REVIEW

Correlating structure and function during the evolution of fibrinogen-related domains

Russell F. Doolittle,* Kyle McNamara, and Kevin Lin

Departments of Chemistry & Biochemistry and Molecular Biology, University of California, San Diego, La Jolla, California 92093-0314

Received 30 August 2012; Revised 4 October 2012; Accepted 5 October 2012

DOI: 10.1002/pro.2177

Published online 17 October 2012 proteinscience.org

Abstract: Fibrinogen-related domains (FReDs) are found in a variety of animal proteins with widely different functions, ranging from non-self recognition to clot formation. All appear to have a common surface where binding of one sort or other occurs. An examination of 19 completed animal genomes—including a sponge and sea anemone, six protostomes, and 11 deuterostomes—has allowed phylogenies to be constructed that show where various types of FReP (proteins containing FReDs) first made their appearance. Comparisons of sequences and structures also reveal particular features that correlate with function, including the influence of neighbor-domains. A particular set of insertions in the carboxyl-terminal subdomain was involved in the transition from structures known to bind sugars to those known to bind amino-terminal peptides. Perhaps not unexpectedly, FReDs with different functions have changed at different rates, with ficolins by far the fastest changing group. Significantly, the greatest amount of change in ficolin FReDs occurs in the third subdomain (“P domain”), the very opposite of the situation in most other vertebrate FReDs. The unbalanced style of change was also observed in FReDs from non-chordates, many of which have been implicated in innate immunity.

Keywords: fibrinogen-related domains; ficolin; angiopoietin; tenascin; fibroleukin; hepassocin; microfibril associated glycoprotein; evolution; phylogenomics

Abbreviations: Fn3 domain, fibronectin type III domain; FReD, fibrinogen-related domain; FReP, fibrinogen-related protein; MFAgp, microfibril-associated glycoprotein; NCBI, National Center for Biotechnology Information.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Russell F. Doolittle, Department of Chemistry and Biochemistry, Univ. Calif., San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0314. E-mail: rdoolittle@ucsd.edu.

Grant sponsor: Academic Senate, University of California, San Diego.

Introduction

Fibrinogen-related domains (FReDs) take their name from the globular portions of vertebrate fibrinogen molecules where they were first recognized.¹ Homologs have since been found in a wide variety of proteins from vertebrates, including tenascins (initially called cytactin² or hexabrachion³), ficolins,⁴ angiopoietins,⁵ fibroleukin,⁶ and some other extracellular proteins.^{7–9} They are also found in nonvertebrate animals of all kinds.^{10–14} Unlike the eponymous fibrinogen molecule—where the domains are directly involved in the formation of

fibrin clots—most of these proteins are engaged in interactions with cell surfaces, either of host or alien origin (Table I).

Reviews dealing with particular kinds of FReP (proteins containing FReDs) have appeared over the years, including several recently published.^{15–17} Here, we report a comprehensive analysis of the phylogenetic distribution of all kinds of FReD, with special attention being paid to structural modifications that correlate with changes in function as new domain combinations emerged during the course of evolution.

Defining FReDs

When FReDs were first described 20 years ago,¹ it was recognized that the term “domain” was being used casually and that FReDs actually consist of definable substructures. Later, the first X-ray determination of a fibrinogen domain to be reported,¹⁸ the γ C domain, showed that there are three distinguishable subdomains, referred to as A, B, and P (Fig. 1). The subdomains are often—but not as a rule—encoded on separate exons.

In the interval since that initial report, crystal structures have been determined for a variety of FReDs, including the β C and α 'C domains from fibrinogen,^{19,20} an angiopoietin,²¹ three ficolins,^{22,23} and a tachylectin from the horseshoe crab.²⁴ All are globular entities built around a central five-stranded beta sheet.

The amino-terminal subdomain (“A domain”) amounts to about 50 residues and is characterized by a disulfide bond, the cysteines of which are typically separated by 28 ± 2 residues. The second and third subdomains are intimately associated, the middle strand of the central β sheet that constitutes the second subdomain actually being provided by a segment that appears sequentially after the third subdomain. The two subdomains together amount to about 150 residues, although there is considerable variation in the carboxyl-terminal region that occurs after the unusually positioned β strand.

Importantly, the third (carboxyl-terminal) subdomain contains a unique disulfide bond, the cysteines of which almost always have 12 residues between them. Moreover, in all but one of the structures that have been determined (upwards of two dozen for eight different kinds of FReD studied)^{18–24} the peptide bond between the second cysteine and its amino-terminal neighbor is *cis* (in the single exception, under some circumstances a *cis-trans* interconversion can take place²³). All known structures contain a bound calcium ion in the vicinity of this segment.

Typically, this subdomain has a recognizable binding site in this same region—a well sculpted “hole” in some—where the known interactions with putative ligands occur. In this regard, many FReDs

Table I. FRePs Found Among Chordates: Properties and Functions

FReP ^a	Length ^b	Function ^c	Neighbor domains ^d	Where made ^e	First appearance
MFApp	255	Microfibril association	(none)	Aorta, vascular system	Bony fish
Fibroleukin	439	Inflammatory response	Uncharacterized, conserved	Leukocytes	Protochordates
FIBCD1	461	Chitin binding	Membrane spanner, coiled coil	Intestinal wall	Agnatha
Ficolins	288–326	Innate immunity	Collagen segments	Liver, lung, leukocytes	Protochordates, amphibia ^f
Tenascins	1299–4242	cell–cell interaction, extracellular matrix	EGF, FN3 domains	Fibroblasts	Protochordates
Angiopoietins	496–503	Angiogenesis	Coiled coils (two-stranded)	Vascular system	Agnatha
Angioarrestins	346–493	Anti-angiogenesis	Coiled coils (two-stranded)	Adrenal, placenta, etc.	Protochordates
Fibrinogen	430–870	Fibrin clot formation	Coiled coils (three-stranded)	Liver	Protochordates ^g
Hepassocin	312	Liver cell mitogen	putative coiled coil	Liver	Agnatha

^a FReP = fibrinogen-related protein.

^b Length of proteins in humans (including signal peptides).

^c Also called angiopoietin-like proteins.

^d Domains or segments on amino-terminal side.

^e Not all sites of expression listed.

^f Independent evolution.

^g Protofibrinogen.

Figure 1. Backbone structure of a typical FReD (human H ficolin, PDB 2J5Z) showing three subdomains (A, violet; B, brown; P, blue). The amino-terminal (N) is usually preceded by one of a variety of connecting domains of various sorts. Interactions usually occur on the outer face of the third subdomain ("P domain"). Note how the central strand of the main sheet (green) is provided by a sequence segment that occurs after the third subdomain.

are designated as lectins and are known to bind sugars or sugar derivatives, including in some cases components of substances found on bacterial surfaces. In marked contrast, the holes in vertebrate fibrinogen "FReDs" bind amino-terminal peptide "knobs" emanating from the central regions of other fibrinogen molecules and account for the initial step in fibrin polymerization.

As far as is known, when they occur in the company of other domains, FReDs are always at the carboxyl ends of polypeptide chains. Past reports of FReDs occurring as internal sequences have been found to be the results of mistakes in the automated prediction of exon splicing, and especially the omission of exons from adjacent genes.^{25,26}

Other criteria can be used to categorize FReDs besides sequence-based phylogenetic trees, including the kinds of domain that occur as their nearest neighbor on the amino-terminal side. These domains, when they occur, can be two- or three-stranded coiled coils, simple tethers, fibronectin type III domains (FN3), or collagen domains; many are uncharacterized, simple connectors without homology to other known proteins. For the most part, sequence-based phylogenies of the FReDs themselves are in accordance with the occurrence of the neighboring segments; additionally, specific and easily recognized motifs are often associated with particular FReD types.

As for function, neighbor domains play key roles in presenting the interactive surfaces of FReDs. For example, the tight triple-helix nature of a collagen segment insures that the three terminal FReDs are packed together as a symmetrical homotrimer, a feature with obvious advantage for binding to surfaces

with numerous, uniformly distributed ligands, as might be expected to occur on the cell walls of bacteria. Coiled-coils are also important in how FReDs are displayed, the two-stranded sort leading to side-by-side FReDs that can bind to and bring together diffusible entities on cell surfaces. The oligomeric association of FReDs can also occur independent of and in the absence of adjacent domains.

The approach

Our study began with an inventory of FReDs in various genome databases. The current version of the SMART database²⁷ lists 2380 occurrences of FReDs in 2354 proteins, 35 of which allegedly occur in bacteria. The Pfam database²⁸ lists 3020 occurrences, including 28 in bacteria, and Superfamily²⁹ reports 3479 FReDs in 3431 FRePs, bacterial entries not specified. These are misleading tallies, however, because virtually all of the non-animal occurrences consist of only the amino-terminal subdomain of FReDs, which in other instances is known as a NEC domain.³⁰ The NEC (an acronym based on the neuro and collagenous environments in which they frequently occur) domain corresponds exactly to the A subdomain of a FReD. NEC domains are more widespread than FReDs and are found not only in animals but also in choanoflagellates, some other protists, and a restricted group of bacteria.*

The functions of NEC domains in those other settings have no bearing on the subjects being considered here, and in this article—unless stated otherwise—FReDs consist of all three subdomains. All told, we identified 564 full-length FReDs in 19 representative genomes. A full listing of FRePs used in this study, complete with their NCBI designations, is provided in the Supporting Information (Supporting Information Table S1).

When the inventory was complete, a phylogeny was constructed of the 23 FReDs found in the human genome, all of which have a history of careful study by actual experiment. The tree of human FReDs was then used as a template for assigning FReDs found in other chordate genomes. Consideration was also given to amino-terminal neighboring domains (Table I).

Results

FRePs in vertebrates

The human genome encodes 23 different, full-sized FReDs, including (a) four tenascins, (b) three ficolins, (c) three angiopoietins, (d) six angioarrestins³²

*In the way of an exception to every rule, we identified a single full-sized FReD in the genome of the bacterium *Bacteriovax marinus* (recently renamed from being *Bdellovibrio stolpii*),³¹ the evolutionary history of which remains mysterious.

Figure 2. Phylogenetic tree (unrooted) constructed from 23 FReD sequences found in human genome. Designations are those used by the NCBI.

(also called angiopoietin-like), (e) three fibrinogen domains (β C, γ C, and the minor form called α 'C) (f) fibroleukin, (g) hepassocin, (h) the fibCD1 protein, and (i) the microfibril associated glycoprotein (MFAgp) (Fig. 2). The distribution of orthologs of these proteins was determined for eight other vertebrates (Table II). Initial assignments were made on the basis of how FReD sequences from those other organisms clustered with their human counterparts in phylogenetic trees. As an example, the results of the human-chicken comparison are shown in Figure 3. Although human and chicken both have 23 FRePs, they are not the same 23; gains and losses were sustained along both lineages. Comparisons with FReD collections from seven other vertebrates are available as Supporting Information (Supporting Information Figures S2–S12); the results for each of

the major types of FReP are summarized in the following sections.

Tenascins. Alternative splicing aside, four different tenascins are encoded in the human genome: tenascin-C, tenascin-W, tenascin-R, and tenascin-X. Genes for these same tenascins are found in almost all of the vertebrate genomes we examined. An additional tenascin-C was previously reported for two species of puffer fish.³³ We now find that zebrafish and mangofish also have two different tenascin-C genes, and additionally, mangofish has two tenascin-X genes, for a total of six tenascin FReD genes. Contrarily, the currently available draft genome sequences for the opossum and platypus genomes both lack a tenascin-X (Supporting Information Figures S2 and S3).

Table II. Distribution of Full-Length FRePs in Chordates

Frep ^a	Sea squirt	Lamprey ^b	Zebrafish	Mangofish	Clawed frog	Green lizard	Chicken	Platypus	Opossum	Human
MFAgp	0	0	15	9	4	1	0	0	1	1
Fibroleukin	1	2	2	3	2	1	1	1	1	1
FIBCD1	0	1	2	2	1	1	1	1	1	1
Ficolins	9	0	0	0	25	4	1	1	3	3
Tenascins	1	2	5	6	4	4	4	3	3	4
Angiopoietins	0	6	3	4	3	4	3	2	3	3
Angioarrestins	1	1	10	6	8	4	5	5	5	6
Fibrinogen	3	3	3	3	3	3	3	3	3	3
Hepassocin	0	1	2	2	2	2	2	1	1	1
Other or unassigned	18	4	0	3	4	3	3	1	0	0
Total FReDs	33	>20 ^b	42	38	56	27	23	18	21	23

^a FReP, fibrinogen-related protein.

^b Lamprey data based on draft assembly; some FReD fragments not included.

Figure 3. Phylogenetic tree of 23 FReDs from human and 23 from chicken. Those cases in which the chicken does not have an ortholog of the human type (e.g., MFAgp) are denoted with asterisks (*), and in those instances where the inverse is true are marked with arrows (←) (e.g., chicken has two hepassocins).

Ficolins. The human genome encodes three different ficolins, usually designated L-, M-, and H- (also called ficolin-1, -2, and -3). The three types differ with regard to the number of collagen triplets (GXX-) in the amino-terminal segment, leading to slightly different molecular weights.

The distribution of ficolins among non-mammalian vertebrates varies greatly. None has been found in any fish (Table II). In direct contrast, the African clawed frog, *Xenopus* (*Silurana*) *tropicalis*, has 25 verifiable ficolins. Four of these do not have collagen segments, but phylogenetic analysis shows they have lost those segments, as opposed to never having had them (Supporting Information Figure S5). The number of collagen-triplets in the remaining 21 ranges from 2 to 36.

The Green anole lizard, *Anolis carolinensis*, has four quite similar ficolins, the result of three recent gene duplications (Supporting Information Figure S4). Chicken (*Gallus gallus*) and platypus (*Ornithorhynchus anatinus*) each have a single ficolin encoded in their genomes, but the opossum (*Monodelphinus domestica*) has orthologs of all three human types (Table II).

Angiopoietins. There has been confusion about the number of angiopoietins in humans and other mammals, some literature citations, as well as the NCBI protein database, listing four angiopoietins (angiopoietin-1, angiopoietin-2, angiopoietin-3, and angiopoietin-4). However, in the NCBI listing, as

well as in structural alignments shown in some publications,²¹ angiopoietins-3 and -4 are identical (NCBI: AAD31728.1 and EAX10654.1). To confuse matters further, there is a set of angiopoietin-like proteins—the ANGPTL- series—that are sometimes called angioarrestins.³²

Authentic angiopoietins, by which we mean angiopoietins-1, -2, and -3 (or -4) are unique among FReDs in having a third disulfide bond, readily recognizable as a C-X-C-X-C motif in a critical region of the third subdomain. By this criterion, we count three authentic angiopoietins in humans (Fig. 2). However a fourth FReD, ANGPTL-7, clusters with these three phylogenetically and also has an extra disulfide bond, one of its cysteines corresponding to those in authentic angiopoietins.

Non-mammalian vertebrate genomes vary with regard to the number of authentic angiopoietin sequences they encode. Opossum, platypus, chicken, and frog have three, like humans, but the green anole lizard and mangofish have four, zebrafish has three (two of which were previously cloned³⁴), and lamprey has six, all of which contain the CXCXC motif with the extra disulfide bond (Table II).

Angiopoietin-like Proteins (Angioarrestins). In humans, three angiopoietin-like FReDs—ANGPTL-1, -2, and -5—form a distinct phylogenetic clade. Like genuine angiopoietins, these angioarrestins have amino-terminal, long two-stranded α -helical

coiled coils, but, unlike the real angiopoietins, the angioarrestins have a proline-rich segment between the coiled coils and the FReD.

Two other angiopoietin-like FReDs, ANGPTL-4 and -6, branch independently in the overall phylogeny (Fig. 2); both have shorter amino-terminal segments. A sixth, the ANGPTL-7, as mentioned above, clusters with the three genuine angiopoietins, but does not have the full CXCXC motif found in the other three. A seventh angiopoietin-related protein (ANGPTL-3) was not included in our trees because it lacks the third (carboxyl-terminal) subdomain.

The criteria for assigning FReDs from non-mammalian vertebrates to the angioarrestin group included clustering with assigned ANGPTL- FReDs from human, as well as the absence of the CXCXC motif in the third domain disulfide region. Additionally, in considering orthologs of the main group of three, we looked for the proline-rich segment that occurs between the coiled coil and the FReD. By these rules, the chicken genome has five of six possible orthologs of the human set, green lizard four, African clawed frog eight, mangofish six, and zebrafish ten. We were only able to make a single positive identification in the lamprey: an authentic angioarrestin that was 80% identical with the sequence of human ANGPTL-2 (Supporting Information Figure S12), did not have CXCXC, and which contained the proline-rich segment between the coiled coil and the FReD.

Fibroleukin. The number of fibroleukin sequences varies from one to three in vertebrate genomes. Single copies are found in human, marsupials, lizard, and chicken, two each in frog, zebra fish, and lamprey, and three in mangofish (Table II).

Intestinal chitin-binding proteins (FIBCD1). The FIBCD-1 FReP, which occurs in the brush border intestinal wall and is known to bind chitin,^{9,35} was found in every vertebrate examined, almost always with a characteristic membrane-spanning sequence in the amino-terminal region (the one exception corresponds to an unsequenced region in the lamprey genome). Zebrafish and mangofish have two genes for this protein.

Microfibril-associated glycoproteins. Phylogenetically, the MFAgPs are most closely related to the intestinal chitin-binding protein (Fig. 2). Although none was found in lamprey, there are 15 in zebrafish and 9 in mangofish, four in the frog, and one each in lizard, opossum, and human. None was found in chicken or in the draft genome of the platypus (Table II). It is thought that asparagine-linked carbohydrate in MFAgP plays a role in binding to integrins,⁸ and indeed the motif for such a site persists throughout the vertebrates. The intestinal chitin-binding protein also consistently exhibits

Figure 4. Different vertebrate FReD types change at different rates. The pairs chosen for time points are: rodent-human (70 mya); human-opossum (110 mya); and human-platypus (130 mya).

such a motif, however, as do many other of the FReDs examined.

Fibrinogen domains. All vertebrates appear to have a set of three orthologous genes corresponding to the β C, γ C, and α 'C FReDs found in human fibrinogen. In this regard, it is firmly established that the “holes” of β C and γ C FReDs interact with “knobs” on other fibrinogen molecules during fibrin formation, but, as far as is known, the carboxyl domain of a minor form of the α chain denoted α 'C does not. In lampreys the α 'C is part of a separately encoded α chain,³⁶ but in jawed vertebrates it is the result of alternative splicing.³⁷

Hepassocins. Most non-mammalian vertebrates have two hepassocins. Only one was found in the lamprey, marsupials, and mammals. Phylogenetically, hepassocin FReDs invariably cluster with those from fibrinogen (Figs. 2 and 3).

Vertebrate FReDs change at different rates

Different FReD types change at vastly different rates. The slowest changing FReDs are the angiopoietins and angioarrestins; the fastest are ficolins (Fig. 4). For example, human ANGPTL-2 and an angioarrestin from chicken have only two differences among their 221 residues, but the lone ficolin in chickens has 75 differences among 211 residues in its best match with a human ficolin.

Not only do ficolins change faster than other FReDs but they also experience more of that change in the third subdomain (“P domain”) than in the first and second subdomains, the very opposite of what occurs in all other FReDs except the microfibril-associated glycoproteins (MFAgPs) (Table III). The differential is most pronounced when orthologs are compared, but it extends to paralogs as well, a relevant observation when evaluating the large numbers of FReDs in non-chordate organisms like the sponge and sea anemone.

Table III. Different Rates of Change for FReD Subdomains as Reflected in Ratio of %IDs: (AB Domains) / (P Domain)

Chordates	# seqs	# comps	(AB)/P (frac IDs $\pm$ SD)
Angiopietins	9	36	0.87 $\pm$ 0.08
Angioarrestins	5	10	0.85 $\pm$ 0.12
FIBCD1	8	28	0.90 $\pm$ 0.05
Tenascins	10	45	0.90 $\pm$ 0.05
Fibroleukins	12	66	0.87 $\pm$ 0.07
Ficolins			
“vert, not frog	7	21	1.16 $\pm$ 0.28
“frog	25	276	1.42 $\pm$ 0.24
“sea squirt	6	15	1.08 $\pm$ 0.13
Mfagps	9	36	1.28 $\pm$ 0.20
Fibrin-gamma	6	15	0.90 $\pm$ 0.06
Fibrin-beta	6	15	0.75 $\pm$ 0.08
Fibrin-alpha'	6	15	0.82 $\pm$ 0.08
Hepassocin	6	15	0.83 $\pm$ 0.05
Non-Chordates			
Paralogs			
Sponges	14	91	1.57 $\pm$ 0.25
Sea anemone	19	171	1.33 $\pm$ 0.29
Snail	6	15	1.46 $\pm$ 0.15
Mosquito	6	15	1.17 $\pm$ 0.20
Orthologs (insects)			
Scabrous	3	3	0.81 $\pm$ 0.02
Unknown protein	3	3	0.92 $\pm$ 0.03

Protochordates

Two protochordate genomes were downloaded and surveyed: the cephalochordate amphioxus (*Bran-chiostoma floridae*) and the urochordate sea squirt (*Ciona intestinalis*). For the most part, however, our study concentrated on the sea squirt; our examination of the amphioxus database, which has an inordinate number of FReDs, was limited to establishing the presence or absence of certain key sequences, as will be noted in passing. Experimental evidence for FRePs with bacterolytic activity has been reported for the closely related species, *B. belcheri*.³⁸

As it happens, the genome of the common sea squirt on its own revealed a set of remarkable evolutionary innovations involving FReDs. These include self-incompatibility factors, proteins that interact with the signaling protein Notch, notochord developmental factors, a protofibrinogen, and a transcription factor, the last named being the only reported case of an intracellular FReD. None of these appears to have a counterpart in the amphioxus genome.

Tenascin. As noted in earlier reports,^{33,39} *C. intestinalis* has one very long FReP with a domain combination typical of tenascin. It has also been noted that amphioxus has a similar gene.³³

Ficolins. Previous workers have identified ficolins in other ascidians⁴⁰; we counted nine in sea squirt with various numbers of collagen GXX- triplets. Phylogenetic trees showed that none of these is ortho-

logs with ficolins found in vertebrates (Supporting Information Figure S14).

Angiopietins and angiopietin-like proteins.

Given the criteria used for vertebrates, we did not find any authentic angiopietins in the sea squirt (*C. intestinalis*) genome. We did find a gene for a protein that clusters with angioarrestins, however, the protein sequence (NCBI entry XP_002126422) of which is almost 60 percent identical with various members of the vertebrate ANGPTL subfamily.

Fibroleukin. The *C. intestinalis* genome has a FReD that clusters with the fibroleukins from vertebrates on phylogenetic trees. However, the ~400-residue long amino-terminal connector segment of this entry (NCBI NP_001027605) is about twice as long as a typical fibroleukin connector and bears no sequence resemblance to vertebrate fibroleukins, nor does it resemble any other known animal protein.

FIBCD1 and microfibril associated glycoprotein.

No genes for either of these FRePs were identified in the sea squirt genome.

Fibrinogen domains. A precursor form of vertebrate fibrinogen had previously been found in the sea squirt genome in which three consecutive genes give rise to a subunit protein containing obvious coiled-coil segments, followed by FReDs with the exact same disulfide bonding pattern that occurs in vertebrate fibrinogen β chains.²⁵ In contrast, no gene was found for hepassocin.

Self-incompatibility factors. In *C. intestinalis*, FReD-containing proteins are part of a self-incompatibility system that prevents self-fertilization by these hermaphroditic creatures.⁴¹ The FRePs, which are called v-Themis A and B, occur in the vitelline coat of eggs and are extremely polymorphic. The complementary membrane-bound receptors, which are not FRePs, occur in sperm and are called s-Themis A and B. Fertilization in these creatures is an external event, and the various combinations of sperm and egg A and B sort themselves out in the company of other individuals in a way that minimizes self-fertilization and inbreeding.⁴¹ The sea squirt self-incompatibility FReDs are very diverse (37–75% identical with each other), a reflection of the extremely rapid rates of change characteristic of proteins involved in gamete recognition.⁴²

Transcription factor. The solitary intracellular occurrence of a FReD, so far, is in a sea squirt transcription factor⁴³ and is worthy of detailed comment. NCBI lists two entries that are more than 99% identical and are apparently the same protein (FFA00181.1 is described as “TPA: transcription

Figure 5. The evolution of a transcription factor in the ascidian lineage (shown here as a region from the sea squirt (*Ciona intestinalis*) genome) resulted from the duplication of a ficolin gene downstream from a histone-2A gene. The duplication led to an unusual alternative splice in which a histone exon connects to a FReD in the duplicated gene, omitting the collagen segment (violet). In the NCBI versions, the exons X₁ and X₂ (light green) have the same DNA sequence but are read in different frames, a happenstance that would befuddle further splicing in the correct frame. The problem would be solved if a putative exon (yellow) in the histone region was used to put the reading frames in register.

factor”, and XP_002132084) as “transcription factor CBF/NF-Y/archaeal histone”). In either case, the sequence has the same amino-terminal histone-2A domain connected to the same FReD.

The NCBI also contains several other entries labeled “Predicted: Similar to transcription factor protein”. Upon phylogenetic analysis of their FReDs, these proteins—the “similars” and the TPA-transcription factor—cluster together tightly. We have found that the “similars,” all have amino-terminal neighboring collagen domains, however, which by any account makes them ficolins and not transcription factors. Indeed, the gene for one of the “similars” lies exactly next to the gene for the transcription factor, and careful inspection of the DNA sequences in this region shows how a coincidental splicing of alternate exons allowed the birth of this unique transcription factor (Fig. 5).

If the NCBI entries are correct, then, in a very unusual occurrence, the first exons of the FReDs in the adjacent genes are read in two different frames, a most unlikely occurrence that poses a dilemma for how one of these exons could be connected to the next exon, because one of the translations would have to be in conflict. More likely, there is yet another small exon (yellow in Supporting Information Fig. 5), properly phased, between the histone gene and the duplicated gene. In any case, the duplication leading to the transcription factor was a very recent event in ascidian evolution and is not likely to be in evidence in other lineages. Nothing similar was found in amphioxus.

Non-chordates

The strategy of using the human set of FReDs as a template for identifying orthologs in other species was less useful in non-chordates because of problems associated with greater evolutionary distances. Instead, trees were constructed with various sets of FReDs from different non-chordates in search of

orthologs. Additionally, sequences neighboring the FReDs were examined in an effort to follow evolutionary appearances.

Sponge. Partial sequences aside, 92 full-length FReDs were identified in the sponge, *Amphimedon queenslandica* (Table IV), the majority of which have short uncharacterized sequences on their amino-terminal sides (<60 residues between the signal sequence and the FReD). None has collagen-like sequences or any of the domains associated with FRePs found in chordates. However, 41 of the 106 had an ~50-residue motif containing five or six cysteines, consistent with a tightly folded, disulfide bonded domain. The motif was not found in any other proteins in the NCBI protein database.

Table IV. Distribution of Full-Length FReDs in Some Animal Genomes

Genome	Number of FReDs
Sponge (<i>Amphimedon queenslandica</i>)	92
Sea anemone (<i>Nematostella vectensis</i>)	25
Round worm (<i>Caenorhabditis elegans</i>)	8
Snail (<i>Biomphalaria glabrata</i>)	>50
Fruit fly (<i>Drosophila melanogaster</i>)	14
Mosquito (<i>Anopheles gambiae</i>)	29
Honey bee (<i>Apis mellifera</i>)	2
Silkworm (<i>Bombyx mori</i>)	3
Sea urchin (<i>Strongylocentrotus purpuratus</i>)	30
Amphioxus (<i>Branchiostoma floridae</i>) ^a	>100
Sea squirt (<i>Ciona intestinalis</i>)	33
Lamprey (<i>Petromyzon marinus</i>)	>20
Mangofish (<i>Oreochromis niloticus</i>)	38
Zebrafish (<i>Danio rerio</i>)	42
African clawed frog (<i>Xenopus tropicalis</i>)	56
Green anole lizard (<i>Anolis carolinensis</i>)	27
Chicken (<i>Gallus gallus</i>)	23
Platypus (<i>Ornithorhynchus anatinus</i>)	18
Opossum (<i>Monodelphis domestica</i>)	21
Human (<i>Homo sapiens</i>)	23

^a FReDs in the amphioxus genome were not rigorously inventoried.

Several sponge FRePs had relatively long amino-terminal sequences, the longest of which is a 789-residue protein precursor (NCBI XP_003387327) called “WD-repeat containing and planar cell polarity effector protein.” The entry also corresponds to the Pfam²⁸ designation “DUF3312” or (pfam11768), “a domain of unknown function” found in eukaryotes. Notably, the protein does not have the “WD” dipeptidyl sequence characteristic of WD-repeats. The nearest homolog to the sponge entry is a 558-residue protein in sea anemone that does not have a FReD.

In another case, an ~300-residue connector corresponds to a nidogen-like domain. Apart from those two, the three other sponge FRePs with relatively long connectors appear to be artifacts resulting from mispredicted splicing.

When included in phylogenetic trees with FReDs from other phyla, sponge entries invariably cluster separately, indicating an independent diaspora that has been occurring since the time of the last common ancestor of sponges and other animals. That said, and neighbor domains aside, sponge FReD sequences tend to occur closest to chordate ficolins and tenascins (Supporting Information Figure S17). A tree that compares a random assortment of 25 sponge FReDs with 25 sea anemone FReDs is shown in Supporting Information Figure S18.

Sea anemone. Again not counting partial sequences, the genome of the sea anemone (*Nematostella vectensis*) contains about 25 full-length FReDs, most assigned as “predicted proteins.” On average, the FReDs are about 45% identical to those found in sponge. Most of the sequences between the signal peptide and the FReD are quite short. Remarkably, the longest of the sea anemone FRePs (XP_001638156, 264 residues) has an amino-terminal collagen segment composed of 15 tripeptidyl GXX- repeats and has to be regarded as an independently evolved ficolin.

Snails. FReDs from the snail *Biomphalaria glabrata* have been extensively studied in the past,^{13,44–46} and our observations are mainly confirmatory. Neighboring domains have previously been found to be immunoglobulin domains and, in one case, a brace of EGF domains.⁴⁶ The lengths of these FRePs range from 372 to 756 residues. One of these (NCBI: AEO50747.1) has a connector domain (COG1196) that is also found in a *Drosophila* FReP (NCBI: NP_001104020) and is a likely ortholog.

Round worms. Because the round worm, *Caenorhabditis elegans*, is a traditional favorite of biologists, we surveyed its FReDs. Most of the amino-terminal connections in eight FRePs are

uncharacterized segments amounting to about 200 residues each.

Insects. Insect FRePs, particularly from mosquitoes and fruit flies, have been the subject of numerous studies in the past.^{47–50} Consistent with those earlier studies, we found a great variation in numbers of FReDs in various insects, ranging from 29 in the mosquito *Anopheles gambia* to only two in the honeybee, *Apis mellifera* (Table IV). It has been reported that in mosquitoes a FReP referred to as FBP9 (ADF80549) is a dimer.⁴⁹ The relatively short amino-terminal segment (about 60 residues) contains a single cysteine that is likely responsible for dimer formation.

In the case of the honeybee, both FRePs appear to be orthologs of proteins found in other insects, one being the well known scabrous protein originally found in *Drosophila melanogaster*.¹⁰ The scabrous protein in *Drosophila* is known to be a homodimer.⁵¹

Sea Urchins. The sea urchin (*Strongylocentrotus purpuratus*) is a basal deuterostome and a reasonable place to search for ancestors of chordate FRePs. Several of the FReDs tend to cluster with human ficolins and tenascins (Supporting Information Figure S15), but none of them had collagen segments or any hint of domains found in combination with vertebrate tenascins. In the end, we were unable to place any of the 35 FReDs as orthologs with chordate counterparts (Supporting Information Figure S16).

Two classes of FReD

Phylogenetic analysis suggested that there are two distinguishable groups of FReD, one of which occurs in all animal groups and the other of which is restricted to protochordates and chordates (Supporting Information Figure S17). The most significant differences between the two groups occur in the third subdomain (“P domain”) where ligand binding occurs. The later emerging group, which we call Class II, has longer sequences in this region, the result of several insertions. As a first approximation, assignment to one of the two classes can be determined simply by counting the number of residues between the two characteristic disulfide bonds that define FReDs, a simple reflection of an assortment of small insertions and/or deletions in this region (Fig. 6); the two classes also differ as a result of other insertions nearer their carboxyl-termini. Both kinds of insertion bear heavily on the nature of the binding site.

X-ray crystallographic structures are available for both classes of FReD, and although much has been written about how similar they are, the differences are what dictate their different functions. The class I structures include all three human ficolins, angiopoietin-2, and horseshoe crab tachylectin. The

Figure 6. Chordate FReDs (top) occur in two general classes (I and II) that can be distinguished simply by the number of intervening residues between the two essential disulfide bonds. Non-chordate FReDs (lower) are predominantly Class I, an anomalous cluster found in *C. elegans* (denoted by W) notwithstanding.

class II structures include the α /C, β C, and γ C domains of a variety of vertebrate fibrinogens.

Previous structural alignments^{21–24} have focused on similarities in ligand binding. For example, the small molecule ligands for mammalian ficolins and horseshoe crab tachylectin—all class I FReDs—are virtually superposable in such alignments.^{22,23} Class II ligands have also been shown to superpose at the same locations,²⁴ although, as discussed below, some distinguishing features for these comparisons have not been remarked upon previously.

In this regard, all FReDs appear to indulge in macromolecular interactions, the complementary non-FReD interactant typically, but not always, protruding from cell surfaces. Many FReDs are designated as lectins because the interactant is a carbohydrate. Experimentally, FReDs have been shown to bind to sialic acids, acetylated sugars, chitin, and D- and L-sugars of bacterial origin (Table V).^{9,12,14,22,52–56}

All known carbohydrate-binding FReDs fall into class I.

Assignment of FReDs to one of the two classes was a first step in correlating structural modifications with changes in function. However from the viewpoint of function, FReDs can be categorized into three groups depending on whether (a) they interact directly with host cells (e.g., angiopoietins), (b) bind to nonhost cells (e.g., ficolin), or (c) bind peptide knobs at the ends of extended tethers (e.g., fibrinogen/fibrin). It might be expected, a priori, that the direct face-to-face nature of host cell signaling would be the most demanding with regard to conserving structure, and certainly angiopoietins and angioarrestin FReDs change the most slowly (Fig. 4).

Sugar-binding FReDs

Although FReDs from all three human ficolins have binding sites at locations corresponding to the polymerization “holes” that occur in vertebrate

Table V. FRePs Known to Bind Sugars or Sugar Derivatives

FReP	Ligand	Organism	Refs. 31–38
Ficolins	GlcNac et al.	<i>Homo sapiens</i>	Ref. 55
Tachylectin	GlcNac	<i>Tachyleus tridentatus</i>	Ref. 12
FIBCD1	Chitin	<i>Homo sapiens</i>	Ref. 9
Slug lectin	Sialic acid	<i>Limax flavus</i>	Ref. 53
Snail lectin	L-Fucose	<i>Biomphalaria glabrata</i>	Ref. 13
Ascidian lectin (ficolin)	GlcNac	<i>Halocynthia roretzi</i>	Ref. 40
Ascidian lectin (P36)	glucose	<i>Halocynthia roretzi</i>	Ref. 54

fibrinogens,^{56,57} the principal ligands known to bind to class I FReDs are carbohydrate in nature. Both M- and L-ficolins have been shown to bind *N*-acetyl glucoseamine (GlcNac) in the same way as has been shown for horseshoe crab tachylectin.²⁴ L-Ficolin has also been shown to bind other acetylated sugars (GalNac, ManNac) as well as *N*-acetyl cysteine. H-ficolin, on the other hand, binds galactose, as well as D- and L-fucose, molecules known to occur among the complex carbohydrates on the surfaces of bacteria.^{22,23}

Other class I FReDs, for which X-ray structures are not available, have been shown to bind a similar set of N-acetylated sugars and other acetylated small molecules (Table V). Sequence comparisons of the chitin-binding protein called FIBCD1 have been made with horseshoe crab tachylectin and ficolins,^{9,35} many of the same residues being found to be essential for binding.

The alignment of essential residues for binding sugars or their derivatives holds up when ficolins of other vertebrates are used, however, as well as

when FReDs from non-chordates are examined (Fig. 7). Numerous other class I FReDs meet the same conditions also, including MFagg, the FReD of which is closely related to that of FIBCD1. Indeed, many sponge and sea anemone FReDs meet the same criteria, as do several sea squirt FReDs (Fig. 7), and it can be anticipated that many of these share the ability to bind N-acetylated moieties.

The question arises, which of the class I FReDs do not have the key residues regarded as essential for acetyl binding? Tenascins, some of which are reported to bind to integrins by way of the RGD motif,⁵⁸ for example, lack a key region in the FReD as the result of a deletion (Fig. 7). In the past, RGDs in tenascins have been mainly associated with their FN3 domains, and only two of the 37 tenascins we examined have an RGD sequence in the FReD (one of the five in zebrafish and one of the three in platypus). In only one instance could we find a report of a possible ligand binding to the FReD portion of a tenascin,⁵⁹ and apparently, a structural characterization of that interaction has not been pursued. In

Figure 7. Alignment of key interfacial region (third subdomain) from a series of FReDs. Residues known to be involved in ligand contacts on the basis of X-ray crystallography (sequences marked with * have X-ray structures) are shaded red; otherwise conserved residues are green. Note that matches with all the shaded residues from human ficolins and the horseshoe crab tachylectin occur in several sponge and sea anemone FReDs, as well as with FIBCD1 and MFagg. Tenascin FReDs have suffered a deletion in a region thought to be essential for binding acetylated sugars.

Figure 8. Class I and II FReDs are structurally distinct at their principal interaction sites. The upper panel is a stereo-depiction of the angiopoietin-2/Tie-2 receptor complex (PDB 2GY7) with two other class I FReDs (human ficolin 1, PDB 2J5Z, blue, and horseshoe crab tachylectin, PDB 1JC9, red) structurally aligned on the angiopoietin (green). In the lower panel two class II FReDs (the β C (brown) and γ C (magenta) domains from human fibrinogen, PDB 1FZC) structurally aligned with angiopoietin in the same way. Unlike the ficolin and tachylectin, the class II FReDs clearly clash with the appropriately positioned angiopoietin-2/Tie-2 receptor (cyan).

our view, the auspicious location of FReDs in tenascins at the very ends of long gangly tethers begs for involvement.

Cell surface receptors and FReDs

The crystal structure of the angiopoietin-2-receptor complex is an example of a FReP acting as a diffusible signal with a host cell receptor protein.⁶⁰ The ligand-receptor interaction itself is mostly spread across a large complementary hydrophobic surface. Interestingly, structural alignment of other class I structures—horseshoe crab tachylectin and human ficolins—with the angiopoietin-2 model results in surprisingly smooth fitting with the angiopoietin receptor, virtually no clashing observed at the backbone level (Fig. 8, upper panel).

Angiopoietins bind to their receptors without the involvement of carbohydrate; however, a variety of polar and nonpolar interactions occurring between amino acid sidechains across the interface. Even so, the results suggest that during evolution there is no fundamental barrier between pattern recognition class I structures and those involved in host cell-activation.

FReDs that bind amino-terminal peptides

In contrast, in the only two instances where FReDs are known to bind peptide ligands, the FReDs are class II (the β C and γ C domains of vertebrate fibrinogens), and superposition of class II structures on the angiopoietin-2 model results in severe clashing with the receptor backbone (Fig. 8, lower panel). The interaction of “knobs and holes” that occurs in vertebrate fibrin formation is very different from the face-to-face interaction observed in the angiopoietin-angiopoietin receptor complex. Besides the inserted loops, the β C and γ C domains of vertebrate fibrinogen have experienced a series of substitutions at key sites whereby hydrophobic residues are replaced by polar ones, the sidechains of which have been shown to be critical for specific peptide binding.^{18,19}

Discussion

FReDs are found in all animals, from sponges to mammals; they do not occur in choanoflagellates, the proposed direct descendant of the last unicellular ancestor of animals.⁶¹ The sponge genome (*A. queenslandica*)⁶² contains almost a hundred full-length FReDs; most of which were spawned by gene duplications that occurred well after the divergence of modern sponges from the common ancestor with other animal phyla. Phylogenetically speaking, there have been drastic expansions and contractions of FReD populations along all lineages. Among insect genomes, for example, the FReD count may be as high as 30 in some species, but silkworms have only three occurrences and honeybees only two.

Although FReDs from non-deuterostomes are frequently called “angiopoietin-like” or “ficolin-like” or “tenascin-like”, etc., these are not genuine ficolins (no collagen domains), nor are they authentic tenascins (no Fn3 domains), nor any of the other types of FReD found in deuterostomes. In the sea anemone genome (*N. vectensis*), a FReD was found adjacent to a collagen segment, but phylogenetic considerations indicate that it is the result of independent evolution and not a strict ortholog of vertebrate ficolin.

Nonetheless, the conventions of nomenclature are such that many FRePs from non-deuterostomes have been named “ficolin-like” or “tenascin-like,” and so forth, based on the kind of sequence they tend most to resemble. Our findings suggest that it would be folly to infer any commonality of function from these names, however.

FReDs and innate immunity

Many FReDs have been implicated in the phenomenon of innate immunity, and especially in the recognition of pathogens.^{12,13,46–48,57,63,64} Chief among these are the ficolins. Although genes for ficolins have been found among the protochordates, notably they have not been found in fish. The frog (*Xenopus tropicalis*) has more than two dozen ficolin genes; lizards have four and birds only one.

On the other hand, fish have multiple genes for microfibril-associating glycoproteins (Table II). Given that MFAgs exhibit the same unusual kind of accelerated change in their third domains, it may be that in fish these proteins are playing the same role as ficolins, even though without a collagen portion they would not participate in the complement pathway, a characteristic of vertebrate ficolins.⁶⁵

Ficolins have evolved independently on at least three different occasions. One was found in the sea anemone genome but not in the sponge or any of the protostomes we examined (round worm, snail, and four arthropods). They are found in protochordates (e.g., sea squirts) and all tetrapods, but not fish.

Among non-chordates, FReD-containing proteins have been shown to defend against infection. In snails, for example, studies have shown that these proteins help resist infection by trematode parasites.¹³ In horseshoe crabs, FReD-containing proteins are known to have strong bacterial agglutinating activity.⁶³

As a general rule, the interactive third subdomain (“P domain”) changes more slowly than the rest of the FReD. In contrast, in ficolins, often cited as being part of innate immunity, the third subdomain changes significantly faster than the other two subdomains. Positive selection associated with this change is implied by the drastic differences in rates of change.

When on occasion a FReD may assume a new, more specific function, the accelerated change in the third subdomain comes to a halt, the new attachment interaction geometry being locked in and conserved. As an example, arthropod FReDs orthologs to the scabrous gene product of *Drosophila* exhibit a slower change for the third subdomain than for the rest of the FReD (Table III). The scabrous gene encodes a diffusible protein thought to interact, directly or indirectly, with photoreceptor cells during development.⁵¹

In summary, FReDs are widely distributed in the animal kingdom, almost entirely as extracellular, diffusible proteins, the interactive sites of which typically involve the carboxyl-terminal subdomain. The broad diversity of functions exhibited by these structures range from cell signaling to the binding of pathogens to fibrin gel formation. A potential for binding alien pathogens is reflected by a very rapid change in the carboxyl-subdomain that is manifested in vertebrate ficolins (and fish MFAgs), as well as numerous non-chordate FReDs, and is likely charac-

teristic of positive selection. On another note, structural changes that gave rise to the ability to bind tethered peptide knobs leading to polymerization arose after the split leading to chordates, the first evidence of their presence occurring in ascidians (sea squirts).

Materials and Methods

Databases

Database holdings available at the National Center for Biotechnology Information (NCBI.gov) were used extensively. Whole sequence genomes for sponge (*A. queenslandica*), sea anemone (*N. vectensis*), honeybee (*A. mellifera*), sea urchin (*S. purpuratus*), amphioxus (*B. floridae*), sea squirt (*C. intestinalis*), mangofish (*Oreochromis niloticus*), and green anole lizard (*Anolis carolina*) were downloaded on to our own computer. These databases were used for resolving anomalies associated with automated splicing errors, as well as for determining the origin of rearrangements leading to the unique transcription factor in *C. intestinalis*.

BLAST⁶⁶ in one form or other was used for all searches. Alignments and trees were calculated by the progressive method⁶⁷ and trees drawn at the Phylodendron website (<http://iubio.indiana.edu/treeapp/treeprint-form.html>). PDB files were downloaded from the Protein Data Bank (www.rcsb.org). Structures were manipulated with PyMol.⁶⁸ All structural alignments were made by specifying the 14-residue segment encompassing the unique disulfide bond in the FReDs being compared.

Database URLs

sponge:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Amphimedon_queenslandica

sea anemone:

<http://genome.jgi.doe.gov/Nemve1>

honeybee:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Apis_mellifera

sea urchin:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Strongylocentrotus_purpuratus

amphioxus:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Branchiostoma_floridae

sea squirt:

<http://www.jgi.doe.gov/ciona>

lamprey:

http://genome.wustl.edu/pub/Petromyzon_marinus

mangofish:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Oreochromis_niloticus

lizard:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Anolis_carolinensis

References

- Doolittle RF (1992) A detailed consideration of a principal domain of vertebrate fibrinogen and its relatives. *Protein Sci* 1:1563–1577.
- Jones FS, Hoffman S, Cunningham BS, Edelman GM (1989). A detailed structural model of cytotoxin: protein homologies, alternative RNA splicing, and binding regions. *Proc Natl Acad Sci USA* 86:1905–1909.
- Erickson H (1993) Tenascin-C, tenascin-R and tenascin-X: a family of talented proteins in search of functions. *Curr Opin Cell Biol* 5:869–876.
- Ichijo H, Hellman U, Wernstedt C, Gonez LJ, Claesson-Welsh L, Heldin C-H, Miyazono K (1993). Molecular cloning and characterization of ficolin, a multimeric protein with fibrinogen- and collagen-like domains. *J Biol Chem* 268:14505–14513.
- Davis S, Aldrich TH, Jones PF, Acheson A, Compton DL, Jain V, Ryan TE, Bruno J, Radziejewski C, Maisonpierre PC, Yancopoulos GD (1996) Isolation of angiopoietin-1, a ligand for the Tie2 receptor, by secretion-trap expression cloning. *Cell* 87:1161–1169.
- Koyama T, Hall LR, Haser WG, Tonegawa S, Saito H (1987) Structure of a cytotoxic T-lymphocyte-specific gene shows a strong homology to fibrinogen. *Proc Natl Acad Sci USA* 84:1609–1613.
- Yamamoto T, Gotoh M, Sasaki H, Terada M, Kitajima M, Hirohashi S (1993) Molecular cloning and initial characterization of a novel fibrinogen-related gene, HFREP-1. *Biochem Biophys Res Commun* 193: 681–687.
- Kobayashi R, Mizutani A, Hidaka H (1994) Isolation and characterization of a 36-kDa microfibril-associated glycoprotein by the newly synthesized isoquinolinesulfonamide affinity chromatography. *Biochem Biophys Res Commun* 198:1262–1266.
- Schlosser A, Thomsen T, Moeller JB, Nielsen O, Tornøe I, Mollenhauer J, Moestrup SK, Holmskov U (2009) Characterization of FIBCD1 as an acetyl group-binding receptor that binds chitin. *J Immunol* 183:3800–3809.
- Baker NE, Mlodzik N, Rubin GM (1990) Spacing differentiation in the developing *Drosophila* eye: a fibrinogen-related lateral inhibitor encoded by scabrous. *Science* 250:1370–1377.
- Xu, X, Doolittle RF (1990) Presence of a vertebrate fibrinogen-like sequence in an echinoderm. *Proc Natl Acad Sci USA* 87:2097–2101.
- Gokudan S, Muta T, Tsuda R, Koori K, Kawahara T, Seki N, Mizunoe Y, Wai S, Iwanaga S, Kawabata S-I (1999) Horseshoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. *Proc Natl Acad Sci USA* 96:10086–10091.
- Adema CM, Hertel LA, Miller RD, Loker ES (1997) A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci USA* 94: 8691–8696.
- Kurachi S, Song Z, Takagaki M, Yang Q, Winter HC, Kurachi K, Goldstein IJ (1998) Sialic acid-binding lectin from the slug *Limax flavus*. Cloning, expression of the polypeptide, and tissue localization. *Eur J Biochem* 254:217–222.
- Chiquet-Ehrismann R, Tucker RP (in press) Tenascins and the importance of adhesion modulation. *Cold Spring Harb Perspect Biol*; doi:10.1101.
- Fagiani E, Christofori G (in press) Angiopoietins in angiogenesis. *Cancer Lett*; doi: 10.1016.
- Kilpatrick DC, Chalmers JD (in press) Human L-ficolin (ficolin-2) and its clinical significance. *J Biomed Biotechnol*.
- Yee VC, Pratt KP, Cote HCF, LeTrong I, Chung D, Davie EW, Stenkamp RE, Teller DC (1997) Crystal structure of a 30 kDa fragment from the gamma chain of human fibrinogen. *Structure* 5:125–138.
- Spraggon G, Everse SJ, Doolittle RF (1997) Crystal structures of fragment D from human fibrinogen and its cross-linked counterpart from fibrin. *Nature* 389: 455–462.
- Spraggon G, Applegate D, Everse SJ, Zhang JZ, Veerapandian L, Redman C, Doolittle RF, Grieneringer G. (1998) Crystal structure of a recombinant alphaEC domain from human fibrinogen-420. *Proc Natl Acad Sci USA* 95:9099–9104.
- Barton WA, Tzvetkova D, Nikolov DB (2005) Structure of the angiopoietin-2 receptor binding domain and identification of surfaces involved in Tie2 recognition. *Structure* 13:825–832.
- Garlatti V, Belloy N, Martin L, Lacroix M, Matsushita M, Endo Y, Fujita T, Fontecilla-Camps JC, Arlaud GJ, Thielens NM, Gaboriaud C (2007) Structural insights into the innate immune recognition specificities of L- and H- ficolin. *EMBO J* 26:623–633.
- Tanio M, Kondo S, Sugio S, Kohno T (2007) Trivalent recognition unit of innate immunity system: crystal structure of trimeric human M-ficolin fibrinogen-like domain. *J Biol Chem* 282:3889–3895.
- Kaires N, Beisel H-G, Fuentes-Prior P, Tsuda R, Muta T, Iwanaga S, Bode W, Huber R, Kawabata S (2001) The 2.0-Å crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and blood coagulation systems. *Proc Natl Acad Sci USA* 98:13519–13524.
- Doolittle RF (2012) The protochordate *Ciona intestinalis* has a protein like full-length vertebrate fibrinogen. *J Inn Immunity* 4:219–222.
- Nagy A, Hegyi H, Farkas K, Tordai H, Banyai L, Patthy L (2008) Identification and correction of abnormal, incomplete and mispredicted proteins in public databases. *BMC Bioinform* 9:353.
- Letunic L, Doerks T, Bork P (2002) SMART 6: recent updates and new developments. *Nucleic Acids Res* 30: 242–244.
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EC (2000) The Pfam protein families database. *Nucleic Acids Res* 28:263–266.
- Wilson D, Madera M, Vogel C, Chothia C, Gough J (2007) The superfamily database in 2007: families and functions. *Nucleic Acids Res* 35:D308–D313.
- Pradel G, Hayton K, Aravind L, Iyer LM, Abrahamsen MS, Bonawitz A, Meija C, Templeton TJ (2004) A multidomain adhesion protein family expressed in *Plasmodium falciparum* is essential for transmission to the mosquito. *J Exp Med* 199:1533–1544.
- Baer ML, Ravel J, Piniero SA, Guether-Borg D, Williams HN (2004) Reclassification of salt-water *Bdellovibrio* sp. As *Bacteriovorax marinus* sp. Nov. and *Bacteriovorax litoralis* sp. Nov. *Int J Syst Evol Microbiol* 54:1011–1016.
- Dhanabal M, Jeffers M, LaRochelle WJ, Lichenstein HS (2005) Angioarrestin: a unique angiopoietin-related protein with anti-angiogenic properties. *Biochem Biophys Res Commun* 333:308–315.
- Tucker RP, Drabikowski K, Hess JF, Ferralli J, Chiquet-Ehrismann R, Adams JC (2006) Phylogenetic analysis of tenascin gene family: evidence of origin in the chordate lineage. *BMC Evol Biol* 6:1–17.
- Pham VN, Roman BL, Weinstein BM (2001) Isolation and expression analysis of three zebrafish angiopoietin genes. *Dev Dynam* 221:407–474.

35. Thomsen T, Moeller JB, Schlosser A, Sorenson GL, Moestrup SK, Palaniyar N, Wallis R, Mollenhauer J, Holmskov U (2010) The recognition unit of FIBCD1 organizes into a noncovalently linked tetrameric structure and uses a hydrophobic funnel (S1) for acetyl group recognition. *J Biol Chem* 285:1229–1238.
36. Pan Y, Doolittle RF (1992) cDNA sequences of a second fibrinogen α chain in lamprey: an archetypal version alignable with full-length β and γ chains. *Proc Natl Acad Sci USA* 89:2066–2070.
37. Fu Y, Weissbach L, Plant PW, Oddoux C, Cao Y, Liang J, Roy SN, Redman CM, Greininger G (1992) Carboxy-terminal-extended variant of the human fibrinogen α subunit: a novel exon conferring marked homology to β and γ subunits. *Biochemistry* 31:11968–11972.
38. Fan C, Zhang S, Li L, Chou Y (2008) Fibrinogen-related protein from amphioxus *Branchiostoma belcheri* is a multivalent pattern recognition receptor with a bacteriolytic activity. *Mol Immunol* 45:3338–3346.
39. Tucker RP, Chiquet-Ehrismann R (2009) Evidence for the evolution of tenascin and fibronectin early in the chordate lineage. *Int J Biochem Cell Biol* 41:424–434.
40. Kenjo A, Takahashi M, Matsushita M, Endo Y, Nakata M, Mizuochi T, Fujita T (2001) Cloning and characterization of novel ficolins from the solitary ascidian *Halocynthia roretzi*. *J Biol Chem* 276:19959–19965.
41. Harada Y, Takagaki Y, Sunagawa M, Saito T, Yamada Y, Taniguchi H, Shoguchi E, Sawada H (2008) Mechanism of self-sterility in a hermaphroditic chordate. *Science* 320:548–550.
42. Vacquier VD, Swanson WJ (2011) Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harb Perspect Biol* 3:a002931.
43. Miwata K, Chiba T, Horii R, Yamada L, Kubo A, Miyamura D, Satoh N, Satou Y (2006) Systematic analysis of embryonic expression profiles of zinc finger genes in *Ciona intestinalis*. *Dev Biol* 292:546–554.
44. Zhang SM, Adema CM, Kepler TB, Loker ES (2004) Diversification of Ig superfamily genes in an invertebrate. *Science* 305:251–254.
45. Leonard PM, Adema CM, Zhang S-M, Loker ES (2001) Structure of two FREP genes that combine IgSF and fibrinogen domains, with comments on diversity of the FREP gene family in the snail *Biomphalaria glabrata*. *Gene* 269:155–165.
46. Zhang S-M, Nian H, Zeng Y, de Jong RJ (2008) Fibrinogen-bearing protein genes in the snail *Biomphalaria glabrata*: characterization of two novel genes and expression studies during ontogenesis and trematode infection. *Dev Comp Immunol* 32:1119–1130.
47. Cirimotich CM, Dong Y, Garver LS, Sim S, Dimopoulos G (2010) Mosquito immune defenses against Plasmodium infection. *Dev Comp Immunol* 34:387–395.
48. Dong Y, Dimopoulos G (2009) Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *J Biol Chem* 284:9835–9844.
49. Wang X, Zhao Q, Christensen BM (2005) Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, *Anopheles gambiae*, and the fruitfly, *Drosophila melanogaster*, genomes. *BMC Genomics* 6:114.
50. Middha S, Wang X (2008) Evolution and potential function of fibrinogen-like domains across twelve Drosophila species. *BMC Genomics* 9:260.
51. Lee E-C, Hu X, Yu S-Y, Baker NE (1996) The scabrous gene encodes a secreted glycoprotein dimer and regulates proneural development in Drosophila eyes. *Mol Cell Biol* 16:1179–1188.
52. Perovic-Ottstadt S, Adell T, Proksch P, Wiens M, Korzhev M, Gamulin V, Muller IM, Muller WEG (2001) A (1 $\rightarrow$ 3)- α -D-glucan recognition protein from the sponge *Suberites domuncula*. *Eur J Biochem* 271:1924–1937.
53. Knibbs RN, Osborne SE, Glick GD, Goldstein IJ (1993) Binding determinants of the sialic acid-specific lectin from the slug *Limax flavus*. *J Biol Chem* 268:18524–18531.
54. Sekine H, Kanjo A, Azumi K, Chi G, Takahashi M, Kasukawa R, Ichikawa N, Nakata M, Mizuochi T, Matsushita M, Endo Y, Fujita T (2001) An ancient lectin-dependent complement system in an ascidian: novel lectin isolated from the plasma of the solitary ascidian, *Halocynthia roretzi*. *J Immunol* 167:4504–4510.
55. Krarup A, Thiel S, Hansen A, Fujita T, Jensenius JC (2004) L-ficolin is a pattern recognition molecule specific for acetyl groups. *J Biol Chem* 279:47513–47519.
56. Gout E, Garlatti V, Smith DF, Lacroix M, Dumestre-Perard C, Lunardi T, Martin L, Cesbron J-Y, Arlaud GJ, Gaboriaud C, Thielens NM (2010) Carbohydrate recognition properties of human ficolins. Glycan array screening reveals the sialic acid binding specificity. *J Biol Chem* 285:6612–6622.
57. Garlatti V, Martin L, Lacroix M, Gout E, Arlaud GJ, Thielens NM, Gaboriaud C (2010) Structural insights into the recognition properties of human ficolins. *J Inn Immunity* 2:17–23.
58. Schnapp LM, Hatch N, Ramos DM, Klimanskaya IV, Sheppard D, Pytela R (1995) The human integrin $\alpha_8\beta_1$ functions as a receptor for tenascin, fibronectin and vitronectin. *J Biol Chem* 270:23196–23202.
59. Schumacher S, Jung M, Norenberg U, Dorner A, Chiquet-Ehrismann R, Stuermer CA, Rathjen FG (2001) CALEB binds via its acidic stretch to the fibrinogen-like domain of tenascin-C or tenascin-R and its expression is dynamically regulated after optic nerve lesion. *J Biol Chem* 276:7337–7345.
60. Barton WA, Tzvetkova D, Miranda EP, Kolev MV, Rajashankar KR, Himanen JP, Nikolov DB (2006) Crystal structures of the Tie2 receptor ectodomain and the angiopoietin-2-Tie2 complex. *Nature Struct Mol Biol* 13:524–532.
61. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
62. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS (2010) The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466:720–726.
63. Kawabata S, Iwanaga S (1999) Role of lectins in the innate immunity of horseshoe crab. *Dev Comp Immunol* 23:391–400.

64. Hanington P, Zhang S-M (2011) The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. *J Inn Immunity* 3:17–27.
65. Matsushita M, Endo Y, Hamasaki N, Fujita T (2001) Activation of the lectin complement pathway by ficolins. *Intern Immunopharm* 1:359–363.
66. Altschul SF, Madden TL, Schaffer AA, Zhang J, Miller W, Zhang Z, Miller DJ (1997) Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
67. Feng D-F, Doolittle RF (1996) Progressive alignment of amino acid sequences and construction of phylogenetic trees from them. *Math Enzymol* 266:368–382.
68. DeLano WL (2009) The PyMOL molecular graphics system. Palo Alto, CA: Delano Scientific LLC.