic adaptations and/or characteristics of proteins. Although the molecular evolutionary analysis using phylogenetic trees and theoretical predictions provide useful information on how proteins/genes have evolved, there are some direct evidence for the molecular properties of ancestral proteins, which can predict the evolutionary process and the selection pressures during the evolutionary event (Thornton 2004; Williams et al. 2006). To date, some examples of experimental reconstruction studies for ancestral proteins, which have significantly deviated sequences from the current ones rather than a few point mutations, have been reported (Jermann et al. 1995; Gaucher et al. 2003; Thornton et al. 2004; Ugalde et al. 2004; Thomson et al. 2005; Zhang 2006; Bridgham et al. 2006; Watanabe et al. 2006; Ortlund et al. 2007).

In the present study, to understand the mechanism of the rapid adaptive evolution process of congerins, we designed an ancestral congerin named Con-anc corresponding to the putative full amino acid sequence at the divergence of ConI and ConII. Both congerins have diversified via accelerated evolution. Interestingly, the properties of rCon-anc were very similar to those of ConII in terms of thermostability and carbohydrate recognition specificity, although Con-anc shares a higher sequence similarity with ConII (76%) than with ConII (61%). However, remarkable differences were observed between Con-anc and ConII with regard to the sugar-binding specificity to GM3(NeuAc), GM3(NeuGc), and GD1a (fig. 4C), which share the common structure of α2,3-sialyl galactose moiety in their nonreducing termini. The α2,3-sialyltransferase and α2,6-sialyltransferase have been recently isolated and cloned from the pathogenic marine bacteria Vibrio sp. and Photobacterium damselae, respectively, suggesting that α2,3-sialyl galactose–containing sugars, such as GM3(NeuAc), GM3(NeuGc), and GD1a may be specifically present in pathogenic marine bacteria (Yamamoto et al. 1998; Takakura et al. 2006). These observations suggest that the ConII gene involved in the biological defense system has evolved by adaptive rapid evolution from the ancestral gene and has acquired a binding ability that is specific to
FIG. 4.—Affinity constants ($K_a = 1/K_d$) of Con-anc, Con-anc-L5, ConI, and ConII toward 34 PA sugars determined by FAC analysis. (A) Schematic representation of PA oligosaccharides used in FAC. Numbers 1–23: N-glycans and numbers 26–50: glycolipid glycans. (B) Affinity constants for all PA sugars. (C) Expanded scale for PA sugars numbered 26–50.
pathogenic marine bacteria via the recognition of α2,3-sialyl carbohydrates. In other words, the selection pressure on ConII may be these pathogenic marine bacteria that possess α2,3-sialyl galactose–containing sugars.

On the other hand, as shown in figure 3A, ConI exhibited a half-activity retention temperature (51°C) that was ~5°C higher than rCon-anc and rConII (46.5°C) and a stable conformation to urea denaturation compared with rCon-anc and rConII. Furthermore, ConI bound to oligosaccharides with approximately 7- to 9-fold higher affinity than both rCon-anc and rConII, on an average to several tested sugars, to which all congerins bound (fig. 4B). Thus, ConI has evolved from the Con-anc under significant selection pressure to increase the structural stability and to acquire the ability to bind to specific carbohydrates. As reported previously (Shirai et al. 1999), the strand-swap structure and increased protein–ligand interactions in ConI account for these observations. The selection pressure on ConI may have improved the cell coagulation activity and structural stability. The observations from this study and previous studies suggest that the 2 congerin genes have structurally diverged from their ancestor in order to adapt to different tasks.

Figure 6A shows the 3D structure of Con-anc predicted by homology modeling; we compared this structure with those of ConI and ConII. Although Con-anc shares a higher sequence similarity with ConI than with ConII, the properties of Con-anc were very similar to those of ConII in terms of thermostability and carbohydrate recognition specificity, except for the recognition of GM3 and GD1a. The amino acid residues of Con-anc that were identical to those of ConII but different from those of ConI (represented as colored residues of space filling in the ribbon models in fig. 6A) are likely to be responsible for the higher thermostability and the strong and rigid carbohydrate recognition properties of ConI. These replacement sites are dispersed over the protein in either its primary or tertiary structure; however, they involve only 20 amino acid residues, namely, Glu5, Lys12, Val22, Ala27, Ile32, Thr38, Leu60, Val64, Lys77, Pro80, Phe81, Ile91, Thr92, Phe93, Thr95, Leu102, Lys115, Ala118, Asn120, and Phe132 (figs. 2A and 6A). In ConI, several of these sites (Gln5, Phe118, and Pro120) are located in contact with the N-terminal S1 strand, suggesting these residues may

FIG. 6.—Comparative 3D structural analysis of Con-anc and congerins. Comparison of the overall structure (A), the structure around the Ile88 (Thr88 in ConII) residue (B), the MES ligand-binding sites (C), and the L5 loop (D) of Con-anc (red) with ConI (green) and ConII (magenta). The Con-anc structure was predicted by homology modeling using the ConI structure as the template. Bound lactose and MES in ConII are represented by the ball and stick model.
be involved in the strand-swap structure. However, it is not clear whether Con-anc adopts the strand-swap conformation or not. Crystallization and X-ray structural analysis of Con-anc are still in progress. Furthermore, Thr38, Leu60, and Val64 (replaced by Met, Leu, and Val in Con-anc, respectively) along with 2 inserted residues, namely, Asp67 and Asn68, were found in the periphery of the carbohydrate-binding site including the loop structure L5 (figs. 2A and 6D). These residues seem to affect the carbohydrate-binding site. Indeed, the Con-anc-L5, of which amino acid residues at L5 loop were substituted into ConI sequence TLKGDN, showed increased binding activity to the specific sugars containing lacto-N-biosyl (Gal[b1-3GlcNAc] and lacto-N-neobiosyl moieties, especially, LNFP-II, LNDHII, and A-hepta saccharide (fig. 4C). As shown in figure 6O, the local conformation changes due to the inserted amino acid residues, Asn67 and Asp68, were observed in the L5 loop region of Con-anc-L5 as well as ConI as but in the corresponding regions of ConII and Con-anc. In ConI, the L5 loop is sifted close to the bound lactose molecule, and Asp68 forms hydrogen bonding and interacts with Trp70 at a hinge of the loop, which interacts directly with the ligand lactose via stacking interaction, to stabilize its conformation. The sugar-binding activity of ConI (D68N) was greatly reduced by 30–60% in almost all sugars (data not shown), suggesting that the interaction between Trp70 and Asp68 is essential for the strong binding ability of ConI. As shown in figure 3A, Con-anc-L5 also showed higher thermostability than Con-anc and ConII. This indicates the L5 loop including inserted amino acid residues, Asn67 and Asp68, can stabilize the conformation of Trp70 residue, which is essential for the interaction with carbohydrates, by interactions. Thus, the interaction between Trp70 and Asp68 is essential for the strong binding ability of ConI. This is also supported by the result from ConI mutant, ConI(D68N), of which carbohydrate-binding activity was reduced by 30–60% in almost all sugars (data not shown). We have recently found that the thermostability of ConII was improved by in vitro evolutionary protein engineering and that 2 amino acid mutations at Tyr16 and Thr89 in ConII to the Phe and Ile residues, which correspond to ConI, participate in their thermostable properties (Shionyu-Mitsuyama et al. 2005). The thermostability of the ConII (Y16F/T89I) in which both sites were replaced with those of ConI was higher than that of the wild-type ConII and was comparable with that of ConI. Based on this observation, it was expected that Con-anc, in which the corresponding amino acid residues are Phe16 and Ile89, might show higher stability. However, Con-anc showed lower thermostability than ConI. This discrepancy may be due to the effect of substitutions of other amino acid residues that lie adjacent to the Phe16 and Ile89 residues (within 4.5 Å), for example, Glu101 (fig. 4B). The crystal structures of the most thermostable double mutant of ConII (Y16S/T89I) revealed that significant contribution to the stabilization of congersins appeared to be the removal of the interior solvent molecules and the hydrophobic packing of the cavity that remained after water expulsion (Shionyu-Mitsuyama et al. 2005). In the predicted Con-anc structure, the hydrogen bonding and the hydrophobic packing patterns around the Ile89 residue were different from those of ConI and thermostable mutants (fig. 6B). X-ray structural analysis of Con-anc is needed to discuss more clearly the relationships between the structure and thermostability of congersins.

On the other hand, the comparison of ConII and Con-anc revealed several replaced residues (Arg3, Ile41, and Tyr123) at a region close to the carbohydrate-binding site; this may be related to the ability of ConII to bind to the α2,3-sialyl galactose moieties in GM3 and GD1a (fig. 6C). A comparative mutagenesis study of ConII indicated that the replacement of Arg3 and Tyr123 residue of ConII with Gly and Ser, respectively, decreased its binding activity to GM3, whereas the replacement of Ile41 with Leu showed no apparent effect on the GM3-binding activity (fig. 5). This suggests that Arg3 and Tyr123 residues are ones of the key residues for the recognition of α2,3-sialyl galactose moieties. These data confirmed that ConII has evolved via accelerated evolution under significant selection pressure to acquire the binding activity to specific carbohydrates such as GM3 and GD1a. Thus, the congerin genes are excellent examples for the study of the generation of new functional genes by natural selection following gene duplication.

In conclusion, protein engineering using a probable ancestral form based on phylogenetic trees is a powerful approach to determine the evolutionary process and natural selection pressures such as protein stabilization and the strong carbohydrate-binding ability for ConI and carbohydrate-recognition specificities to marine pathogens for Con-II. Furthermore, this method could be a useful and efficient tool for analyzing the structure–activity relationships of proteins. To our knowledge, this is the first report of the probable ancestral form encoding for the full-length sequence of predicted ancestor protein.

Supplementary Material

Supplementary figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, 14360045 to T.O., and Japan Society for the Promotion of Science research fellowship to A.K.

Literature Cited


Takashi Gojobori, Associate Editor

Accepted August 30, 2007