

Chapter 2

Domain Shuffling and the Evolution of Vertebrate Extracellular Matrix

Hiroshi Wada

Abstract Contribution of domain shuffling events in the evolution of the vertebrate extracellular matrix (ECM) is reviewed. In some cases, domain shuffling contributed to the evolution of molecules via minor modification of existed molecules by fusing a domain to existing proteins. In contrast, some ECM molecules such as aggrecan and fibronectin seem to be created rather de novo via domain shuffling. In the case of occludin, a domain of completely different function (ELL: elongation of RNA synthesis) was recruited for use in cell adhesion molecules. These cases are spectacular examples of molecular tinkering.

2.1 Introduction

Evolution of the extracellular matrix (ECM) has played a major role in the evolution of multicellularity of animals. Many components of the ECM, including laminin and type IV collagen, were acquired in common ancestors of metazoans (Patthy 2003). Since then ECMs have continued to evolve. The evolution of ECM molecules contributed to the diversification of various metazoans (Patthy 1999, 2003).

Because most ECM molecules show multidomain architectures, the evolution of ECM molecules likely progressed through two steps, namely, fixation of functional amino acid domains (domain emergence) and subsequent shuffling of functional domains (domain shuffling). These mechanisms have generated the variety of functional ECM molecules found today. Ekman et al. (2007) indicated that the evolutionary rate of domain emergence was relatively constant during eukaryotic evolution or possibly relatively high in eukaryotic ancestors prior to the divergence

H. Wada (✉)

Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
e-mail: hwada@biol.tsukuba.ac.jp

of plants, fungi, and metazoans. However, domain shuffling has accelerated in metazoan lineages.

Duplication divergence is one of the most widely accepted mechanisms for the molecular evolution of domain emergence. According to this theory, novel functional sequences were established by gene duplication followed by divergence of amino acid sequences. In cases where one of the duplicated genes performs an ancestral function, the other copy is typically under less functional constraint and thus free to fix distinct amino acid sequences. On the other hand, the recent accumulation of genomic data may increase our understanding of the other aspect of gene evolution. The genome sequences analyzed to date contain a large number of genes that show no homology to any genomes of other species. Some authors such as Schmidt and Davies (2007) and Tautz and Domazet-Loso (2011) proposed that these sequences emerged from noncoding genome sequences and that novel functional amino acid sequences continue to emerge.

In this chapter we focus on the second step of ECM molecular evolution, namely, domain shuffling. Our previous research identified several novel ECM molecules produced via domain shuffling in vertebrate ancestors (Kawashima et al. 2009). Our basic strategy to evaluate domain shuffling events was as follows. First, novel domain architectures in gene models were evaluated by listing domain pairs found in single-gene models of several representative genomes. Using this method, we identified 47 vertebrate genes created by domain shuffling. These genes possessed domain pairs conserved in all vertebrate species but were not found in any invertebrate genomes. We found that most of the domain shuffling events were likely due to exon shuffling, because most of the new domain pairs were split by intron(s) (Kawashima et al. 2009), and nice correlation was observed between exon and domain boundaries (Liu and Grigoriev 2004). Among the exons involved in domain shuffling, a number were split by phase 1 introns at both the 5' and 3' termini (exons of intron phase type 1-1), while 0-0 phase exons were abundant in all exons (Patthy 1999; Kawashima et al. 2009). The mechanism behind this type of shuffling remains unclear but Patthy (1999) suggested that it is due to the larger number of exons encoding extracellular protein domains encoded by 1-1 phase exons. Below we review how these novel genes were created by domain shuffling and how they contributed to evolution of the vertebrate body.

2.2 Lecticans and Evolution of Vertebrate Cartilage

Cartilage, one of the most important features of vertebrates, was established through ECM evolution. The most abundant component of vertebrate cartilage is fibrillar collagen, and gene duplication of fibrillar collagens has contributed to the formation of various vertebrate skeletal structures (Wada et al. 2006). An interesting exception is agnathan cartilage, whose main component is elastin-like molecules (Wright and Youson 1983; Robson et al. 1993, 1997); therefore, the evolutionary relationship between agnathan and gnathostomes skeletons remains

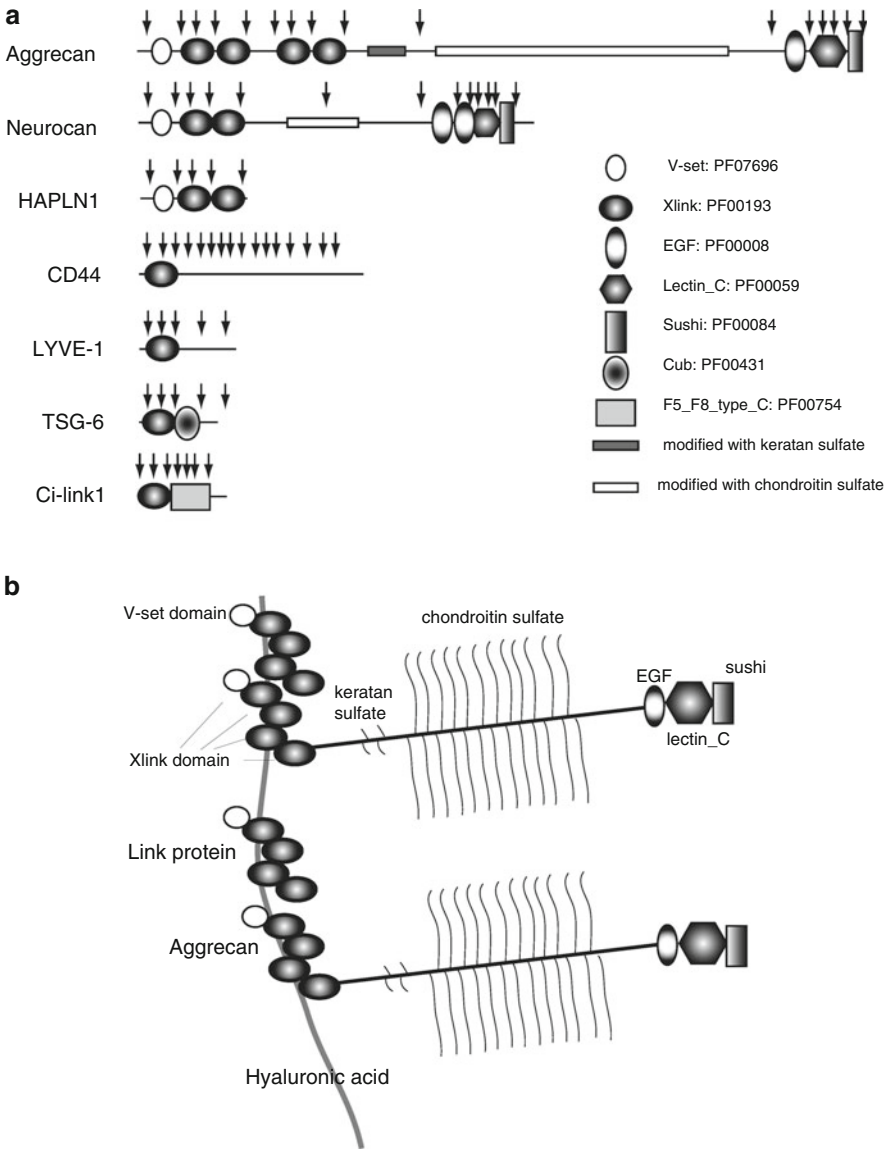


Fig. 2.1 Schematic illustration of the domain architecture of aggrecan. **(a)** Comparison of the domain architecture to proteins with Xlink domain. **(b)** Schematic illustration of the domain architecture of aggrecan and HAPLN and their interactions with hyaluronic acid

unclear (McCauley and Bronner-Fraser 2006; Zhang et al. 2006; Ohtani et al. 2008). Aggrecan is the most abundant non-collagenous protein component of gnathostome cartilage. The aggrecan domain architecture is shown in Fig. 2.1a; the domain pair V-set and Xlink is novel in vertebrates. This domain pair is also

found in other lecticans, such as neurocan and bevicin, as well as in HAPLNs (Fig. 2.1a). By examining each domain in the amphioxus and ascidian genomes, we found that these domain architectures were established *de novo*. The V-set, Lectin C, EGF, and Sushi domains were abundant in both genomes. For example, V-set domains were found in 269 and 26 gene models of the amphioxus and ascidian genomes, respectively, and Sushi domains were found in 365 and 61 gene models. In contrast, 20 amphioxus gene models and 2 ascidian gene models contain Xlink domains. One of the ascidian gene models (*Ci-link1*) consisted of a single Xlink domain and the F5/8 type C domain (Fig. 2.1a). In vertebrates, the Xlink domain was also found in CD44, TSG-6, LYVE-1, and stabilin-1. These genes contain single Xlink domains with architectures similar to *Ci-link1* (Fig. 2.1a).

The vertebrate molecules with single Xlink domains are involved in lymphocyte migration (Prevo et al. 2001; Cichy and Pure 2003; Ponta et al. 2003; Lesley et al. 2004), and *Ci-link1* is expressed in some juvenile blood cells (Kawashima et al. 2009; Yoneda et al. 2010). These observations suggest that the Xlink domain is a component of the surface molecules of blood cells in protochordate ancestors. The Xlink domain then combined with another component, such as the V-set domain in a vertebrate ancestor, and was recruited as a component of cartilage, a novel vertebrate structure (Kawashima et al. 2009; Yoneda et al. 2010).

This scenario introduced other aspects of the evolution of the Xlink domain. The Xlink domain functions with the aggrecan molecule as the interface with hyaluronan (HA) (Fig. 2.1b). The complex consisting of HA, link protein/HAPLN1, and aggrecan provides tensile strength for cartilage to absorb shock and resist compression in the joint. Mutations in aggrecan or the link protein/HAPLN1 cause severe defects in cartilage and result in dwarfism (Watanabe et al. 1994; Watanabe and Yamada 1999). Although Xlink is present in various metazoans, including nematodes, no HA had been identified in invertebrates (Yoneda et al. 2010). Indeed, hyaluronan synthase (HAS) was believed to have been acquired in vertebrate ancestors via horizontal gene transfer (Salzberg et al. 2001). Our recent studies of HAS have revealed that amphioxus also possesses HAS genes, and thus HAS was perhaps acquired in the common ancestors of chordates (Yoneda et al. 2010). On the other hand, ascidians do not possess HA, and the ascidian Xlink domain binds heparin (Yoneda et al. 2010).

The above observations reveal rather dramatic evolutionary events during the establishment of the molecular complex of Aggrecan and HA for important cartilage components (Yoneda et al. 2010). Chordate ancestors acquired HAS possibly by horizontal gene transfer, but their Xlink domain was not utilized for HA binding, instead functioning as a heparan-binding domain. In the common ancestors of vertebrates, Xlink acquired binding specificity against HA. Subsequently, through domain shuffling, genes encoding lecticans were established *de novo*. After gene duplications in the common ancestors of vertebrates (perhaps two rounds of genome duplications), one of the lectican genes is utilized as a cartilage matrix component.

Another novel molecule, cartilage acidic protein 1 precursor, was recruited as cartilage matrix (Kawashima et al. 2009). This molecule was established via minor

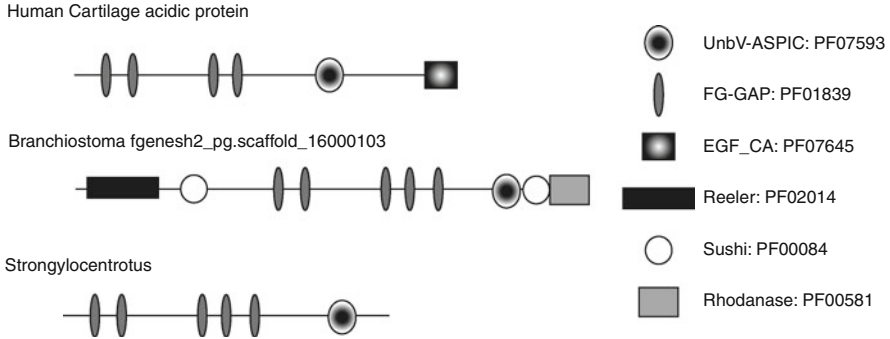


Fig. 2.2 Schematic illustration of the domain architecture of the cartilage acidic protein. Amphioxus (*Branchiostoma*) and sea urchin (*Strongylocentrotus*) possess genes with similar domain architecture, but they lack EGF_CA domain in their C-terminus

modifications of the domain architecture. Amphioxus and sea urchins possess genes of similar domain architectures with FG-GAP domain and UnbV-ASPIC repeats but lack EGF_CA (Fig. 2.2). The combination of the UnbV-ASPIC and FG-GAP domains may have occurred in common ancestors of bilaterians, and subsequently this gene acquired the EGF_CA domain in ancestral vertebrates, which allowed it to function as the cartilage acidic protein.

2.3 Domain Shuffling and Evolution of the Tight Junction

The tight junction exists only in vertebrates and ascidians, and the major component of the tight junction, claudin, is found in these species (Kollmar et al. 2001). We previously demonstrated that the other tight junction component, occludin (Tsukita et al. 2001), is found only in vertebrates (Kawashima et al. 2009). The combination of the MARVEL domain and occluding ELL domains is specific to vertebrates (Fig. 2.3a). Amphioxus and ascidian possess the occludin_ELL domain in RNA polymerase II elongation factor ELL (Fig. 2.3b). Thus, it is likely that, after genome duplication in the vertebrate lineage, one of the paralogs was combined with the MARVEL domain and the occludin protein was created. Seventeen amphioxus genes and four ascidian genes contain the MARVEL domain, and all of the genes from amphioxus and ascidians contain only the MARVEL domain (Fig. 2.3a, b). The structures of protochordate genes containing the MARVEL domain are more similar to MAL and physins involved in vesicle trafficking, in which the MARVEL domain is thought to be involved in cholesterol-rich membrane apposition events (Sánchez-Pulido et al. 2002) (Fig. 2.3a). The protochordate genes containing the MARVEL domain might perform roles similar to MAL or physins. Subsequently, in an ancestral vertebrate, a gene acquired the occludin_ELL domain and was committed to a novel function as a component of the tight junction.

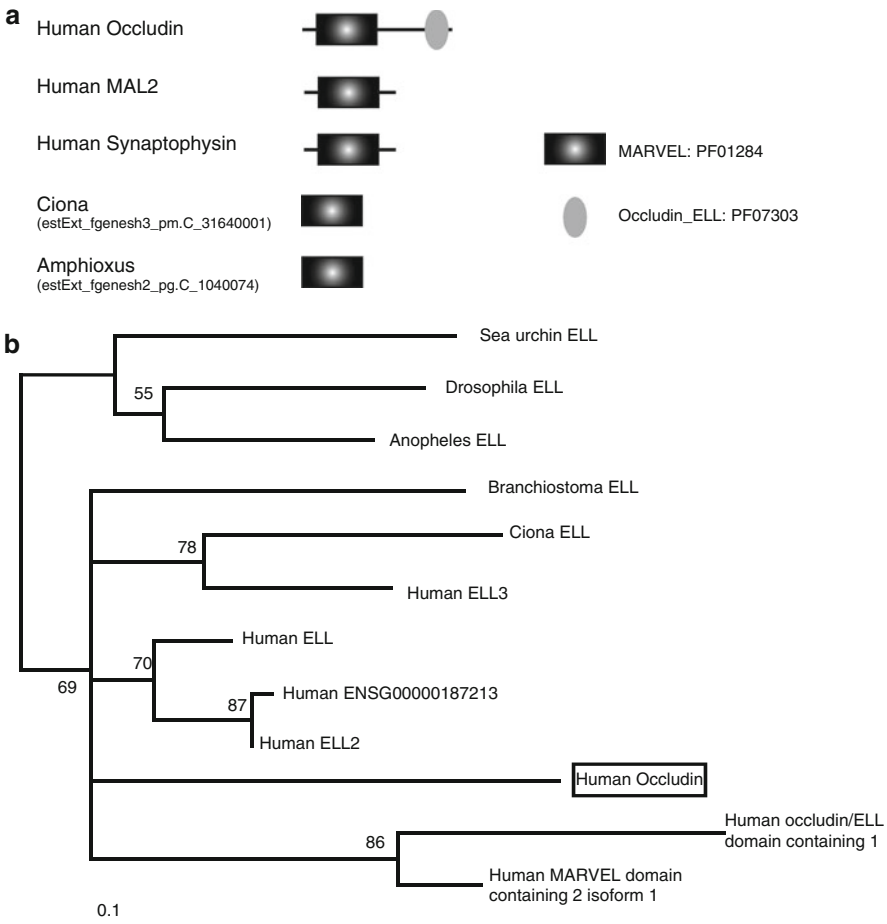


Fig. 2.3 Domain architecture of occludin and evolutionary origin of ELL domain. **(a)** Comparison of domain architecture of occludin to some proteins with MARVEL domain. **(b)** Maximum likelihood phylogenetic tree of occludin_ELL domains. Numbers on the nodes show bootstrap values

Occludin is not an essential component of the tight junction because occludin-deficient knockout mice lack any detectable barrier defects (Saitou et al. 2000; Schulzke et al. 2005). Rather, occludin is thought to be involved in a regulatory aspect of the tight junction such as mediating extracellular signals for cell adhesion (Saitou et al. 2000). Because the *Ciona* genome lacks occludin, it is likely that the tight junction did not contain occludin when it originated in the common ancestor of ascidians and vertebrates. Upon acquiring occludin as a novel component of the tight junction, vertebrates may have used the junction in a more flexible manner.

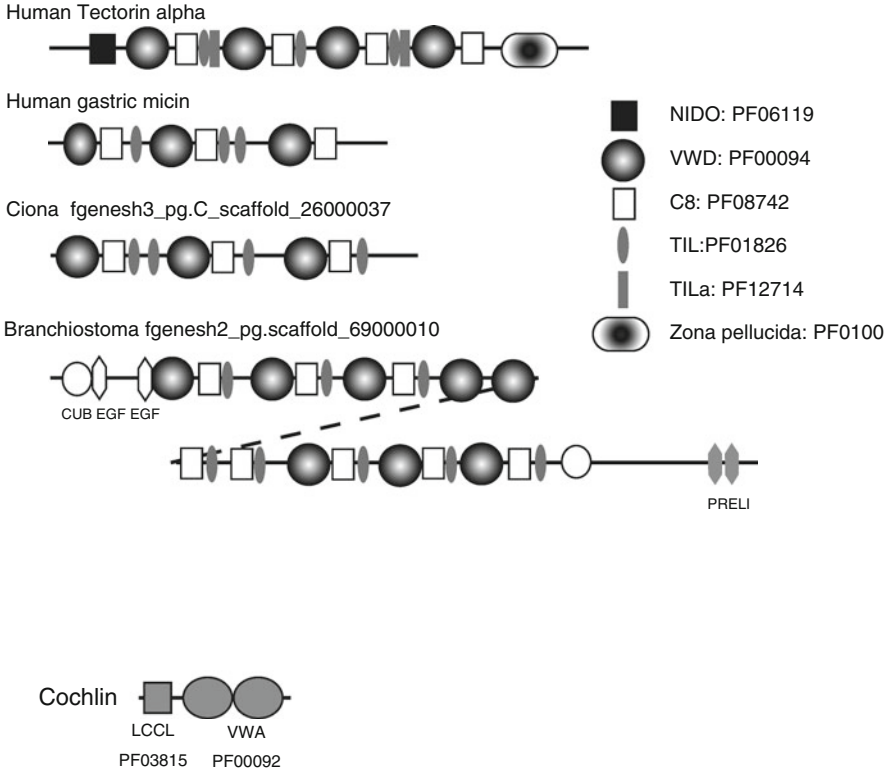


Fig. 2.4 Domain architecture of tectorin-alpha and cochlin. Ascidian (*Ciona*) and amphioxus (*Branchiostoma*) also possess proteins with similar domain architecture of VWD-C8-TIL repeats

2.4 Domain Shuffling and Evolution of the Vertebrate Auditory System

Two new genes established by domain shuffling contributed to the evolution of the vertebrate auditory system (Kawashima et al. 2009). Tectorin-alpha is one of the major components of the tectorial membrane in the mammalian inner ear. In humans, a missense mutation in tectorin-alpha causes hearing loss of about 50–80 dB (Verhoeve et al. 1998). Tectorin-alpha consists of three or four repeats of vWD-C8-TIL domains with N-terminal NIDO domain and C-terminal zona_pellucida-like domain (Fig. 2.4). Domain combinations between the zona pellucida-like and NIDO domains and between the NIDO and TIL are only observed in vertebrate genomes. The repeats of vWD-C8-TIL domains were found in mucins, and protochordates possessed genes with similar domain architectures (Fig. 2.4). Therefore, ancestral molecules may have functioned as

mucin, and the novel molecule recruited as a component of the tectorial membrane fused with NIDO and zona pellucida domains.

Cochlin is also an essential component of the vertebrate auditory system, and a mutation in *cochlin* causes deafness (Robertson et al. 2006). Cochlin shows domain architecture of the LCCL with the repeat of VWA domain, whose combinations are found only in vertebrates (Fig. 2.4) (Kawashima et al. 2009). Thus, this molecule was apparently established de novo. These observations clearly indicate that domain shuffling played critical roles in the evolution of vertebrate auditory systems.

2.5 Other ECM Molecules Established in Vertebrates

Fibronectin and vitronectin also show domain architectures unique to vertebrates (Kawashima et al. 2009). Fibronectins consist of three types of domain repeats: FN1, FN2, and FN3. FN3 repeats were found in the tyrosine phosphatase gene of various metazoans including *Drosophila* or *Hirudo*. FN1 and FN2 repeats were also found in respective amphioxus genes (Fig. 2.5). Therefore, fusions of these repeats in the vertebrate ancestors established fibronectin as a novel ECM molecule.

Vitronectin consists of Hemopexin domain repeats with an N-terminal somatomedin domain (Fig. 2.6). Hemopexin domain repeats were found in several metalloproteinases, including MMP9 (Fig. 2.6). Vitronectin may have originated from one of these metalloproteinases genes by fusing with the somatomedin domain.

2.6 Future Prospects

We previously performed a comprehensive survey of molecules established by domain shuffling in the common ancestors of vertebrates (Kawashima et al. 2009). However, this list may remain incomplete. As more genome sequences are determined, we will identify other ECM molecules or domain pairs thought to be exclusive to vertebrates in invertebrate species. Indeed, our further analyses revealed two ECM molecules (fibulin and nidogen) thought to be exclusive to vertebrates in other invertebrates. There are an unexpected number of genes found in vertebrates and cnidarians, but not in genomes of *Drosophila* or *C. elegans*, indicating that multiple gene loss occurs frequently (Putnam et al. 2007). Also, incomplete gene model prediction may hinder the identification of true domain pairs. Prediction of gene models remains one of the most technically difficult issues in genome analyses. Although several algorithms for gene model prediction are available, gene model prediction is dependent on transcriptome data. Thus, the list of domain pairs in single-gene models is not complete.

Human Fibronectin

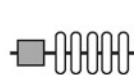


Branchiostoma fgenes2_pg.scaffold_467000013



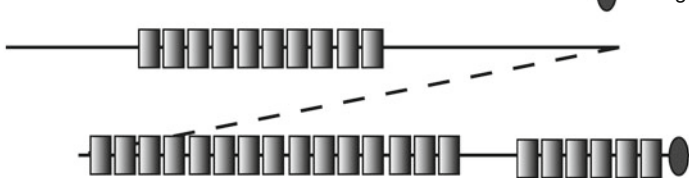
FN1: PF00039

Branchiostoma fgenes2_pg.scaffold_508000014



FN2: PF00040

Branchiostoma estExt_fgenes2_pg.C_2330030



FN3: PF00041

Lectin_C: PF00059

Kringle: PF00051

Branchiostoma estExt_gwp.C_1010065



Drosophila Tyrosine phosphatase AAA28484



Y_phosphatase: PF00102

I-set: PF07679

Fig. 2.5 Domain architecture of fibronectin. Amphioxus (*Branchiostoma*) possess FN1 repeat, FN2 repeat, and FN3 repeat in different proteins

Human Vitronectin



Human MMP9



Branchiostoma estExt_fgenes2_pg.C_490126



Somatomedin_B: PF01033

Hemopexin: PF00045

Peptidase_M10: PF00413

PG_binding_1: PF01471

Fig. 2.6 Domain architecture of vitronectin. The repeat of hemopexin was found in MMP9, and similar domain architecture is found in some amphioxus genes

While keeping these incompleteness in mind, we found that domain shuffling contributed to the evolution of molecules, not only via minor modification of molecules by fusing a domain to existing proteins, but also by producing a novel molecule *de novo*, such as aggrecan and fibronectin. In the case of occludin, a domain of completely different function (ELL: elongation of RNA synthesis) was recruited for use in cell adhesion molecules. This may be a spectacular example of molecular tinkering.

It is also important to understand how novel protein domains are established during evolution. Unique domains continue to emerge, and there are 278 domains identified to be unique to vertebrates (Ekman et al. 2007). It was widely believed that most of these new domains were the result of replication divergence. However, recent reviews by Schmidt and Davies (2007) and Tautz and Domazet-Łoso (2011) make the point that novel sequence may have emerged from noncoding genomic sequences. Further analyses of the mechanistic aspects of gene evolution may reveal dynamic aspects of molecular evolution.

References

- Cichy J, Pure E (2003) The liberation of CD44. *J Cell Biol* 161:839–843
- Ekman D, Bjorkund A, Elofsson A (2007) Quantification of the elevated rate of domain rearrangements in Metazoa. *J Mol Biol* 372:1337–1348
- Kawashima T, Kawashima S, Tanaka C, Murai M, Yoneda M, Putnam NH, Rokhsar DS, Kanehisa M, Satoh N, Wada H (2009) Domain shuffling and the evolution of vertebrates. *Genome Res* 19:1393–1403
- Kollmar R, Nakamura SK, Kappler JA, Hudesbeth AJ (2001) Expression and phylogeny of claudins in vertebrate primordia. *Proc Natl Acad Sci USA* 98:10196–10201
- Lesley J, Gal I, Mahoney DJ, Cordell MR, Rugg MS, Hyman R, Day AJ, Mikecz K (2004) TSG-6 modulates the interaction between hyaluronan and cell surface CD44. *J Biol Chem* 279:25745–25754
- Liu M, Grigoriev A (2004) Protein domains correlate strongly with exons in multiple eukaryotic genomes—evidence of exon shuffling? *Trends Genet* 20:399–403
- McCauley DW, Bronner-Fraser M (2006) Importance of SoxE in neural crest development and the evolution of the pharynx. *Nature* 441:750–752
- Ohtani K, Yao T, Kobayashi M, Kusakabe R, Kuratani S, Wada H (2008) Expression of Sox and fibrillar collagen genes in lamprey larval chondrogenesis with implications for the evolution of vertebrate cartilage. *J Exp Zool* 310B:596–607
- Pathy L (1999) Genome evolution and the evolution of exon-shuffling—a review. *Gene* 238:103–114
- Pathy L (2003) Modular assembly of genes and the evolution of new functions. *Genetica* 118:217–231
- Ponta H, Sherman L, Herrlich PA (2003) CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol* 4:33–45
- Prevo R, Banerji S, Ferguson DJP, Clasper S, Jackson DG (2001) Mouse LYVE-1 is an endocytic receptor for hyaluronan in lymphatic endothelium. *J Biol Chem* 276:19420–19430
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94

- Robertson N, Hamaker S, Patriub V, Aster J, Morton C (2006) Subcellular localisation, secretion, and post-translational processing of normal cochlin, and of mutants causing the sensorineural deafness and vestibular disorder, DFNA9. *J Med Genet* 40:479–486
- Robson P, Wright G, Sitarz E, Maiti A, Rawat M, Youson J, Keeley F (1993) Characterization of lamprin, an unusual matrix protein from lamprey cartilage. Implications for evolution, structure, and assembly of elastin and other fibrillar proteins. *J Biol Chem* 268:1440–1447
- Robson P, Wright GM, Youson JH, Keely FW (1997) A family of non-collagen-based cartilages in the skeleton of the sea lamprey, *Petromyzon marinus*. *Comp Biochem Physiol* 118B:71–78
- Saitou M, Furuse M, Sasaki H, Schulzke J-D, Fromm M, Takano H, Noda T, Tsukita S (2000) Complex phenotype of mice lacking occludin, a component of tight junction strands. *Mol Biol Cell* 11:4131–4142
- Salzberg SL, White O, Peterson J, Eisen JA (2001) Microbial genes in the human genome: lateral transfer or gene loss. *Science* 292:1903–1906
- Sánchez-Pulido L, Martín-Belmonte F, Valencia A, Alonso MA (2002) MARVEL: a conserved domain involved in membrane apposition events. *Trends Biochem Sci* 27:599–601
- Schmidt EE, Davies CJ (2007) The origin of polypeptide domains. *Bioessays* 29:262–270
- Schulzke JD, Gitter AH, Mankertz J, Spiegel S, Seidler U, Amasheh S, Saitou M, Tsukita S, Fromm M (2005) Epithelial transport and barrier function in occludin-deficient mice. *Biochim Biophys Acta* 1669:34–42
- Tautz D, Domazet-Loso T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12 (10):692–702
- Tsukita S, Furuse M, Itoh M (2001) Multifunctional strands in tight junctions. *Nat Rev Mol Cell Biol* 2:285–293
- Verhoeve K, Van Laer L, Kirschhofer K, Legan P, Hughes D, Schatteman I, Verstreken M, Van Hauwe P, Coucke P, Chen A, Smith R, Somers T, Offeciers F, Van de Heyning P, Richardson G, Wachtler F, Kimberling W, Willems P, Govaerts P, Van Camp G (1998) Mutations in the human alpha-tectorin gene cause autosomal dominant non-syndromic hearing impairment. *Nat Genet* 19:60–62
- Wada H, Okuyama M, Satoh N, Zhang S (2006) Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol Dev* 8:370–377
- Watanabe H, Yamada Y (1999) Mice lacking link protein develop dwarfism and craniofacial abnormalities. *Nat Genet* 21:225–229
- Watanabe H, Kimata K, Line S, Strong D, Gao L, Kozak C, Yamada Y (1994) Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. *Nat Genet* 7:154–157
- Wright GM, Youson JH (1983) Ultrastructure of cartilage from young adult sea lamprey, *Petromyzon marinus* L: a new type of vertebrate cartilage. *Am J Anat* 167:59–70
- Yoneda M, Nakamura T, Murai M, Wada H (2010) Evidence for the heparin-binding ability of the ascidian Xlink domain and insight into the evolution of the Xlink domain in chordates. *J Mol Evol* 71:51–59
- Zhang G, Miyamoto MM, Cohn MJ (2006) Lamprey type II collagen and *Sox9* reveal an ancient origin of the vertebrate collagenous skeleton. *Proc Natl Acad Sci USA* 103:3180–3185

Evolution of Extracellular Matrix

Keeley, F.W.; Mecham, R. (Eds.)

2013, X, 290 p.,

ISBN: 978-3-642-36002-2