Received: 19 July 2022 | Accepted: 22 September 2022

DOI: 10.1111/jth.15897

ORIGINAL ARTICLE

Comparative sequence analysis of vitamin K-dependent coagulation factors

Bosko M. Stojanovski | **Enrico Di Cera**

Edward A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, St. Louis, Missouri, USA

Correspondence
Enrico Di Cera, Edward A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, 1100 S Grand Blvd, St. Louis, MO 63104 USA.
Email: enrico@slu.edu

Funding information
National Heart, Lung, and Blood Institute, Grant/Award Number: HL049413, HL139554 and HL147821

Abstract

Background: Prothrombin, protein C, and factors VII, IX, and X are vitamin K (VK)-dependent coagulation proteins that play an important role in the initiation, amplification, and subsequent attenuation of the coagulation response. Blood coagulation evolved in the common vertebrate ancestor as a specialization of the complement system and immune response, which in turn bear close evolutionary ties with developmental enzyme cascades. There is currently no comprehensive analysis of the evolutionary changes experienced by these coagulation proteins during the radiation of vertebrates and little is known about conservation of residues that are important for zymogen activation and catalysis.

Objectives: To characterize the conservation level of functionally important residues among VK-dependent coagulation proteins from different vertebrate lineages.

Methods: The conservation level of residues important for zymogen activation and catalysis was analyzed in >1600 primary sequences of VK-dependent proteins.

Results: Functionally important residues are most conserved in prothrombin and least conserved in protein C. Some of the most profound functional modifications in protein C occurred in the ancestor of bony fish when the basic residue in the activation site was replaced by an aromatic residue. Furthermore, during the radiation of placental mammals from marsupials, protein C acquired a cysteine-rich insert that introduced an additional disulfide in the EGF1 domain and evolved a proprotein convertase cleavage site in the activation peptide linker that also became significantly elongated.

Conclusions: Sequence variabilities at functionally important residues may lead to interspecies differences in the zymogen activation and catalytic properties of orthologous VK-dependent proteins.

KEYWORDS
blood coagulation, evolution, protein C, prothrombin, vitamin K-dependent clotting factors

Manuscript handled by: Alan Mast Final decision: Alan Mast, 22 September 2022

1 | INTRODUCTION

Blood coagulation is initiated either extrinsically through exposure of tissue factor at the site of injury and complex formation with factor (F) VIIa[1] or intrinsically through contact activation where a cascade of proteolytic reactions (involving the proteases FXIIa, kallikrein and FXIa) results in FIX activation.[2] Both pathways lead to FX activation and subsequent thrombin formation, which amplifies its own generation through activation of FV and FVIII. Once activated, FVIIIa associates with FIXa to increase formation of FXa, which in turn assembles with FVa on membrane surfaces to form the prothrombinase complex responsible for an explosive conversion of prothrombin into thrombin.[1] Besides its procoagulant function that promotes fibrin clot formation, thrombin also performs an important anticoagulant function through activation of protein C (PC) in a reaction enhanced by thrombomodulin.[3,4] Activated PC (APC) downregulates the coagulation response through inactivation of FVa and FVIIIa.[4]

Protein C, FVII, FIX, and FX are vitamin K (VK)-dependent coagulation factors whose multidomain architecture comprises an N-terminal γ-carboxyglutamic acid (Gla) domain that promotes membrane binding, two epidermal growth factor (EGF1 and EGF2) domains that serve as spacers and a serine protease domain that hosts the active site (Figure S1).[5–12] Prothrombin also belongs to the family of VK-dependent coagulation proteins, but its structure comprises a Gla, kringle 1, kringle 2, and serine protease domains (Figure S1).[13,14] All VK-dependent coagulation proteins circulate in plasma as inactive zymogens. Activation proceeds through proteolytic cleavage at a conserved site, creating a new N-terminus at position 16 (chymotrypsin numbering) that ion pairs to the conserved D194 and organizes the active site architecture.[15–17] Activation of PC, FVII, FIX, and FX proceeds with retention of the Gla and EGF domains, which remain connected to the protease domain through a disulfide linkage.[5–12] In contrast, the conversion of prothrombin to thrombin proceeds with shedding of the Gla and kringle domains.[14,18,19] The physiological importance of releasing these auxiliary domains is documented by the compromised procoagulant activity of meizothrombin, a thrombin precursor that retains the Gla and kringle domains.[20,21]

Blood coagulation evolved as a specialization of the complement system and immune response, which in turn bear close evolutionary ties with developmental enzyme cascades.[22,23] The coagulation cascade evolved in the common vertebrate ancestor through a series of gene duplication and exon shuffling events during a short evolutionary time that separates the divergence of vertebrates from urochordates.[24–30] Primitive vertebrates apparently had an abridged version of the coagulation cascade, lacking some important coagulation factors (i.e., FIX, FVIII, and contact pathway proteins), as documented by the absence of such genes in jawless fish which are the earliest diverging extant vertebrates.[24–26] As vertebrates radiated into more classes, the coagulation cascade increased in complexity, resulting in the introduction of novel factors whose role was to preserve efficient thrombin formation.[24–26,31,32] Previous studies on the evolution of the coagulation cascade have focused on when specific coagulation factors first emerged during vertebrate expansion.[24–32] Consequently, there are no comprehensive comparative reports that measure the conservation level of functionally important residues among VK-dependent coagulation factors from different vertebrate lineages. Here, we address these questions through analysis of >1600 primary sequences of VK-dependent coagulation proteins obtained from all classes of vertebrates that diversified during specific evolutionary periods.

Essentials

- Vitamin K-dependent factors play an important role in the coagulation cascade.
- A comprehensive sequence analysis of vitamin K-dependent coagulation factors is presented.
- Prothrombin is the most and protein C the least conserved protein across vertebrate lineages.
- Sequence variabilities modulate the functional properties of orthologous proteins.

2 | METHODS

The number of predicted primary sequences downloaded from the National Center for Biotechnology Information and UniProt databases were: fish ([FVII = 71], [FIX = 100], [FX = 105], [PC = 122], [prothrombin = 76]), amphibians ([FVII = 9], [FIX = 12], [FX =], [PC =], [prothrombin = 11]), reptiles ([FVII = 28], [FIX = 29], [FX = 30], [PC = 30], [prothrombin = 27]), birds ([FVII = 95], [FIX = 100], [FX = 69], [PC = 93], [prothrombin = 98]), and mammals ([FVII = 112], [FIX = 98], [FX = 100], [PC = 76], [prothrombin = 99]). All accession IDs are provided as supporting information. Multiple sequence alignments were performed using Constraint-based Multiple Alignment Tool (COBALT).[33]

3 | RESULTS

3.1 | Conservation of the zymogen activation site

The VK-dependent coagulation proteins of the EGF type are synthesized as inactive zymogens that are activated through proteolytic cleavage at position 15 (the scissile site occupies the P1 position; Table S1 presents chymotrypsin and Human Genome Variation Society [HGVS][34] numbering).[8–12] The position of the activation site is highly conserved among >1300 primary sequences of PC, FVII, FIX, and FX, with the scissile residue being always spaced by exactly 12 amino acids from the N-terminus of the perfectly conserved P28 in the protease domain (Figures S2–S5). G19 can be used as another marker in locating the position of the scissile site, although

this residue is less conserved than P28. Our analysis reveals that the zymogen activation site is predominantly occupied by a basic residue (e.g., Arg and Lys in snake and lizard PC; Figure 1). A notable exception is the presence of Trp at the scissile site in PC from ray-finned fish and the presence of Tyr in PC from lobe-finned fish (Figures 1 and S2). Strikingly, the predicted PC sequences of all 116 species of ray-fins express Trp at the zymogen activation site. Because coagulation proteases have trypsin-like specificity and preferentially cleave at basic rather than aromatic residues, it is unclear what enzyme activates PC in bony fish. Unlike bony fish, the earlier diverging cartilaginous fish express PC with P1-Arg at the zymogen activation site (Figure S2).

Activation of prothrombin into thrombin requires cleavage at two scissile sites and proceeds along two alternative pathways, depending on whether the cleavage is initiated at the R271 or R320 site (Figure S1).[18,19] Initial cleavage at R320 results in formation of the active intermediate meizothrombin, whereas initial cleavage at R271 leads to shedding of the auxiliary Gla and kringle domains and formation of the zymogen prethrombin-2. Subsequent cleavage at the alternative bond generates thrombin as final product in either pathway. The activation of prothrombin along the prethrombin-2 and meizothrombin pathways appears to be shared by all vertebrates as R271 and R320 are absolutely conserved in >300 primary sequences (Figure 2). Once thrombin is generated, the newly created amino terminus is attacked by autoproteolytic cleavage at R284 to remove 13 amino acids from the A chain.[35] With some minor exceptions, position 284 is occupied by a basic residue (Figure 2), implicating evolutionary conservation of the autoproteolytic site. However, the propensity for auto-proteolysis at position 284 likely varies across vertebrates, because amphibians, birds, and many mammals contain an acidic residue at position P2 (Figure 2), which should compromise cleavage by thrombin.[36,37] Interestingly, regardless of whether thrombin is generated after cleavage at R271 or R284, there appears to be strong selective pressure for a Thr (or less commonly Ser) as the first residue in the newly created N-terminus (Figure 2), although the functional significance for such requirement is yet to be elucidated.

The rate of zymogen activation is strongly influenced by the physicochemical properties of the residues that precede the scissile site. Among the immediate residues preceding the activation site, it is only the P2 position that is partly conserved across vertebrates, whereas the P3 and P4 residues are highly variable (Figures 1 and 2). The P2 positions preceding the R271 and R320 cleavage sites of prothrombin and the activation site of FVII are almost always occupied by Gly (exceptions are fish; Figures 1 and 2). Because FXa shows a preference for substrates with P2-Gly[38] and activates prothrombin and FVII with highest specificity in mammals,[18,39] it is likely that FXa also activates these zymogens in most vertebrates. The P2 position in the activation sites of FIX and FX is predominantly occupied by residues with branched side chains (Thr, Ile, and Val), known to restrict the flexibility of the main chain torsion angles (Figure 1). This imposed rigidity around the scissile site probably influences the zymogen activation rate, considering the evolutionary conservation of such residues at the P2 position in FIX and FX. An interesting exception is FIX from ray-finned fish, on which we noted P2-Lys in nearly half of sequences (Figure 1). The presence of basic residues at the P2 position is often seen among substrates of proprotein convertases which are involved in intracellular processing of pro-peptides.[40] This

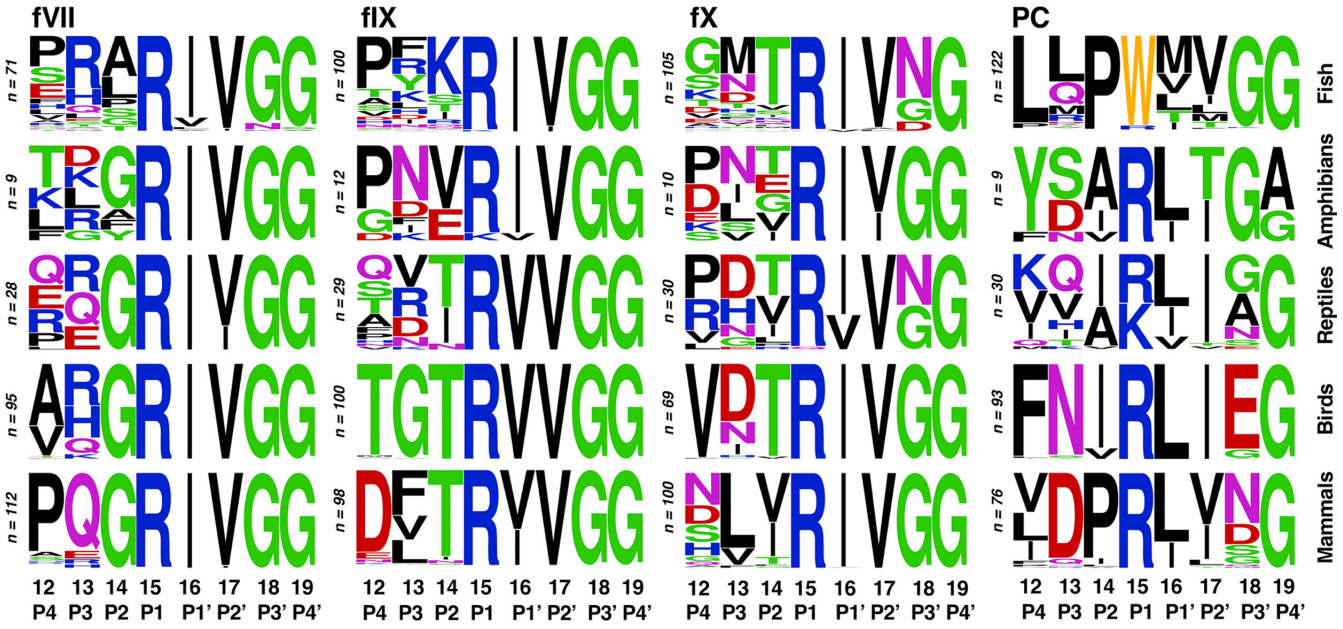


FIGURE 1 The residues surrounding the zymogen activation sites of FVII, FIX, FX, and PC are shown in a web logo format. All VK-dependent coagulation proteins of the EGF type are synthesized as inactive zymogens and are activated through proteolytic cleavage at position 15 (chymotrypsin numbering; see Table S1 for HGVS numbering). Based on the nomenclature of Schechter and Berger, the scissile residue occupies the P1 position. Following activation, the newly created N-terminus becomes inserted into the active site where residues 16 (P1′) ion-pairs with D194. The number of analyzed sequences from each vertebrate class are indicated on the *Y*-axis.

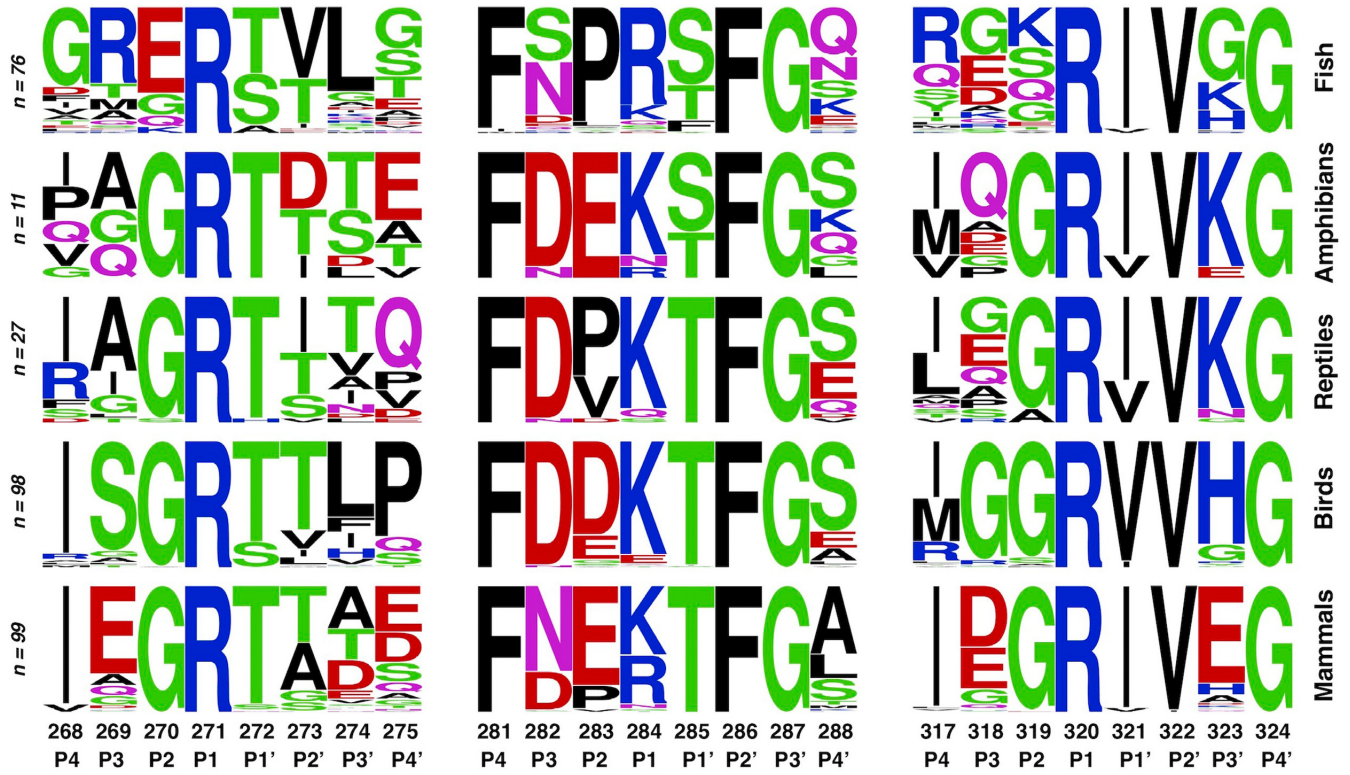


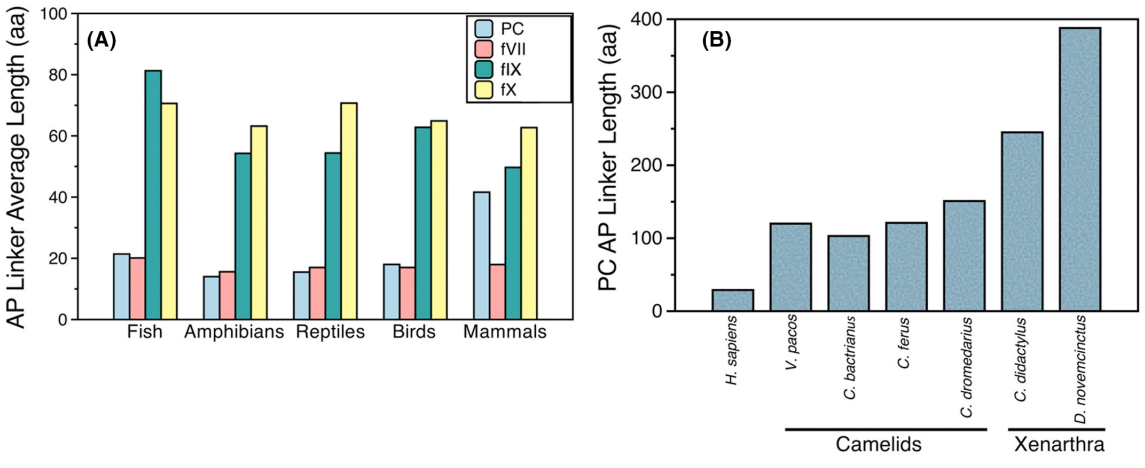
FIGURE 2 The residues surrounding the 271 (left), 284 (middle), and 320 (right) cleavage sites of prothrombin are shown in a web logo format. Activation of human prothrombin requires alternative cleavage at positions R320 and R271 to generate the protease thrombin. Once thrombin is generated, the newly created amino terminus is susceptible to autoproteolytic cleavage at the residue that corresponds to R284 of human prothrombin. The numbering of residues corresponds to the positions of human mature prothrombin. Based on the nomenclature of Schechter and Berger, the scissile residue occupies the P1 position. Residue 320 occupies the analogous position as residue 15 in other serine proteases and cleavage at this site results in insertion of residue 16 (i.e., 321) into the active site and organization of the active site architecture. The number of analyzed sequences from each vertebrate class are indicated on the *Y*-axis.

raises the possibility that FIX might be activated prior to secretion in the circulatory system of fish due to intracellular processing. Finally, an interesting example of sequence convergence can be noted at the P2 position in the zymogen activation site of PC, where the Pro that is present in all species of fish was substituted by a hydrophobic amino acid during the emergence of tetrapods only to converge once again into Pro when placentals radiated from marsupials (Figures 1 and S2). The physiological importance of the sequence convergence process is evident from the appearance of thrombotic complications in carriers of the PC P168L mutation.[41,42]

3.2 | Conservation of the activation peptide linker

All VK-dependent zymogens contain an activation peptide (AP) linker between the EGF2 and protease domains (Figure S1).[8,9,12,43–45] The N-terminal border of the AP linker is defined by the perfectly conserved cysteine that covalently cross-links the EGF2 and serine protease domains, whereas the C-terminal border is delineated by the zymogen activation site (i.e., residue 15). Among vertebrates, the AP linkers of FIX and FX are significantly longer than those of FVII and PC, with average lengths that can be nearly four times greater (Figure 3A). Throughout most of their evolutionary history, FVII and PC had AP linkers of nearly identical lengths, but the similarity in length changed with the emergence of placental mammals when the AP linker of PC became elongated (Figure 3A). The AP linker of PC among placental mammals comprises about 30 amino acids on average, but some extreme exceptions can be noted among Camelidae and Xenarthra, with the southern two-toed sloth (*Choloepus didactylus*; XP_037706319.1) and nine-banded armadillo (*Dasypus novemcinctus*; XP_012380873.1) having the longest AP linkers identified thus far comprised of 245 and 388 amino acids, respectively (Figure 3B,C). Notably, all placentals have a dipeptide repeat in the AP linker of PC comprised of residues XQ (where X is usually Asp) that appears at precisely spaced intervals. In most placentals, the XQ dipeptide is repeated about two to three times after one residue interval, whereas in the Xenarthra the XQ dipeptide is repeated >30 times at an interval of five residues (Figure 3C).

Before secretion into circulation, the AP linkers of FX and PC are processed intracellularly at a proprotein convertase cleavage site, producing two-chain zymogens whose light and heavy chains remain connected through a disulfide bond (Figure S1).[43–47] An AP is then released when the zymogens are activated through cleavage at position 15.[43–47] In the case of PC, a proprotein convertase cleavage site with consensus sequence K-K-R-3X-K-**R**↓ is only found in the AP linker of placental mammals, with bats being an exception



(C)

```
Human      APGYKLGDDLLQCHPAVKFPCGRPWKRMEKKRSHLKRDTE-------------------- 160
Sloth      APGYKLADDLRECQPVVKFPCGRVAKPTEKKRHKLKRDVGQVGQVNLPTDQVDSPTDQVS 180
Armadillo VPGYQLEDDFRKCQPSVKFPCGRLVKGTDKKRYNMKRDLDHVSPVDLPADQASPPTDQAS 180
          .***:* **: :*:* *******  *  :*** ::***
Human      ------------------------------------------------------------
Sloth      PPTGQVDLPTGQVDSTTDQVYPPTSQLDSPTDQVDPPAYQVDPPTGQLDSPTGQVDS--- 237
Armadillo PPTDQASPPMGQASSPTDQASPPADKARPPTDQASPPTDQASPPTDQASPPMDQASPPTD 240

Human      ------------------------------------------------------------
Sloth      --------------------------------PTDQVNSTTDQVDSPADQVDPPAYQVDP 265
Armadillo QASPPADKARPPTDQASPPTDQASPPADKASPPTDQASPPADQVSPPMDQASPPADQVSP 300

Human      ------------------------------------------------------------
Sloth      PTGQLDSPTGQVD--------------SPTDQVNSTTDQVDPPTDQVDSPTGQVDSPTDQ 311
Armadillo PVDQVSPPINQASPPTDQASPSINQTSPPTDQVSPPADQVSPPVDQVSPPTSQASPPTDQ 360

Human      ------------------------------------------------------------
Sloth      VNSTTDQVDPPTDQVDSTTDQVDSPADQVDPPADQVDPPADQVDPPT------------ 358
Armadillo ASPSINQTSPPVDQVSLPTDQVSPPINQTSPPTDQASPPTNQVSPPTDQASPSMNQTSPP 420

Human      ------------------------------------------------------------
Sloth      ------------------------------------------------------------GQL 361
Armadillo TDQASPPADQASLPTDQVSPSINQTSPPTDRASPSINQTSPPTDRASPPTDQASPSINQT 480

Human      ----------------------------------------DQEDQVDPRLIDGKMTRRGDS 181
Sloth      DSPTDEVRRPP----------------------DQLDQLDPRLNPRLIQGKPTKWGDS 397
Armadillo SPPMDQVSPPTDQASPSINQTSPPTDWASPSINQTSPPTDQKGQFDPRIIQGQPAKMGDS 540
                                                 : . :.:**:*:*: :: ***
Human      PWQVVLLDSKKKLACGAVLIHPSWVLTAAHCMDESKKLLVRLGEYDLRRWEKWELDLDIK 241
Sloth      PWQVILLDSKKKLACGAVLIHTHWVLTAAHCMEDGKKLFVRLGEYDLRRFEKWEVNVDIK 457
Armadillo PWQVILLDSKKKLACGAVLIHASWVLTAAHCMEDAKKLSVSSGEYDLRRPDKWEVNVDIQ 600
          ****:****************  *********::.*** *  ******* :***:::**:
```

FIGURE 3 Length variabilities in the AP linkers of the VK-dependent coagulation proteins across vertebrate classes. (A) Shown are the average number of amino acids comprising the AP linkers of the VK-dependent coagulation proteins. The N-terminal border of the AP linker is defined by the perfectly conserved Cys that forms the disulfide bridge between the EGF2 and serine protease domains, whereas the C-terminal border is delineated by the zymogen activation site (e.g., residue 15). The number of analyzed sequences from each vertebrate class are listed under methods. (B) The AP linker length of PC from members of Camelidae and Xenarthra are significantly longer than the average length observed in most placental mammals. (C) Multiple sequence alignments of PC from humans, southern two-toed sloth, and nine-banded armadillo. The AP linker is shown in orange; the proprotein convertase cleavage and zymogen activation sites are in blue and red, respectively. Note how a dipeptide comprised of XQ (where X is usually Asp) is repeated at exactly spaced intervals. The accession IDs for the primary sequences of PC that were analyzed are: human (*Homo sapiens*; UniProtID: P04070), alpaca (*Vicugna pacos*; XP_031534569.1), wild Bactrian camel (*Camelus ferus*; XP_032335106.1), dromedary camel (*Camelus dromedarius*; XP_031307064.1), Bactrian camel (*Camelus bactrianus*; XP_010953100.1), southern two-toed sloth (*Choloepus didactylus*; XP_037706319.1), and nine-banded armadillo (*Dasypus novemcinctus*; XP_012380873.1).

because of a cysteine substitution that replaces the scissile arginine (Figure 4A,B). All other vertebrates, including monotremes and marsupials, lack the K-K-R-3X-K-**R**↓ consensus sequence in the AP linker (Figure 4A,B). This implicates that the proprotein convertase processing site is a recent evolutionary innovation that first appeared in PC during the radiation of placentals from marsupials. Interestingly, the evolutionary emergence of the K-K-R-3X-K-**R**↓ cleavage site coincided with elongation of the AP linker in placentals, but it is unclear if these sequence modifications were interrelated or coincidental (Figures 3 and 4). The physiological relevance of cleavage at the K-K-R-3X-K-**R**↓ site is not known, with conflicting reports existing about the importance of processing at this site for subsequent activation of PC.[46,48]

The type I proprotein convertase consensus sequence[40] of R-X-K/R-**R**↓ that is located in the AP linker of fX is conserved in all vertebrates, implicating evolutionary emergence in the common vertebrate ancestor (Figure 4A,C). An exception are the ray-finned fish that have either completely lost (66% of species) the consensus sequence in FX or have acquired (44% of species) a modified type III R-X-X-**R**↓ sequence[40] that lacks a basic residue at the putative P2 position and is positioned at a distance of 17 as opposed to 11 amino acids (as in other vertebrates) from the conserved cysteine that holds the two chains together (Figure 4A,C). We further note that crocodilians and some lizards have acquired a single residue insertion in the proprotein convertase processing site of FX (data not shown).

We have also identified a putative type II proprotein convertase processing site[40] with a consensus sequence K**R**↓ or K**K**↓ in the AP linker of FVII (Figure 4A,D). This putative cleavage site is present in all vertebrates except for ray-finned fish and some reptiles (50% of species). However, there is no evidence that the K**R**↓ site is cleaved intracellularly, with human FVII being the only protein of its class that does not release an AP upon zymogen activation.[49] It is unclear if the lack of cleavage at the proprotein convertase site in FVII is caused by the short length of the AP linker or results from the absence of additional basic residues at the P4 (as present in FX) or P6–P8 (as present in PC) positions in the nonprimed region of the K**R**↓ cleavage site. Cleavage of the consensus sequence of PC can be

(A)

	Protein C	Factor VII	Factor X
Jawless Fish		No	RXK/RR
Cartilaginous Fish	No		RXK/RR$^{50\%}$ RXXR$^{25\%}$
Ray-finned Fish	No	No	RXXR$^{44\%}$ No$^{56\%}$
Amphibians	No	KR$^{22\%}$ KK$^{55\%}$	RXK/RR$^{67\%}$ RXXR$^{33\%}$
Reptiles	No	KR$^{32\%}$ KK$^{18\%}$	RXK/RR
Birds	No	KK	RXK/RR
Monotremes	No	KR	RXK/RR
Marsupials	No	KR	RXK/RR
Placentals	KKR3XKR	KR	RXK/RR

(B)
```
EGF2-----CGRPWKRMEKKRSHLKRDTEDQEDQVDPRLIDGKMTRRGDSPWQ-30 Human PC
EGF2------CGRPGKQMEKKRKHLKRDTNQTDQIDPRLVNGKVTRRGESPWQ-30 Dog PC
EGF2---CGRLGKRMEKKRKTLKRDTNQVDQKDQLDPRIVDGQEAGWGESPWQ-30 Bovine PC
EGF2------CGRPWNPKEKKSKTLKCCTGQRDQVDPRLINGKLTLQGESPWQ-30 Bat PC
EGF2---------------CGKPKVKYPTPNFEELRIRISGGKPANKGDSPWQ-30 Koala PC
EGF2-------------------CGRSVVNSPSPFDLRLIDGKKGLKGMSPWQ-30 Platypus PC
EGF2----------------CGRVKTDYMEAEAGFNIRLIEGKAGRRGDSPWQ-30 Mallard PC
EGF2--------------------CGRVVESKLKVAKIIGGNPGKRGDSPWQ-30 Corn snake PC
EGF2--------------------CGKMKIFNSNYSARLTGAKKGKKGDSPWQ-30 Toad PC
EGF2---------------CGVLQVGEPLFQFVNMKPYVMGGTTGTRGHSPWQ-30 Coelacanth PC
EGF2-------------CGRILIHKSSFNTGPIQGLLPWLIGGEVGMKGESPWQ-30 Ray-fin PC
EGF2-CGMLKGIRMLRNTKPHIGKGAALDDFEFLLDPRITKGRVMPKGGSPWQ-30 Shark PC
```

(C)
```
EGF2-CGKQTLERRKRSVAQATSSS...AP   Human fX
EGF2-CGKYTVERRKRSVTEPATQS...AP   Koala fX
EGF2-CGKLTVGRRKRSRELPEEQD...AP   Platypus fX
EGF2-CGKVLVKRKKRSVILPTDSS...AP   Egret fX
EGF2-CGKVYVRRKKRSVISFGNSS...AP   Turtle fX
EGF2-CGRRAGFRFKRSVSEEVDNS...AP   Toad fX
EGF2-CGAIITGDTRTVFRYERQNT...AP   Ray-fin fX
EGF2-CGIAAEHRARRAVVELDPVS...AP   Shark fX
EGF2-CGRIQVQVKKPRSKREASSV...AP   Lamprey fX
```

(D)
```
EGF2----CGKIPILEKRNASKPQGRIVGGKVCPKGECPWQ-30 Human fVII
EGF2----CGRIPVLEKRNGSIPEGRIVGGYACPKGECPWQ-30 Koala fVII
EGF2----CGKIPVLRKRNESKPEGKIVGGYVCPKGECPWQ-30 Platypus fVII
EGF2-----CGKIPVLAKKNATARGRIVGGFICPPGECPWQ-30 Finch fVII
EGF2-----CGKIPVLAKRNKTEEGRIVGGKECPPGVCPWQ-30 Alligator fVII
EGF2-----CGKIPVLEKKDKALDGRIVGGDECLKGECPWQ-30 Caecilian fVII
EGF2-CGSVPVLQEQNKSQQLDPRGRIVGGTECPKGECPWQ-30 Ray-fin fVII
```

FIGURE 4 The proprotein convertase cleavage sites of the VK-dependent coagulation zymogens. (A) Distribution of the proprotein convertase cleavage sites of PC (K-K-R-3X-K-**R**↓), fVII (K**R**↓ or K**K**↓) and fX (R-X-K/R-**R**↓ or R-X-X-**R**↓) in vertebrates. Representative vertebrate sequences containing the consensus sequences of the proprotein convertase cleavage sites (shown in blue) in (B) PC, (C) FX, and (D) FVII. Note how the K-K-R-3X-K-**R**↓ consensus sequence is missing in PC sequences from nonplacentals and how the scissile Arg is mutated in bats. The proprotein convertase cleavage site is located in the AP linker region that is delineated at the N-terminus by the interdomain bridging Cys (orange) and at the C-terminus by the zymogen activation site (red). Shown are also the conserved markers G19 and P28 (green) in the serine protease domain. Complete alignment of sequences and their respective accession IDs are provided as supporting information.

enhanced by introducing Arg at the P4 position[47] or diminished by mutating the basic residues at the P6-P8 positions.[46]

3.3 | Conserved residues in the protease domain

A new N-terminus at position 16 is generated during zymogen activation that ion pairs with D194 in the active site and organizes the oxyanion hole and primary specificity pocket (Table S1 presents chymotrypsin and HGVS[34] numbering).[17,50] Position 16 (P1′ residue) is predominantly occupied by hydrophobic β-branched amino acids such as I16 and V16 (Figures 1 and 2). The evolutionary selection of Leu, Thr, and Met at position 16 in PC (Figure 1) are interesting exceptions considering that such residues are rarely seen among serine proteases.[7,51] It is unclear how the activity of APC is affected by the polar T16 or M16 residues, but the analogous I16T mutation in thrombin destroys the structural integrity of the oxyanion hole and drastically compromises catalysis.[50] The residues at positions 17 (P2′) and 19 (P4′) also have an important role in promoting effective organization of the active site. Residue 17 stabilizes the orientation of D189 in the primary specificity pocket through a set of H-bonds, whereas G19 serves as a hinge that promotes insertion of the newly created N-terminus into the active site.[17] Nearly all coagulation proteases contain a hydrophobic β-branched amino acids at position 17 (V17 and I17) and G19 except for PC from non-caecilian amphibians in which the selection of T17 and A19 probably adversely affects catalysis (Figures 1 and S2).

The accessibility of substrates to the active site is regulated by the flexible 215–217 segment which exists in equilibrium between open (E) and closed (E*) conformations.[52–54] Among the residues that comprise the 215–217 segment, W215 and G216 are perfectly conserved in all VK-dependent proteases (data not shown). E217 is also highly conserved in all proteases except in FVII, where this position is predominantly occupied by K217 or L217 in nonmammalian vertebrates (Table 1). Most mammals express FVII with E217, but in primates and bats the analogous position is occupied by Q217 or V/L217, respectively (Table 1). Considering the importance of residue 217 in affecting the E*-E equilibrium,[53] the variability at this position likely contributes to interspecies differences in how ligands bind to the active site of FVIIa.

The catalytic activity of all coagulation proteases is enhanced by Na^+ binding to a site that is defined by the 186 and 220 loops and is located >15 Å from the active site residues.[3,11,12,55–60] Whether a protease can effectively bind Na^+ depends on the residue that occupies position 225.[61,62] Serine proteases with F/Y225 bind Na^+, whereas those with P225 do not.[61,62] Across vertebrates, Y225 is nearly exclusively present in all sequences of prothrombin, FIX, and FX (Table 1). Y225 is also present in FVII from most vertebrates (Table 1), except for placental mammals that typically have F225 or hydrophobic amino acids (Val, Ile, or Leu are present in rodents,

TABLE 1 Residues at positions 70, 217, and 225 in the serine protease domain of VK-dependent coagulation proteins

	Prothrombin			Protein C			Factor VII			Factor IX			Factor X		
	70	217	225	70	217	225	70	217	225	70	217	225	70	217	225
Jawless fish	K	E	Y	–	–	–	E	K	F	–	–	–	E	E	$Y^{50\%}$ $F^{50\%}$
Cartilaginous fish	K	E	$Y^{84\%}$ $F^{16\%}$	E	E	F	–	–	–	E	$D^{66\%}$ $F^{34\%}$	Y	E	E	Y
Ray-finned fish	K	E	$Y^{97\%}$ $F^{3\%}$	$D^{99\%}$	$E^{95\%}$ $D^{5\%}$	$L^{99\%}$	$E^{98\%}$	$K^{90\%}$ $R^{10\%}$	Y	$E^{99\%}$	$E^{97\%}$	$Y^{94\%}$ $F^{6\%}$	E	E	$Y^{98\%}$
Amphibians	K	E	Y	E	$E^{84\%}$ $D^{16\%}$	F	$E^{55\%}$ $D^{45\%}$	$L^{67\%}$ $K^{33\%}$	Y	$E^{84\%}$ $D^{16\%}$	$E^{67\%}$ $D^{33\%}$	Y	$E^{97\%}$	E	$Y^{50\%}$ $F^{50\%}$
Reptiles	K	E	$Y^{97\%}$ $F^{3\%}$	$K^{65\%}$ $E^{35\%}$	E	F	$E^{79\%}$ $K^{14\%}$	$K^{93\%}$ $R^{7\%}$	Y	E	E	Y	$E^{87\%}$ $K^{7\%}$	E	Y
Birds	K	E	Y	K	E	F	E	$K^{97\%}$ $M^{3\%}$	Y	E	E	Y	E	E	Y
Monotremes	K	E	$Y^{50\%}$ $F^{50\%}$	E	E	F	E	E	Y	E	E	Y	E	E	F
Marsupials	K	E	$Y^{40\%}$ $F^{60\%}$	$E^{84\%}$ $K^{16\%}$	E	F	$E^{80\%}$ $K^{20\%}$	E	Y	E	E	Y	E	E	Y
Placental mammals	K	E	Y	E	E	Y	$E^{98\%}$ $D^{2\%}$	$E^{69\%}$ $Q^{15\%}$ $V/L^{16\%}$	$F^{75\%}$ $V/L/I^{25\%}$	$E^{96\%}$ $K^{4\%}$	E	Y	$D^{83\%}$ $E^{17\%}$	E	$Y^{90\%}$ $F^{10\%}$

Note: Percent values denote number of sequences in which variability at the specified position exists. The total number of analyzed sequences from each vertebrate class is listed under Methods.

afrotheria, Xenarthra, and some bats). PC is an exception insofar as F225 is the most commonly found residue in vertebrates, but in placentals and ray-finned fish the analogous position is occupied by Y225 and L225, respectively (Table 1). Saturation mutagenesis of position 225 in thrombin has shown profound decrease in activity on introduction of hydrophobic residues,[62] which could implicate interspecies differences in catalytic efficiencies between orthologues that carry V/L225 and F/Y225.

Binding of Ca^{2+} to the protease domain is also known to allosterically enhance the activity of all coagulation proteases other than thrombin.[58,59,63,64] The locale for Ca^{2+} binding is the 70-loop in the protease domain where two of the six ligands that coordinate the divalent metal are provided by the carboxylates of E70 and E80. Thrombin does not bind Ca^{2+} because of the presence of K70 in its 70-loop that forms a salt bridge with E80.[3] The lack of a Ca^{2+}-binding site in thrombin appears to be a common feature in all vertebrates because of perfect conservation of K70 (Table 1). On the other hand, the Ca^{2+} binding site is largely conserved in all other VK-dependent enzymes as supported by the presence of an acidic residue at position 70 (Table 1). A notable exception is the presence of K70 among avian and reptilian PC, with the exception of turtles (Table 1 and Figure S2). Because K70 is expected to form a salt bridge with E80, it is likely that the protease domain of avian and reptilian PC cannot bind Ca^{2+}. Such lack of Ca^{2+} binding is seen with PC mutants that have engineered ionic bridges between residues 70 and 80.[65]

Various insertions in the serine protease domain contribute toward the functional specialization of coagulation proteases. Thrombin contains three long insertions that are present in the 60 loop that delineates the S2-subsite, the 186 loop that defines the Na^+-binding site and the flexible autolysis loop that delimits the lower bound of the active site (Figure S6). The lengths of insertions in the 60 and 186 loop are perfectly conserved in all vertebrates, whereas minor variability in the length of the autolysis loop persists (±2 residues; Figure S6). Importantly, these inserts are absent in recently identified serine proteases from nonvertebrate deuterostomes that have high homology with thrombin and likely evolved from a common thrombin-like protease ancestor.[30] Considering their absence in thrombin-like homologues from primitive deuterostomes, it appears that thrombin acquired its unique insertions in the common ancestor of vertebrates, most likely to improve specificity toward fibrinogen which also evolved around the same time period. PC also has an insert in the autolysis loop, but the length of the insert shows high variability among vertebrates (Figure S2). Last, the 170-loop of FVII contains a 5 residue insertion that is found only in mammals (Figure S3). These results corroborate earlier observation about lack of insertion in the 170-loop of zebrafish fVII[66] and demonstrate that the loop became elongated in the common mammalian ancestor.

3.4 | Conserved residues in the light chain

The light chain of the VK-coagulation proteins is composed of the auxiliary Gla and two EGF or kringle domains (Figure S1). EGF domains comprise 40–50 amino acids and the fold is stabilized by three disulfides that pair with 1–3, 2–4, and 5–6 connectivity.[67,68] The six

cysteines that participate in disulfide formation are highly conserved in the EGF1 domains of FVII, FIX, and FX (Table 2). In contrast, when placentals diverged from marsupials, the EGF1 domain of PC acquired a cysteine rich insert after P54 (human mature PC numbering) that introduces two additional cysteines (Figure 5A). Consequently, all placentals have eight as opposed to the six cysteines that are found in the EGF1 domains of PC from other vertebrates (Figure 5A and Table 2). Although its functional significance is not known, the structure of APC reveals that the cysteine-rich insert causes a lateral bulge and creates two-lobed appearance not seen in other EGF1 domains.[7] When it comes to the EGF2 domain, the only deviation in the potential pairing of disulfides is the presence of seven cysteines that we identified in PC sequences from crocodilia, monotremes, and many birds (67% of sequences; Table 3). Whether the extra cysteine modulates the classical pairing of the EGF2 domain in these animals is something that requires experimental validation.

On a structural level, EGF domains are classified into human-like and C1r-like types depending on whether the last cysteine is located on the turn of the β-hairpin (human) or on the minor sheet itself (C1r).[67] On a functional level, EGF domains are differentiated according to their ability to bind Ca^{2+} and on the presence of an erythro-β-hydroxyaspartic acid (Hya) posttranslational modification.[68] All human VK-dependent coagulation proteases express EGF1 domains that are classified into the human-like and Ca^{2+} binding types, whereas the EGF2 domains belong to the C1r-like type and cannot bind Ca^{2+}.[67,68] To examine whether these functionalities have been conserved during vertebrate evolution, the primary sequences were analyzed for the presence of (1) consensus sequence for Ca^{2+} binding comprising **D**-X-X-**Q**-C (located before the first cysteine[68]), (2) consensus sequence for β-hydroxylation comprising C-X-**D/N**-X-X-X-X-**Y/F**-X-C (located between the third and fourth cysteines with the β-hydroxylation occurring at position **D/N**[68]) and (3) whether the EGF domain belongs to the human or C1r-like type. Identification of the EGF domain type can be deduced from the number of residues that are present between the last two cysteine in the sequence because human-like and C1r-like domains nearly always have eight and twelve residues that are interspersed between the last two cysteines, respectively.[67] Using these criteria, we have identified that vertebrates express EGF1 domains that belong to the human-like type and have largely conserved consensus sequences for Ca^{2+} binding and β-hydroxylation (Table 2 and Figures S2–S5). An exception are fish that express FVII whose consensus sequences for Ca^{2+} binding and β-hydroxylation have been largely mutated (Table 2 and Figure 5B). Binding of Ca^{2+} to the EGF1 domain of human FVII is partly coordinated by the side chains of D46, Q49 (from the **D**-X-X-**Q**-C sequence), and Hya63.[69] Because all these residues are mutated in fish (Figure 5B), we predict that Ca^{2+} does not interact with the EGF1 domain of FVII. These three residues are also important in binding of tissue factor to FVIIa,[69] implicating perturbed interactions with the cofactor. Ray-finned fish also express FX whose consensus sequence for Ca^{2+} binding in

TABLE 2 Total number of Cys, domain type, and presence of Ca^{2+} and Hya consensus sequences in the EGF1 domains of VK-dependent coagulation proteins

	Protein C			Factor VII			Factor IX			Factor X		
	#Cys	Ca^{2+} Hya	EGF Type	#Cys	Ca^{2+} Hya	EGF Type	#Cys	Ca^{2+} Hya	EGF Type	#Cys	Ca^{2+} Hya	EGF Type
Jawless fish	–	–	–	6	No^{Ca2+} No^{Hya}	hEGF	–	–	–	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Cartilaginous fish	6	Yes^{Ca2+} Yes^{Hya}	hEGF	–	–	–	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Ray-finned fish	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	No^{Ca2+} No^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	DXXAC Yes^{Hya}	hEGF
Amphibians	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Reptiles	6	DXXQ/MC Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Birds	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	D/NXXQC Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Monotremes	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Marsupials	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF
Placental mammals	8	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF	6	Yes^{Ca2+} Yes^{Hya}	hEGF

Note: The total number of analyzed sequences from each vertebrate class are listed under Methods. Ca^{2+} binding consensus sequence = DXXQC; Hya consensus sequence = CXD/NXXXXY/FXC.

Abbreviations: hEGF, human-like EGF domain type; Hya, erythro-β-hydroxyaspartic acid.

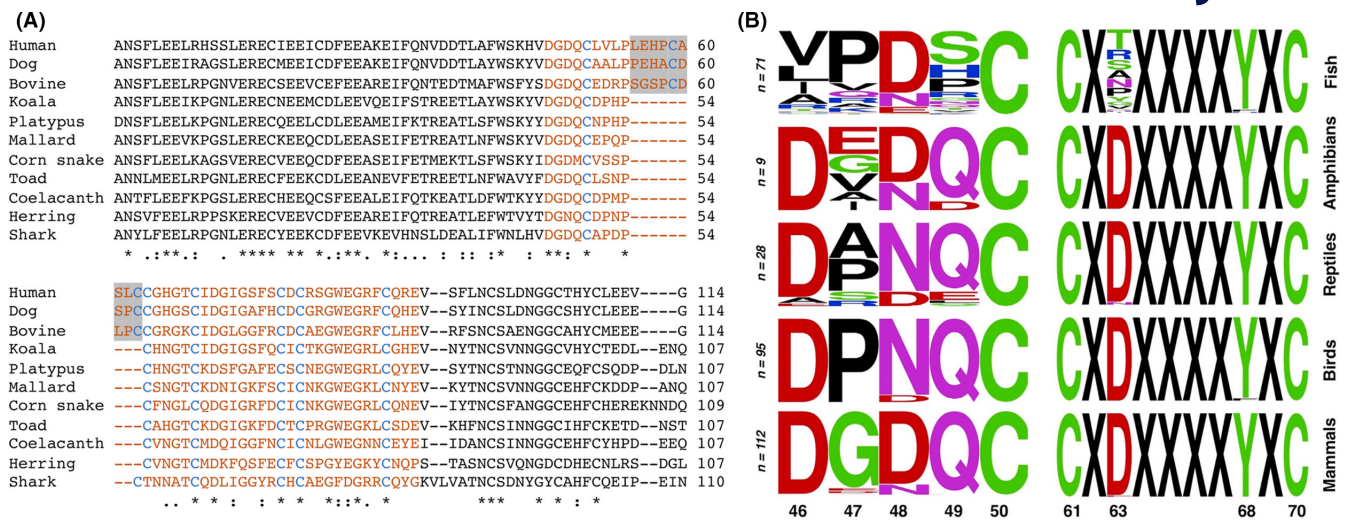


FIGURE 5 Sequence modifications in the EGF1 domains of VK-dependent coagulation factors from different vertebrates. (A) The evolutionary appearance of the cysteine rich insert in the EGF1 domain of PC. Multiple sequence alignments of representative PC sequences demonstrate the presence of a cysteine rich insert in placental mammals that introduces an extra disulfide bond in the EGF1 domain. Such an insert is present in all PC sequences of placental mammals and absent from all other vertebrates (data not shown). The sequence of the EGF1 domain is shown in orange and the cysteine rich insert is highlighted. Complete sequences of PC with their respective accession IDs are provided as supporting information. (B) Web logos of the FVII consensus sequences for Ca^{2+} binding (**D**-X-X-**Q**-C) and β-hydroxylation (C-X-**D**/**N**-X-X-X-X-**Y**/**F**-X-C) that are located in the EGF1 domain. Binding of Ca^{2+} to the EGF1 domain of human FVII is partly coordinated by the side chains of D46, Q49 (from the **D**-X-X-**Q**-C sequence), and D63, with the latter also being the site of β-hydroxylation . Note how the residues in fish that occupy positions 46, 49, and 63 are mutated, indicating lack of Ca^{2+} binding to the EGF1 domain. Numbering of residues corresponds to human mature FVII. The total number of sequences used to create the web logos are shown on the *Y*-axis for each vertebrate class.

the EGF1 domain is modified into D-X-X-**A**-C, probably resulting in compromised binding affinity (Table 2 and Figure S5). In contrast to the modifications seen in the EGF1 domain, we observed strong conservation in all primary sequences that correspond to the EGF2 domain, indicating that the absence of Ca^{2+} binding and β-hydroxylation sites and folding into the C1r-like type are common features that are shared by the EGF2 domains of all vertebrate VK-dependent coagulation proteases (Table 3).

Each kringle domain of prothrombin is stabilized by three disulfides with 1–6, 2–4, and 3–5 connectivity.[14,70] The number of cysteines in each kringle is perfectly conserved across vertebrates (Table S2 and Figure S6). The light chain of human prothrombin also contains three linkers—referred to as Lnk1, Lnk2, and Lnk3—that connect kringle-1 to the Gla domain (residues 47–64), the two kringles (residues 144–169), and kringle-2 to the protease domain (residues 249–284), respectively (Figure S1).[13,14] These three linkers are present in all prothrombin sequences, albeit their sequence and length are poorly conserved (Table S2 and Figure S6). The average length of Lnk1 is longest among ray-fins and shortest in placentals, whereas the opposite is mostly true for the average length of Lnk2 (Table S2). Placentals also have longer than average Lnk3, whereas this linker is shortest in reptiles and amphibians (Table S2). Interestingly, all three linkers of prothrombin have evolved proteolytic sites (Figure S1) that are largely conserved. For instance, the proteolytic sites corresponding to R271 and R284 of human prothrombin are located in Lnk3. Basic residues are typically found at the analogous positions in the respective Lnk3 regions of prothrombin in nearly all vertebrates, implicating functional conservation (Figure 2). A high level of conservation is also seen for the arginine located in Lnk1 (e.g., R54 of human prothrombin) that is susceptible to autoproteolytic cleavage under Ca^{2+}-depleted conditions[71] (Figure S6). Another scissile site that is susceptible to autoproteolytic cleavage is located in Lnk2 (e.g., R155 of human prothrombin) and cleavage at this position results in shedding of the Gla and kringle 1 domains and formation of prethrombin-1.[14,35] Arg is usually present at the analogous position in the Lnk2 region in most vertebrates (Figure S6). Fish and amphibians typically have multiple Lys in the Lnk2 region and is unclear which (if any) of these residues are susceptible to auto-proteolytic cleavage by thrombin.

When it comes to the Gla domain, the total number of glutamates that could be γ-carboxylated varies across vertebrates, with prothrombin, FVII and PC having anywhere between nine and 12 glutamates, whereas this number ranges from 11 to 13 in the Gla domains of FIX and FX. Variability in the number of Gla residues could lead to interspecies differences in binding affinity for membrane surfaces among orthologous proteins. We also noted variability in the ω-loop's length (which is important for promoting interactions with membranes[70,72]), with fIX and prothrombin from most amniotes containing a single residue insertion (Figure S7). Apart from several invariable glutamates, the most conserved positions in the Gla domain are the two cysteines that form a disulfide bond and the aromatic F/YW dipeptide that is part of the so-called helical stretch (Figures S2–S6), which likely have an important role in stabilizing the domain fold.[70]

TABLE 3 Total number of Cys, domain type, and presence of Ca^{2+} and Hya consensus sequences in the EGF2 domains of VK-dependent coagulation proteins

	Protein C			Factor VII			Factor IX			Factor X		
	#Cys	Ca^{2+} Hya	Type	#Cys	Ca^{2+} Hya	Type	#Cys	Ca^{2+} Hya	Type	#Cys	Ca^{2+} Hya	Type
Jawless fish	–	–	–	6	No^{Ca2+} No^{Hya}	cEGF	–	–	–	6	No^{Ca2+} No^{Hya}	cEGF
Cartilaginous fish	6	No^{Ca2+} No^{Hya}	cEGF	–	–	–	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Ray-finned fish	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Lobe-finned fish	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Amphibians	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Reptiles	$6^{87\%}$ $7^{13\%}$	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Birds	$6^{33\%}$ $7^{67\%}$	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Monotremes	$6^{50\%}$ $7^{50\%}$	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Marsupials	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF
Placental mammals	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF	6	No^{Ca2+} No^{Hya}	cEGF

Note: The total number of analyzed sequences from each vertebrate class are listed under methods. Ca^{2+} binding consensus sequence = DXXQC; Hya consensus sequence = CXD/NXXXXY/FXC.

Abbreviations: cEGF, C1r-like EGF domain type; Hya, erythro-β-hydroxyaspartic acid.

4 | DISCUSSION

Much of our knowledge of the functional and structural properties of the VK-dependent coagulation factors has come from studies of the human proteins, which makes it difficult to establish what functional features have been retained during the evolutionary diversification of vertebrates. To understand the level of functional conservation, we analyzed >1600 primary sequences from vertebrates that have diverged during specific evolutionary periods. Overall, we found that prothrombin is the most evolutionary conserved protein, with the functional features that were considered being conserved in all vertebrates. For example, the sites of activation at R271 and R320 of human prothrombin have been identified in all examined sequences, implicating that the activation of prothrombin along the prethrombin-2 and meizothrombin pathways is shared by all vertebrates. The Na^+-binding site is present and Ca^{2+} site absent in all prothrombin sequences, suggesting that these functionalities evolved in the common ancestral protein. So did the insertions in the protease domain that are unique to thrombin and some structural features in the light chain such as the three linkers that connect the various domains.

In contrast, PC represents the least conserved protein, with the positions that were examined often being replaced by residues with drastically different physiochemical properties; however, it remains to be determined whether such substitutions perturb the functional properties of PC considering the context dependent effect of particular mutations. Perhaps the most perplexing modification in PC is the presence of an aromatic residue in the activation site of bony fish, which was identified in all predicted sequences from the 116 species of ray-finned fish that were analyzed. These observations agree with earlier reports about the presence of P1-Trp in the zymogen activation site of PC from *Fugu rubripes*.[26,27] Because all coagulation proteases are trypsin-like enzymes that preferentially cleave at basic residues, the identity of the enzyme that activates PC in bony fish may not be thrombin-like. In mammals, thrombin activates PC in a reaction that is accelerated by thrombomodulin,[3,4] but the R169W mutation abolishes activation by thrombin and promotes effective processing by chymotrypsin.[73] Incubation of fish plasma with activators of PC isolated from snake venom have provided potential hints of zymogen activation,[74,75] but definitive proof as whether thrombin can activate PC in bony fish would require functional studies with purified proteins.

More than any other coagulation protein, PC accumulated many functional modifications when placentals diverged from marsupials. It is during this evolutionary period that PC acquired a cysteine rich insert that introduces an additional disulfide in the EGF1 domain, evolved a proprotein convertase processing site in its AP linker, which also became elongated, its zymogen activation site converged back to the ancient P2-Pro that is present in fish but disappeared when tetrapods emerged and also acquired the F225Y substitution in its Na^+-binding site. There must have been selective pressure to

support introduction of all these functional features in the ancestral placental mammal. Were these modifications acquired to improve zymogen activation or to enhance the enzymatic specificity of one substrate at the expense of another? Answers to such questions would come from comparative structural and biochemical studies between orthologous coagulation proteases from different vertebrates, underscoring the urgency of investing more effort in characterizing the functional properties of VK-dependent coagulation proteins from non-placental vertebrates.

AUTHOR CONTRIBUTIONS

B.M. Stojanovski and E. Di Cera designed the research, analyzed the data, and wrote and reviewed the manuscript; B.M. Stojanovski performed the research.

ACKNOWLEDGMENTS

This study was supported in part by the National Institutes of Health Research Grants HL049413, HL139554, and HL147821.

CONFLICT OF INTEREST

The authors declare no financial interests.

ORCID

Bosko M. Stojanovski https://orcid.org/0000-0002-2016-5369
Enrico Di Cera https://orcid.org/0000-0003-2300-4891

REFERENCES

1. Mann KG. Thrombin generation in hemorrhage control and vascular occlusion. *Circulation*. 2011;124:225-235. doi:10.1161/CIRCULATIONAHA.110.952648
2. Gailani D, Renne T. The intrinsic pathway of coagulation: a target for treating thromboembolic disease? *J Thromb Haemost*. 2007;5:1106-1112.
3. Di Cera E. Thrombin. *Mol Aspects Med*. 2008;29:203-254.
4. Esmon CT. The protein C pathway. *Chest*. 2003;124:26S-32S.
5. Brandstetter H, Bauer M, Huber R, Lollar P, Bode W. X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and hemophilia B. *Proc Natl Acad Sci USA*. 1995;92:9796-9800. doi:10.1073/pnas.92.21.9796
6. Brandstetter H, Kuhne A, Bode W, et al. X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J Biol Chem*. 1996;271:29988-29992.
7. Mather T, Oganessyan V, Hof P, et al. The 2.8 A crystal structure of Gla-domainless activated protein C. *EMBO J*. 1996;15:6822-6831.
8. Perera L, Darden TA, Pedersen LG. Predicted solution structure of zymogen human coagulation FVII. *J Comput Chem*. 2002;23:35-47. doi:10.1002/jcc.1155
9. Venkateswarlu D, Perera L, Darden T, Pedersen LG. Structure and dynamics of zymogen human blood coagulation factor X. *Biophys J*. 2002;82:1190-1206. doi:10.1016/s0006-3495(02)75476-3
10. Stojanovski BM, Pelc LA, Zuo X, Di Cera E. Zymogen and activated protein C have similar structural architecture. *J Biol Chem*. 2020;295:15236-15244. doi:10.1074/jbc.RA120.014789
11. Vadivel K, Bajaj SP. Structural biology of factor VIIa/tissue factor initiated coagulation. *Front Biosci (Landmark Ed)*. 2012;17:2476-2494. doi:10.2741/4066
12. Zogg T, Brandstetter H. Activation mechanisms of coagulation factor IX. *Biol Chem*. 2009;390:391-400. doi:10.1515/BC.2009.057
13. Pozzi N, Chen Z, Di Cera E. How the linker connecting the two kringles influences activation and conformational plasticity of prothrombin. *J Biol Chem*. 2016;291:6071-6082. doi:10.1074/jbc.M115.700401
14. Pozzi N, Chen Z, Pelc LA, Shropshire DB, Di Cera E. The linker connecting the two kringles plays a key role in prothrombin activation. *Proc Natl Acad Sci USA*. 2014;111:7630-7635. doi:10.1073/pnas.1403779111
15. Hedstrom L. Serine protease mechanism and specificity. *Chem Rev*. 2002;102:4501-4524.
16. Page MJ, Di Cera E. Serine peptidases: classification, structure and function. *Cell Mol Life Sci*. 2008;65:1220-1236.
17. Huber R, Bode W. Structural basis of the activation and action of trypsin. *Acc Chem Res*. 1978;11:114-122.
18. Brufatto N, Nesheim ME. Analysis of the kinetics of prothrombin activation and evidence that two equilibrating forms of prothrombinase are involved in the process. *J Biol Chem*. 2003;278:6755-6764. doi:10.1074/jbc.M206413200
19. Stojanovski BM, Di Cera E. Role of sequence and position of the cleavage sites in prothrombin activation. *J Biol Chem*. 2021;297:100955. doi:10.1016/j.jbc.2021.100955
20. Stojanovski BM, Pelc LA, Zuo X, Pozzi N, Di Cera E. Enhancing the anticoagulant profile of meizothrombin. *Biomol Concepts*. 2018;9:169-175.
21. Shaw MA, Kombrinck KW, McElhinney KE, et al. Limiting prothrombin activation to meizothrombin is compatible with survival but significantly alters hemostasis in mice. *Blood*. 2016;128:721-731.
22. Krem MM, Di Cera E. Molecular markers of serine protease evolution. *EMBO J*. 2001;20:3036-3045. doi:10.1093/emboj/20.12.3036
23. Krem MM, Di Cera E. Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem Sci*. 2002;27:67-74.
24. Doolittle RF. Step-by-step evolution of vertebrate blood coagulation. *Cold Spring Harb Symp Quant Biol*. 2009;74:35-40. doi:10.1101/sqb.2009.74.001
25. Doolittle RF. Coagulation in vertebrates with a focus on evolution and inflammation. *J Innate Immun*. 2011;3:9-16. doi:10.1159/000321005
26. Jiang Y, Doolittle RF. The evolution of vertebrate blood coagulation as viewed from a comparison of puffer fish and sea squirt genomes. *Proc Natl Acad Sci USA*. 2003;100:7527-7532. doi:10.1073/pnas.0932632100
27. Davidson CJ, Hirt RP, Lal K, et al. Molecular evolution of the vertebrate blood coagulation network. *Thromb Haemost*. 2003;89:420-428.
28. Davidson CJ, Tuddenham EG, McVey JH. 450 million years of hemostasis. *J Thromb Haemost*. 2003;1:1487-1494. doi:10.1046/j.1538-7836.2003.00334.x
29. Kulman JD, Harris JE, Nakazawa N, Ogasawara M, Satake M, Davie EW. Vitamin K-dependent proteins in Ciona intestinalis, a basal chordate lacking a blood coagulation cascade. *Proc Natl Acad Sci USA*. 2006;103:15794-15799. doi:10.1073/pnas.0607543103
30. Ponczek MB, Bijak MZ, Nowak PZ. Evolution of thrombin and other hemostatic proteases by survey of protochordate, hemichordate, and echinoderm genomes. *J Mol Evol*. 2012;74:319-331. doi:10.1007/s00239-012-9509-0
31. Ponczek MB, Gailani D, Doolittle RF. Evolution of the contact phase of vertebrate blood coagulation. *J Thromb Haemost*. 2008;6:1876-1883. doi:10.1111/j.1538-7836.2008.03143.x
32. Ponczek MB, Shamanaev A, LaPlace A, et al. The evolution of factor XI and the kallikrein-kinin system. *Blood Adv*. 2020;4:6135-6147. doi:10.1182/bloodadvances.2020002456
33. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23:1073-1079.

34. Goodeve A, Reitsma P, McVey J, Working Group on Nomenclature of the Scientific and Standardisation Committee of the International Society on Thrombosis and Haemostasis. Nomenclature of genetic variants in hemostasis. *J Thromb Haemost*. 2011;9:852-855.
35. Petrovan RJ, Govers-Riemslag JW, Nowak G, Hemker HC, Tans G, Rosing J. Autocatalytic peptide bond cleavages in prothrombin and meizothrombin. *Biochemistry*. 1998;37:1185-1191. doi:10.1021/bi971948h
36. Gallwitz M, Enoksson M, Thorpe M, Hellman L. The extended cleavage specificity of human thrombin. *PloS One*. 2012;7:e31756.
37. Uliana F, Vizovišek M, Acquasaliente L, et al. Mapping specificity, cleavage entropy, allosteric changes and substrates of blood proteases in a high-throughput screen. *Nat Commun*. 2021;12:1-18.
38. Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, Craik CS. Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc Natl Acad Sci USA*. 2000;97:7754-7759.
39. Butenas S, Mann KG. Kinetics of human factor VII activation. *Biochemistry*. 1996;35:1904-1910. doi:10.1021/bi951768c
40. Seidah NG, Day R, Marcinkiewicz M, Chretien M. Precursor convertases: an evolutionary ancient, cell-specific, combinatorial mechanism yielding diverse bioactive peptides and proteins. *Ann N Y Acad Sci*. 1998;839:9-24. doi:10.1111/j.1749-6632.1998.tb10727.x
41. Conard J, Horellou MH, van Dreden P, et al. Homozygous protein C deficiency with late onset and recurrent coumarin-induced skin necrosis. *Lancet (London, England)*. 1992;339:743-744. doi:10.1016/0140-6736(92)90641-f
42. Reitsma PH, Bernardi F, Doig RG, et al. Protein C deficiency: a database of mutations, 1995 update. On behalf of the subcommittee on plasma coagulation inhibitors of the scientific and standardization committee of the ISTH. *Thromb Haemost*. 1995;73:876-889.
43. Rudolph AE, Mullane MP, Porche-Sorbet R, Daust HA, Miletich JP. The role of the factor X activation peptide: a deletion mutagenesis approach. *Thromb Haemost*. 2002;88:756-762.
44. Stojanovski BM, Pelc LA, Di Cera E. Role of the activation peptide in the mechanism of protein C activation. *Sci Rep*. 2020;10:11079. doi:10.1038/s41598-020-68078-z
45. Carnbring Bonde A, Rosenorn Hansen S, Johansson E, Rose Bjelke J, Lund J. Site-specific functional roles of the factor X activation peptide in the intrinsic tenase-mediated factor X activation. *FEBS Lett*. 2022;596:1567-1575. doi:10.1002/1873-3468.14321
46. Essalmani R, Susan-Resiga D, Guillemot J, et al. Thrombin activation of protein C requires prior processing by a liver proprotein convertase. *J Biol Chem*. 2017;292:13565. doi:10.1074/jbc.A116.77040
47. Foster DC, Sprecher CA, Holly RD, Gambee JE, Walker KM, Kumar AA. Endoproteolytic processing of the dibasic cleavage site in the human protein C precursor in transfected mammalian cells: effects of sequence alterations on efficiency of cleavage. *Biochemistry*. 1990;29:347-354. doi:10.1021/bi00454a007
48. Rezaie AR, Esmon CT. Tryptophans 231 and 234 in protein C report the ca(2+)-dependent conformational change required for activation by the thrombin-thrombomodulin complex. *Biochemistry*. 1995;34:12221-12226. doi:10.1021/bi00038a016
49. Hagen FS, Gray CL, O'Hara P, et al. Characterization of a cDNA coding for human factor VII. *Proc Natl Acad Sci USA*. 1986;83:2412-2416. doi:10.1073/pnas.83.8.2412
50. Stojanovski BM, Chen Z, Koester SK, Pelc LA, Di Cera E. Role of the I16-D194 ionic interaction in the trypsin fold. *Sci Rep*. 2019;9:18035. doi:10.1038/s41598-019-54564-6
51. Yousef GM, Elliott MB, Kopolovic AD, Serry E, Diamandis EP. Sequence and evolutionary analysis of the human trypsin subfamily of serine peptidases. *Biochim Biophys Acta*. 2004;1698:77-86.
52. Gohara DW, Di Cera E. Allostery in trypsin-like proteases suggests new therapeutic strategies. *Trends Biotechnol*. 2011;29:577-585. doi:10.1016/j.tibtech.2011.06.001
53. Pelc LA, Koester SK, Chen Z, Gistover NE, Di Cera E. Residues W215, E217 and E192 control the allosteric E*-E equilibrium of thrombin. *Sci Rep*. 2019;9:12304. doi:10.1038/s41598-019-48839-1
54. Vogt AD, Chakraborty P, Di Cera E. Kinetic dissection of the pre-existing conformational equilibrium in the trypsin fold. *J Biol Chem*. 2015;290:22435-22445. doi:10.1074/jbc.M115.675538
55. Gohara DW, Di Cera E. Molecular mechanisms of enzyme activation by monovalent cations. *J Biol Chem*. 2016;291:20840-20848. doi:10.1074/jbc.R116.737833
56. Pelc LA, Koester SK, Kukla CR, Chen Z, Di Cera E. The active site region plays a critical role in Na(+) binding to thrombin. *J Biol Chem*. 2022;298:101458. doi:10.1016/j.jbc.2021.101458
57. Rezaie AR, He X. Sodium binding site of factor Xa: role of sodium in the prothrombinase complex. *Biochemistry*. 2000;39:1817-1825. doi:10.1021/bi992006a
58. Schmidt AE, Stewart JE, Mathur A, Krishnaswamy S, Bajaj SP. Na+ site in blood coagulation factor IXa: effect on catalysis and factor VIIIa binding. *J Mol Biol*. 2005;350:78-91. doi:10.1016/j.jmb.2005.04.052
59. Underwood MC, Zhong D, Mathur A, Heyduk T, Bajaj SP. Thermodynamic linkage between the S1 site, the Na+ site, and the Ca2+ site in the protease domain of human coagulation factor Xa: studies on catalytic efficiency and inhibitor binding. *J Biol Chem*. 2000;275:36876-36884.
60. Di Cera E, Guinto ER, Vindigni A, et al. The Na+ binding site of thrombin. *J Biol Chem*. 1995;270:22089-22092. doi:10.1074/jbc.270.38.22089
61. Dang QD, Di Cera E. Residue 225 determines the Na(+)-induced allosteric regulation of catalytic activity in serine proteases. *Proc Natl Acad Sci USA*. 1996;93:10653-10656. doi:10.1073/pnas.93.20.10653
62. Guinto ER, Caccia S, Rose T, Futterer K, Waksman G, Di Cera E. Unexpected crucial role of residue 225 in serine proteases. *Proc Natl Acad Sci USA*. 1999;96:1852-1857. doi:10.1073/pnas.96.5.1852
63. Schmidt AE, Padmanabhan K, Underwood MC, Bode W, Mather T, Bajaj SP. Thermodynamic linkage between the S1 site, the Na+ site, and the Ca2+ site in the protease domain of human activated protein C (APC). Sodium ion in the APC crystal structure is coordinated to four carbonyl groups from two separate loops. *J Biol Chem*. 2002;277:28987-28995. doi:10.1074/jbc.M201892200
64. Wildgoose P, Foster D, Schioedt J, Wiberg FC, Birktoft JJ, Petersen LC. Identification of a calcium binding site in the protease domain of human blood coagulation factor VII: evidence for its role in factor VII-tissue factor interaction. *Biochemistry*. 1993;32:114-119.
65. Rezaie AR, Mather T, Sussman F, Esmon CT. Mutation of Glu-80–>Lys results in a protein C mutant that no longer requires Ca2+ for rapid activation by the thrombin-thrombomodulin complex. *J Biol Chem*. 1994;269:3151-3154.
66. Freguia CF, Toso R, Ferraresi P, Pinotti M, Bernardi F. Variation of factorVII 140s and 170s loops in fishes: evolutionary aspects and comparison with mutations found in FVII deficiency. *Thromb Haemost*. 2003;90:1220-1222.
67. Wouters MA, Rigoutsos I, Chu CK, Feng LL, Sparrow DB, Dunwoodie SL. Evolution of distinct EGF domains with specific functions. *Protein Sci*. 2005;14:1091-1103.
68. Stenflo J. Contributions of Gla and EGF-like domains to the function of vitamin K-dependent coagulation factors. *Crit Rev Eukaryot Gene Expr*. 1999;9:59-88.
69. Kelly CR, Dickinson CD, Ruf W. Ca2+ binding to the first epidermal growth factor module of coagulation factor VIIa is

important for cofactor interaction and proteolytic function. *J Biol Chem*. 1997;272:17467-17472.
70. Soriano-Garcia M, Padmanabhan K, De Vos A, Tulinsky A. The calcium ion and membrane binding structure of the Gla domain of calcium-prothrombin fragment 1. *Biochemistry*. 1992;31:2554-2566.
71. Stevens WK, Côté HC, MacGillivray RT, Nesheim ME. Calcium ion modulation of meizothrombin autolysis at Arg55-Asp56 and catalytic activity. *J Biol Chem*. 1996;271:8062-8067.
72. Huang M, Rigby AC, Morelli X, et al. Structural basis of membrane binding by Gla domains of vitamin K–dependent proteins. *Nat Struct Mol Biol*. 2003;10:751-756.
73. Yang L, Prasad S, Di Cera E, Rezaie AR. The conformation of the activation peptide of protein C is influenced by Ca2+ and Na+ binding. *J Biol Chem*. 2004;279:38519-38524. doi:10.1074/jbc.M407304200
74. Jagadeeswaran P, Sheehan JP. Analysis of blood coagulation in the zebrafish. *Blood Cells Mol Dis*. 1999;25:239-249.
75. Salte R, Norberg K, Ødegaard OR. Evidence of a protein C-like anticoagulant system in bony fish. *Thromb Res*. 1996;83:389-397.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Stojanovski BM, Di Cera E. Comparative sequence analysis of vitamin K-dependent coagulation factors. *J Thromb Haemost*. 2022;20:2837-2849. doi: 10.1111/jth.15897