Cold Spring Harbor Symposia
on Quantitative Biology

Step-by-Step Evolution of Vertebrate Blood Coagulation

R.F. Doolittle

References	This article cites 21 articles, 6 of which can be accessed free at: **http://symposium.cshlp.org/content/early/2009/08/06/sqb.2009.74.001.refs.html**
P<P	Published online August 10, 2009 in advance of the print volume.
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Advance online articles have not yet appeared in the print volume. Citations to Advance online articles must include the digital object identifier (DOI) and date of initial publication.

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
http://symposium.cshlp.org/subscriptions

Step-by-Step Evolution of Vertebrate Blood Coagulation

R.F. DOOLITTLE

Department of Chemistry and Biochemistry and Molecular Biology, University of California,
San Diego, La Jolla, California 92093–0314

Correspondence: rdoolittle@ucsd.edu

The availability of whole-genome sequences for a variety of vertebrates is making it possible to reconstruct the step-by-step evolution of complex phenomena such as blood coagulation, an event that in mammals involves the interplay of more than two dozen genetically encoded factors. Gene inventories for different organisms are revealing when during vertebrate evolution certain factors first made their appearance in or, on occasion, disappeared from some lineages. The whole-genome sequence databases of two protochordates and seven nonmammalian vertebrates were examined in search of ~20 genes known to be associated with blood clotting in mammals. No genuine orthologs were found in the protochordate genomes (sea squirt and amphioxus). As for vertebrates, although the jawless fish have genes for generating the thrombin-catalyzed conversion of fibrinogen to fibrin, they lack several clotting factors, including two thought to be essential for the activation of thrombin in mammals. Fish in general lack genes for the "contact factor" proteases, the predecessor forms of which make their first appearance in tetrapods. The full complement of factors known to be operating in humans does not occur until pouched marsupials (opossum), at least one key factor still absent in egg-laying mammals such as platypus.

Blood coagulation in humans is a delicately balanced process involving more than two dozen extracellular proteins, many of which need to be converted from precursor forms during the process (Fig. 1). Almost half of the components are members of the serine protease family. Briefly put, the process acts as a biochemical amplifier in the forward direction, a small number of newly exposed tissue molecules acting as an input stimulus for an avalanche response converting a large number of fibrinogen molecules into a gelatinous clot. A delicate balance exists between the need for a fluid circulating state and the polymeric gel at a wound site; of necessity, the process is highly regulated, with various protease inhibitors and counteracting proteases involved. The question arises of how and when this complex process evolved.

Many of the proteins involved are clearly related to one another by gene duplications, and in the past, sequence-based phylogenies have offered insights into the relative order in which certain factors appeared (see, e.g., Doolittle and Feng 1987; Doolittle 1993; Hughes 2000). The relationship of paralogs can also be inferred by the kinds and arrangements of subsidiary domains associated with the catalytic domains (Fig. 2).

The subsidiary domains have very important roles in the blood-clotting process, serving as protein–protein interaction sites and localizing the clot at the site of injury. With regard to localization, GLA (γ-carboxyglutamic acid) domains and discoidin domains anchor certain of the factors to platelets at wound sites. Other subsidiary domains, including kringles, epidermal growth factor (EGF), and plasminogen-apple-nematode (PAN) domains, are well known to promote protein–protein interactions, leading in this case to the assemblage of factors needed for clot formation.

Which of the factors appeared first? As far as known, the thrombin-catalyzed conversion of fibrinogen to fibrin is unique to vertebrate animals (Doolittle 1961; Doolittle and Surgenor 1962), a group for which an accurately determined fossil record is available. As such, the times of appearance or disappearance of various genes can be gauged by examining genomes from various classes of extant vertebrates.

Although blood clotting is mostly an extracellular event centering around the thrombin-catalyzed conversion of fibrinogen into fibrin, blood platelets are also intimately evolved, being both thrombin sensitive, on the one hand, and having a strong affinity for fibrinogen and fibrin, on the other hand. Platelets also sequester some other coagulation factors as well. It must be kept in mind that in nonmammalian vertebrates, the equivalents of platelets are white cells commonly called thrombocytes.

SEARCHING WHOLE-GENOME SEQUENCE DATABASES

The whole-genome sequence (WGS) databases of sea squirt, amphioxus (lancelet), lamprey, puffer fish, frog, green anole lizard, chicken, platypus, and opossum were examined in search of ~20 genes known to be associated with blood clotting in mammals (Table 1).

The general strategy for identifying putative coagulation genes in WGS databases for nonmammalian organisms began with BLAST searching (Altschul et al. 1997) of the protein sequences of a queried factor as it occurs in mammals. The strongest "hits" were then "back-searched" against the standard NCBI nonrecombinant (nr) database. If the latter search returns the target protein as the highest scorer, the hit is presumed to be positive. If other proteins
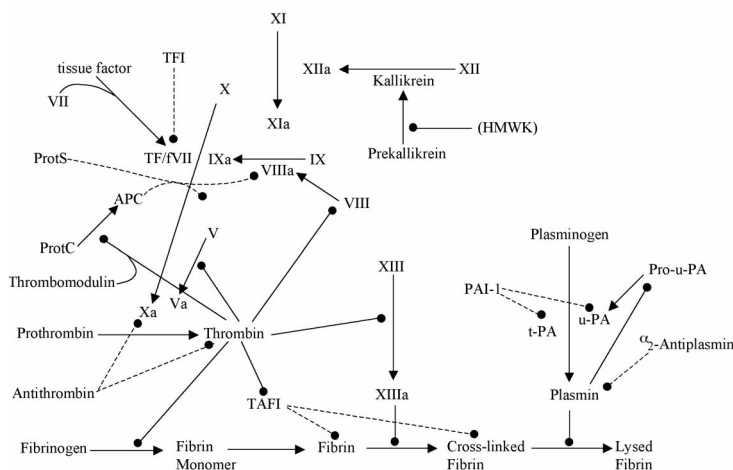
Figure 1. Schematic representation of blood coagulation pathway as it occurs in humans. (Dashed lines) Inhibitory actions. (TFI) Tissue factor inhibitor, (APC) activated protein C, (HMWK) high-molecular-weight kininogen, (TAFI) thrombin-activated fibrinolysis inhibitor, (t-PA) tissue plasminogen activator, (u-PA) urinary plasminogen activator.

have higher scores, it is generally thought that the hit is a paralog and not the targeted query.

The strategy is not foolproof. Changes in the rates of evolution after gene duplications can confound the process, and caution is the byword. The further back in time one tries to probe, the more challenging the process. Obviously, it is more difficult to prove the absence of a gene than its presence. Not all WGS databases are wholly complete or fully assembled. Genes can be missed.

The major difficulty in identifying genes in genome databases, however, is distinguishing orthologs from recently diverged paralogs. Several of the gene duplications that gave rise to new clotting factors took place very soon after the process got started in the narrow window between the appearance of protochordates and jawless vertebrates.

BASAL CHORDATES

Basal chordates, or protochordates, consist of a few groups of organisms that have a notocord at some stage in their development but do not have backbones. Two of these animals, the ascidian called the sea squirt and the cephalochordate called amphioxus or lancelet, have had their genomes sequenced. Until recently, the ascidians were regarded as the earlier diverging species, compared with cephalochordates, but on the basis of their genomes, it is now argued that cephalochordates like amphioxus are the more primitive (Putnam et al. 2008).

Creatures in both groups have simple circulatory systems and primitive hearts, and it has long been thought that the clumping of circulating cells at wound sites is the only hemostatic mechanism needed, there being no evi-
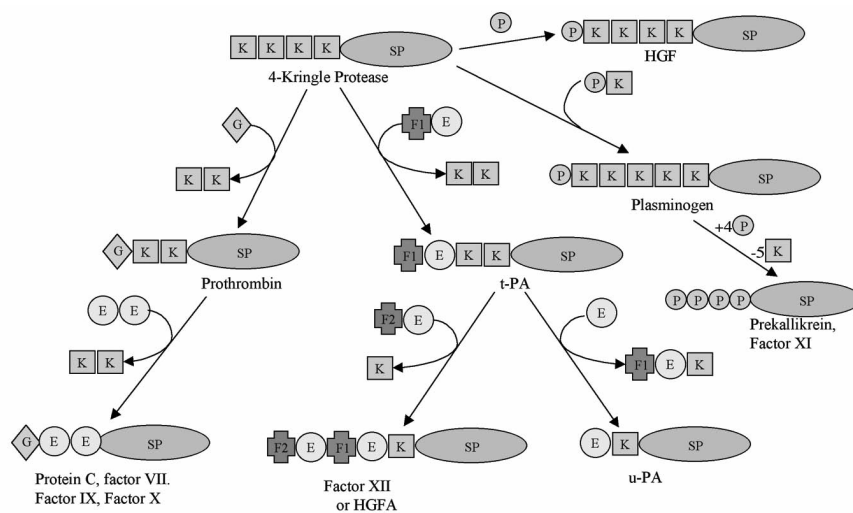

Figure 2. Step-by-step rearrangement of domains during early vertebrate evolution that could have given rise to assorted clotting factor proteases. Other pathways are possible. (K) Kringle, (G) GLA domain, (E) EGF domain, (SP) serine protease, (P) PAN domain, (F1) fibronectin type I, (F2) fibronectin type II. (Adapted from Jiang and Doolittle 2003.)

Table 1. Chordate genome sequences used in this study

Protochordates	*Ciona intestinales*	sea squirt
	Branchiostoma floridae	amphioxus
Jawless fish	*Petromyzon marinus*	lamprey
Bony fish	*Fugu rubripes*	puffer fish
Amphibian	*Xenopus laevis*	frog
Reptile	*Anolis carolinensis*	green anole lizard
Bird	*Gallus gallus*	chicken
Monotreme	*Ornithorrhychus anatinus*	platypus
Marsupial	*Monodelphis domestica*	opossum
Mammal	*Homo sapiens*	human

dence of a fibrin clot (Fry 1909). This view was borne out by a systematic survey of genes in the sea squirt (*Ciona intestinalis*), with no bona fide clotting factor genes being found (Jiang and Doolittle 2003).

The amphioxus genome is somewhat larger than that of the sea squirt (540 Mb vs. 330 Mb), the latter having lost many genes, and it was therefore important to search the amphioxus genome for putative clotting factor genes in case the reduced genome of sea squirts had lost them. In the end, no genes for authentic coagulation factors were found in amphioxus. There *are* numerous genes for fibrinogen-related domains (FREDs) in both genomes (sea squirt and amphioxus), some of which have been cloned (Fan et al. 2008), but none are clustered with other domains that are characteristic of fibrinogen molecules.

Similarly, there are genes that have sequences which resemble the serine protease domains of vertebrate prothrombins, but they do not have coding regions for the subsidiary domains that typify this factor, such as the GLA domain (Fig. 2).

Because fibrinogen, the mainstay of the clot, has never been found in any protochordate by any method, we must conclude that the invention of thrombin-catalyzed fibrin formation took place in the ~50–100-million-year window between the appearance of protochordates and the jawless fish.

JAWLESS FISH

Jawless fish are the earliest appearing vertebrates. Two genera are extant: hagfish and lamprey. Although there has been a long-standing debate as to whether these two

are monophyletic, current opinion favors separate divergences, with hagfish being the more primitive (Janvier 1996). Genomic data for hagfish are sparse, however, and for the moment, our study of clotting factors in this group is mostly limited to the lamprey. Even so, the system in these creatures is decidedly simpler than in mammals and serves as an illustration of how such a system can become more complex.

As an example, mammals have two large, multidomain nonenzyme proteins—factors V and VIII—that have key roles in thrombin generation. These homologous proteins are descended from another blood plasma protein called ceruloplasmin. Sequence differences aside, factors V and VIII differ from ceruloplasmin and some other homologs in having two discoidin domains at their carboxy-terminal ends. The discoidin domains bind to platelet (thrombocyte) surfaces, whereas the main bodies of these proteins are situated normal to the surface and serve as "holders" for factor X in the case of factor V, or factor IX in the case of factor VIII, and bring about the proper orientation that allows these proteases to convert prothrombin to thrombin.

The situation in lampreys is simpler in that only one of these protease-holder duos appears to be present (Fig. 3). Neither factor IX (protease) nor factor VIII (holder protein) was found by searching the lamprey genomic databases (Doolittle et al. 2008), and efforts to clone these factors have been negative, even though other related clotting factors were found (Kimura et al. 2009).

The question may be asked, how can new factors be introduced into an existing pathway? It was long ago suggested that in the case of clotting pathways, new factors that are the products of gene duplications could easily be sandwiched into the middle of pathways where they initially were only performing the same operation as the original gene product. Only a few amino acid replacements were likely needed to broaden the proteolytic specificity to the point where the duplicon could itself activate the other surviving gene product (Doolittle 1961). In line with this thinking, all of the vitamin-K-dependent proteases (prothrombin, factors VII, IX, and X, and protein C) cleave after arginine residues in the same general regions of their homologous substrates. Comparison of thrombin generation in lampreys with that in other vertebrates illustrates such a scheme of events perfectly (Fig. 3).
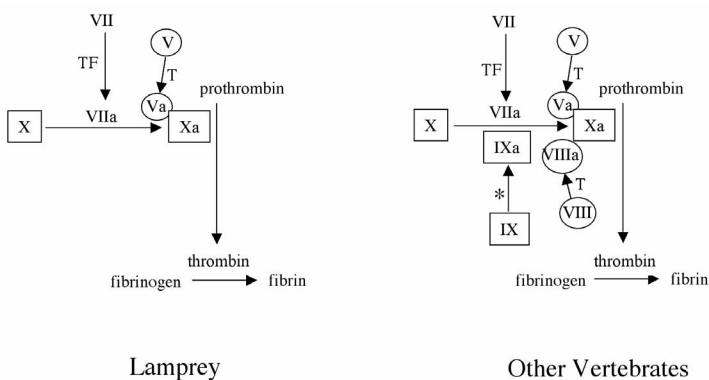

Figure 3. Comparison of thrombin generation in a jawless fish (lamprey) with that in other vertebrates. The lamprey system lacks factor IX (a protease) and factor VIII (a cofactor protein). (T) Thrombin, (TF) tissue factor. The asterisk (*) denoting the activator for converting factor IX to IXa would be factor VIIa in fish and a contact phase factor in tetrapods. (Adapted from Doolittle et al. 2008.)

Lamprey

Other Vertebrates

In passing, we may also note that the lamprey also has some "extra" factors. The genome data make it clear that there are three factors VII, just as was found in the puffer fish (Davidson et al. 2003; Jiang and Doolittle 2003). Additionally, the lamprey has two factors X, the duplication in this case having occurred well after the divergence of lampreys from the rest of the vertebrate lineage.

THE EMERGENCE OF CONTACT FACTORS

Historically, thrombin generation has been regarded as a two-pronged process made up of an *extrinsic* pathway and an *intrinsic* pathway, although it is now recognized that considerable "cross-talk" exists between the two pathways. Classically, the extrinsic pathway begins with tissue factor combining with factor VII to activate factor X in conjunction with factor V. The intrinsic system involves platelet stimulation of factor IX in conjunction with factor VIII to activate factor X similarly. In mammals, factor IX is greatly activated by a subpathway called the "contact factor system," provoked by contact with "foreign" surfaces, as might occur in an injury circumstance. The contact system is composed of a high-molecular-weight kininogen (HMWK) and three proteases: factors XI and XII and prekallikrein (Fig. 4).

Although HMWK is found in fish (Zhou et al. 2008), the contact system proteases do not make their phyletic appearance until after the evolution of tetrapods (Ponczek et al. 2008). Moreover, the HMWK found in fish lacks the particular domain thought to be responsible for the contact activation, with that domain first appearing in HMWK at the level of tetrapods (Zhou et al. 2008).

The contact factor proteases obviously arose from gene duplications of other proteins. Factor XII has the same domainal arrangement as a protease known as hepatocyte growth factor activator (HGFA in Fig. 2). The amino acid sequences of prekallikrein and factor XI, which have strings of four PAN domains (Tordai et al. 1999), cluster with those of plasminogens and hepatocyte growth factor (HGF), both of which have PAN domains at their amino termini (Fig. 2).

Interestingly, reptiles have the gene for factor XII, but birds have lost it (Table 2). There has been a long-standing debate about whether or not factor XII in mammals is truly essential to effective blood clotting, a concern that
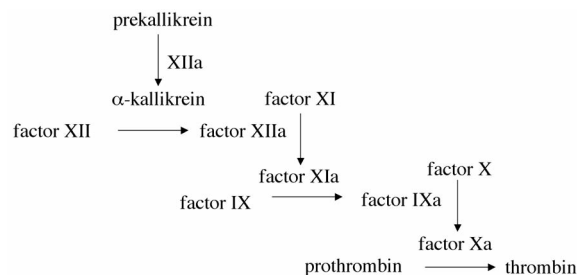
intensified with the finding that it is also absent in some marine mammals (Robinson et al. 1969).

MORE REGULATION IN MAMMALS

Several additional regulatory devices for the clotting scheme have evolved among mammals during the past 100 million years. One of these was the appearance of a large protein called apolipoprotein(a), or apo(a), a recognizable paralog of plasminogen. But whereas plasminogen has five kringles, apo(a) is a polymorphic protein that can have as many as 50 kringles, almost all of which resemble a single kringle found in plasminogen (Lawn et al. 1997). It also has an inactive relic of a serine protease domain, having lost some key active site residues. The protein is intimately bound up with low-density lipopoteins (LDLs), a well-known threat to the vascular circulation. Somehow, apo(a) modulates the destruction of clots, interfering with the action of plasmin. Remarkably, it seems to have evolved independently on two occasions, once among insectivores such as the hedgehog, and again among higher primates (Lawn et al. 1997).

EVOLUTION OF THE FIBRINOGEN γ′ CHAIN

Another recently evolved device for regulating clot formation is the result of a new splicing site in the gene for the γ chain of fibrinogen. As it happens, the splicing at this site in humans is not wholly efficient, and 5%–10% of the time, a nonsplice leads to an alternative chain. The alternative chain, called γ′, has been shown to have two important functions; binding factor XIII in fibrinogen and binding thrombin once it has become a part of fibrin. Not all species have active γ′ sequences exhibiting both functions; some species such as the mouse do not bind thrombin at this site at all (Mosesson et al. 2009). WGS studies now reveal that the intron involved in this particular γ-chain splicing did not make its appearance until the evolution of reptiles and birds (Doolittle et al. 2009).

THE ORIGINS OF VERTEBRATE CLOTTING

The total absence of clotting factors in basal chordates and the existence of a relatively complex system in lampreys indicate that it is not likely we will find any extant creatures with truly rudimentary clotting systems, leaving us to speculate on how the system might have gotten started. With regard to the numerous serine proteases involved, sequence-based phylogenetic trees invariably show thrombin to be earliest appearing of the GLA-containing proteases. Moreover, in mammals, thrombin is known to have a relatively broad range of activities, targeting not only fibrinogen, but also factors V, VIII, and XIII and platelets. It seems unlikely that thrombin and fibrinogen would appear simultaneously; more reasonably, one already existed with an alternative function. For

Figure 4. In mammals, contact factor proteases lead to activation of factor IX.

Table 2. Occurrence of genes for contact factor proteases and some paralogs

Organism	Factor XI	Prekallikrein	Factor XII	HGFA	HGF
Human	yes	yes	yes	yes	yes
Opossum	yes	yes	yes	yes	yes
Platypus	no	yes	yes	yes	yes
Chicken	no	yes	no	yes	yes
Green lizard	no	yes	yes	yes	yes
Frog	no	yes	yes	yes	yes
Zebra fish	no	no	no	?	yes
Puffer fish	no	no	no	yes	yes
Lamprey	no	no	no	yes	yes

(HGF) Hepatocyte growth factor; (HGFA) hepatocyte growth factor activator. HGFA and factor XII have the same domainal arrangement. Factor XI, prekallikrein, and HGF contain PAN modules. (Adapted and updated from Ponczek et al. 2008.)

example, fibrinogen may have had a role in cell–cell interactions, a property of many proteins with fibrinogen-related domains.

A more likely scenario, however, is that thrombin had an early role in agglutinating thrombocytes by proteolyzing cell surface proteins, something it is known to do today, attacking a set of G-protein-coupled receptors called PAR proteins (Vu et al. 1991; Coughlin 2005). According to this scenario, a tissue factor would become exposed during the course of injury, activating prothrombin that would then clump cells which were the ancient ancestors of mammalian platelets (Doolittle 1993). A GLA domain could have helped to keep thrombin localized on the surface of the thrombocytes.

The subsequent appearance of fibrinogen would allow thrombin to broaden its attack, generating a more durable clot composed of fibrin (Fig. 5). Duplications of the prothrombin gene would lead to the appearance of factors VII, X, and eventually IX.

The main problem with this simple scenario has to do with the kinds of subsidiary domains found in thrombin. Besides the GLA domain, thrombin has two kringle domains, usually thought to have an affinity for fibrin. The kinds of domains that interact with tissue factor, however, are the EGF domains found in factors VII, X, and IX. It may be that there was much domain shuffling in the early stages and that thrombin originally had EGF domains, or no peripheral domains at all. Kringle-containing serine proteases have been identified in both sea squirts (Jiang and Doolittle 2003) and amphioxus (Liu and Zhang 2009)

and may have served as starting points for many of the proteases later to be involved in clotting (Fig. 2).

WHY DID THE CLOTTING SYSTEM BECOME MORE COMPLEX?

It is fair to ask why the clotting system is more complex in mammals than it is in earlier diverging vertebrates. Indeed, the quality and character of the fibrin clots generated in fish and mammals do not appear to be significantly different. The answers must have to do with regulating the response under a wider spectrum of environments. Higher blood pressure, more complicated cardiovascular systems, higher metabolic rates, and new organs such as lungs or the placenta all introduce more challenges for maintaining the balance between liquidity and gelation. Having more components in the amplification system may also make it possible to deliver a maximum response with a smaller input stimulus, and this in turn would necessitate having a faster turn-off or lytic response.

CONCLUSIONS

Vertebrate blood clotting evolved in parallel with the development of a pressurized and closed circulatory system with its all important red cell content. Its origin and evolution occurred within the 50–100-million-year window between the appearance of protochordates and vertebrates (Fig. 6). A census of component genes in the
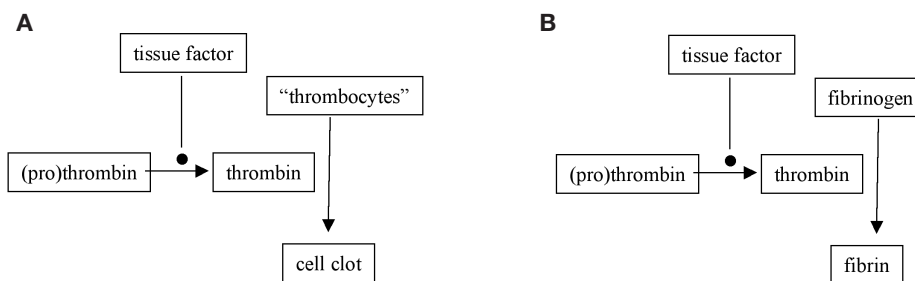

Figure 5. Two simple systems depicting how vertebrate clotting may have arisen. In *A*, a serine protease activated by exposure of a tissue factor provokes cell clumping at the site by attacking a cell surface protein. In *B*, the specificity of the protease is broadened to include an attack on fibrinogen that leads to its polymerization into fibrin.
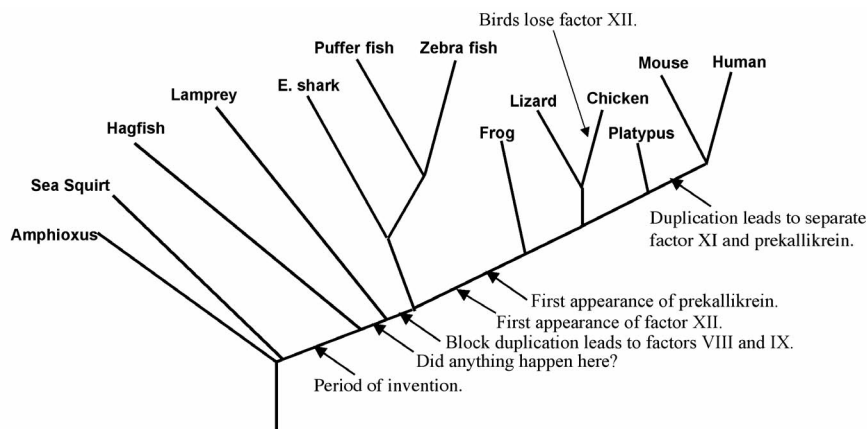
Figure 6. Time-line phylogeny for appearance (and disappearance) of various clotting factors during the course of vertebrate evolution.

lamprey genome has revealed a simpler, but still complex, system in jawed vertebrates. The system is expanded in tetrapods, more factors being added step by step until the time of marsupials. Even then, some additional regulatory devices have appeared along lineages leading to various mammalian groups, including the vulnerability of fibrin to fibrinolytic enzymes, a process that can be regulated by alternative splicing.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Coughlin SR. 2005. Protease-activated receptors in hemostasis, thrombosis and vascular biology. *J Thromb Haemost* **3:** 1800–1814.

Davidson CJ, Tuddenham EG, McVey JH. 2003. 450 Million years of hemostasis. *J Thromb Haemost* **1:** 1487–1494.

Doolittle RF. 1961. "The comparative biochemistry of blood coagulation." PhD thesis, Harvard University, Cambridge.

Doolittle RF. 1993. The evolution of vertebrate blood coagulation: A case of Yin and Yang. *Thromb Haemost* **70:** 24–28.

Doolittle RF, Feng D-F. 1987. Reconstructing the evolution of vertebrate blood coagulation from a consideration of the amino acid sequences of clotting proteins. *Cold Spring Harbor Symp Quant Biol* **52:** 869–874.

Doolittle RF, Surgenor DM. 1962. Blood coagulation in fish. *Am J Physiol* **203:** 964–970.

Doolittle RF, Jiang Y, Nand J. 2008. Genomic evidence for a simpler clotting scheme in jawless vertebrates. *J Mol Evol* **66:** 185–196.

Doolittle RF, Hong S, Wilcox D. 2009. Evolution of the fibrinogen γ′ chain: Implications for the binding of factor XIII, thrombin and platelets. *J Thromb Haemost* (in press).

Fan C, Zhang S, Li L, Chao Y. 2008. Fibrinogen-related protein from amphioxus *Branchiostoma belcheri* is a multivalent pattern recognition receptor with a bacteriolytic activity. *Mol Immunol* **45:** 3338–3346.

Fry HJB. 1909. Blood platelets and coagulation of the blood in marine chordates. *Folia Hematol* **8:** 467–503.

Hughes AL. 2000. Modes of evolution in the protease and kringle domains of the plasminogen-prothrombin family. *Mol Phylogenet Evol* **14:** 469–478.

Janvier P. 1996. The dawn of the vertebrates: Characters versus common ascent in current vertebrate phylogenies. *Paleontology* **39:** 259–287.

Jiang Y, Doolittle RF. 2003. The evolution of vertebrate blood coagulation as viewed from a compassion of puffer fish and sea squirt genomes. *Proc Natl Acad Sci* **100:** 7527–7532.

Kimura A, Ikeo K, Nonaka M. 2009. Evolutionary origin of the vertebrate blood complement and coagulation systems inferred from liver EST analysis of lamprey. *Dev Comp Immunol* **33:** 77–87.

Lawn RM, Schwartz K, Patthy L. 1997. Convergent evolution of apolipoprotein(a) in primates and hedgehog. *Proc Natl Acad Sci* **94:** 11992–11997.

Liu M, Zhang S. 2009. A kringle-containing protease with plasminogen-like activity in the basal chordate *Branchiostoma belcheri*. *Biosci Rep* (in press).

Mosesson MW, Cooley BC, Hernandez I, Diorio JP, Weiler H. 2009. Thrombosis risk modification in transgenic mice containing the human fibrinogen thrombin-binding γ′ chain sequence. *J Thromb Haemost* **7:** 102–110.

Ponczek M, Gailani D, Doolittle RF. 2008. Evolution of the contact phase of vertebrate blood coagulation. *J Thromb Haemost* **6:** 1876–1883.

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453:** 1064–1071.

Robinson AJ, Kropatkin M, Aggeler PM. 1969. Hageman factor (fXII) deficiency in marine mammals. *Science* **166:** 1420–1422.

Tordai H, Banyai L, Patthy L. 1999. The PAN module: The N-terminal domains of plasminogen and growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. *FEBS Lett* **461:** 63–67.

Vu T-K, Hung DT, Wheaton VI, Coughlin SR. 1991. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* **64:** 1057–1068.

Zhou L, Li-Ling J, Huang H, Fei M, Li Q. 2008. Phylogenetic analysis of vertebrate kininogen genes. *Genomics* **91:** 129–141.