EGF-like domains with a C-x-D-x(4)-Y-x-C motif

Abstract

The epidermal growth factor (EGF)-like domain is one of most common domains found in extracellular proteins and many proteins contain multiple EGF-like domains in tandem. Variations in the amino acid sequences of EGF-like domains make their study difficult and the characteristics of EGF-like domains are poorly understood. This mini-review introduces a putative subfamily of EGF-like domains with the consensus amino acid sequence C-x-D-x(4)-Y-x-C. This type of EGF-like domain in Del-1 and vitamin-K dependent coagulation factors has several features in common, such as increasing the efficiency of endocytosis-dependent transfection, attenuating cell adhesion, and inducing apoptosis. This motif is also found in other clinically important proteins such as Notches, Notch ligands, and low-density lipoprotein receptor. Finally, translational research on this motif is discussed.

Keywords: EGF, del-1, vitamin-K dependent coagulation factor, notch, LDLR

Introduction

The epidermal growth factor (EGF)-like domain was first identified in pro-epidermal growth factor in 1983 and since then many proteins with EGF-like domains have been reported to be involved in pathogenesis.1-3 Pro-EGF is significantly related with oncogenesis.4-6 Mutations of EGF-like domains in fibrillin-1, low-density lipoprotein receptor (LDLR), and coagulation factor IX (FIX) are the causes of Marfan syndrome, the autosomal dominant form of familial hypercholesterolemia (FH), and hemophilia B, respectively.7-9 EGF-like domains are found in numerous secreted extracellular proteins, except for the endoplasmic transmembrane protein prostaglandin G/H synthase.8 Extracellular proteins are easily accessed by drugs and thus are good therapeutic targets, making studies of EGF-like domains medically significant.

Currently, 2,662 EGF-like domains are identified in about 640 proteins in the SwissPro database.10 Most of these proteins are from animals but some are from plants and slim molds. Several proteins contain multiple EGF-like domains in tandem. For example, fibrillin-1 has 47 and Notch-1 has 36 EGF-like domains. EGF-like domains comprise a rich variety of amino acid sequences between conserved cysteine residues. The number and variation in EGF-like domain sequences complicates understanding the overall character of EGF-like domains.30 Dividing EGF-like domains into subfamilies would aid investigation of their functions.

This short manuscript reviews reports regarding a putative novel subfamily of EGF-like domains containing the amino acid sequence C-x-D-x(4)-Y-x-C (referred to as the DXXXXY motif in this review) and describes the characteristics of this motif. First, we describe the general properties of EGF-like domains, and then we review reports regarding the DXXXXY motif using information from the UniProdatabase. The DXXXXY motif shows a variety of functions in vitro and in vivo, and is found in several medically important proteins such as coagulation factors, low density lipoprotein receptor, and Notch related proteins. We conclude by discussing the significance of the DXXXXY motif in these proteins.

EGF-like domain

The EGF-like domain typically consists of 30-45 amino acids, including six cysteine residues that form three disulfide bonds. The domain has two β-sheets, with the major sheet at the N-terminus and the minor sheet at the C-terminus. The three disulfides are in the major sheet. In addition to the basic structure, distinct domain subtypes and post translational modifications have been identified.

Human EGF-like (hEGF) domain and compliment C1r (cEGF) domain

The two main types of EGF-like domains are the human EGF-like (hEGF) domain and the compliment C1r EGF-like (cEGF) domain. Wouters et al. discussed the differences between these two groups with regards to their amino acid sequences and biological characteristics based on their experimental data and database research.31 The two domains have clear differences in tertiary structure.12 Additionally, the 6th cysteine in hEGF is located in the β-turn of the minor sheet whereas the 6th cysteine in cEGF is located on the second strand of the sheet itself. There are 8-9 amino acids between the 5th and 6th cysteines in hEGF but more than 10 amino acids between these cysteines in cEGF. Some proteins contain only hEGFs (e.g., Notches), some proteins contain only cEGFs (e.g., LDLR), and some proteins contain both (e.g., fibrillins). The different functions of these two subtypes are poorly understood.

Calcium binding

The calcium binding EGF-like domain (cbEGF) consists of 45 amino acids and its consensus sequence is [DEQN]x-[DEQN](2)-C-x(3, 14)-C-x(3, 7)-C-x-[DN]-x(4)-[FY]-x-C.13 The affinity of this sequence for calcium is independent of differences between hEGF and cEGF. Calcium binding to cbEGF is critical for the functions of proteins. For example, fibrillin-1 contains 47 EGF domains, of
which 43 belong to cbEGF, and mutations affecting calcium binding cause Marfan syndrome. These mutations increase susceptibility to proteolysis and thus calcium binding may help maintain the native conformations of these proteins.14

**O-glycosylation**

O-glycosylation modifications of amino acid residues have been reported in hEGFs, including Notch, and the functions of these modifications have been well studied in Notch EGF-like domains.15–17 Typically, threonine and serine just before the 4th cysteine in EGF-like domains are modified with O-glucose. O-linked glycosylation by GALNT-11 is involved in determining left/right symmetry: glycosylation promotes the activation of Notch-1, possibly by promoting cleavage by ADAM17, modulating the balance between motile and immotile (sensory) cilia at the left-right organizer (LRO).14 O-glycosylations on EGF-like domains 11, 12 and 13 contain both O-linked fucose and O-linked glucose which interact with residues in DLL4.15

**β-hydroxylation**

Post-translational hydroxylation of aspartic acid or asparagine to form erythro-β-hydroxyaspartic acid or erythro-β-hydroxyasparagine has been identified in several EGF-like domains. The consensus amino acid sequence for β-hydroxylation is C-x-[DN]-x(4)-[FY]-x-C. There is a one-to-one correspondence between the hydroxylation of aspartic acid and hEGF, and of the hydroxylation of asparagine and eEGF.10 The significance of β-hydroxylation is poorly understood. Aspartic acid residues in the DXXXXY motifs of Notch are β-hydroxylated by aspartyl β-hydroxylase. Gene targeting of aspartyl β-hydroxylase for β-hydroxylation results in clef palate and the fusion of digits. Additionally, in the presence of APC (adenomatous polyposis coli) mutations, gene targeting of aspartyl β-hydroxylase increases the gene transfer efficiency.

**DXXXXY motif**

The consensus amino acid sequence for β-hydroxylation of EGF-like domains, C-x-[DN]-x(4)-[FY]-x-C, indicates four ways, DXXXXF, DXXXXY, NXXXXX, and NXXXXY. One sequence, C-x-D-x(4)-Y-x-C, may form a putative subfamily. The EGF-like domains from Del-1 and FIX and increase the gene transfer efficiency of non-viral vectors via endocytosis in vitro and in vivo.21 The EGF-like domains in Del-1 and FIX are hEGF and cbEGF types with the DXXXXY motif. The mutation of aspartic acid to asparagine or tyrosine to phenylalanine decreases the activities of Del-1. Other proteins containing the DXXXXY motif are discussed below.

**Del-1**

Del-1 is an extracellular matrix (ECM) protein expressed by embryonic endothelial cells.22 Over-expression of Del-1 enhances the remodeling of blood vessels in vasculogenesis. In adults, Del-1 regulates leucocyte adhesion to pulmonary endothelial cells in inflamed tissues.24 Del-1 contains three EGF-like domains in the N-terminus and two discoidin domains in the C-terminus. The EGF-like domains in Del-1 are all hEGF. The second EGF-like domain contains an RGD integrin-binding motif and the third EGF-like domain has a DXXXXY motif. Discoidin domains have been reported to bind to anionic phospholipids on the surface of cells.25

A recombinant version of the 3rd EGF-like domain of Del-1 induces endocytosis, pinocytosis, and phagocytosis, attenuates cell adhesion to ECM, and induces caspase-dependent apoptosis in vivo.20–26 The EGF-like domain with a DXXXXY motif serves as a trans-element for these functions. Del-1 accumulates in ECM, and a domain for adhesion to ECM has been localized to the 1st discoidin domain.20 The adhesion activity of the 1st discoidin domain is significantly enhanced by the adjacent 3rd EGF-like domain. The EGF-like domain with a DXXXXY motif serves as a cis-element for this function. Based on the finding that the 3rd EGF-like domain and the 1st discoidin domain accumulate in ECM, induce apoptosis, and improve transfection efficiency, Kitano et al. treated tumor explant model mice with cDNA of the 3rd EGF-like domain and the 1st discoidin domain using a non-viral vector.30,31 The transgene product accumulated in ECM and induced apoptosis in tumors, indicating that such gene therapy can significantly suppresses tumor growth and improve life prognosis.

**Coagulation factors VII, IX, and X**

Coagulation factors VII (FVII), IX (FIX) and X (FX) are highly homologous vitamin-K dependent enzymes that are essential for coagulation.32 They have a γ-carboxyglutamic acid (Gla) domain at the N-terminus that is essential for binding to phosphatidyserine in the cell membrane. Two EGF-like domains are adjacent to the Gla domain. The 1st EGF-like domain is a hEGF and cbEGF domain with the DXXXXY motif and the 2nd is a cbEGF but not a cbEGF domain.13 A peptidase domain that cleaves other coagulation factors is adjacent to the 2nd EGF-like domain. Mutation of FIX causes hemophilia B and mutation of the 1st EGF-like domain disrupts calcium binding and may cause hemophilia B.33 In the absence of calcium, the 1st EGF-like domain in FIX adopts a conformation that locks the Gla domain in a fixed position, restricting movement of the Gla and cbEGF domains.34,36 Calcium binding to cbEGF induces a conformational change that may change protein function, such as enzymatic activity and binding to other molecules. For this function, the EGF-like domain with a DXXXXY motif serves as a cis-element.

A recombinant version of the 1st EGF-like domain of FIX suppresses cell adhesion to ECM, increases cell migration, and activates caspase-3.37 The 1st EGF-like domains of FVII and FX share many functions. The EGF-like domain with a DXXXXY motif serves as a trans-element for these functions. Additionally, the recombinant 1st EGF-like domain of FIX induces the exposure of phosphatidyserine and the clustering of lipid rafts in the cell membrane that likely contribute to endocytosis.38 Because coagulation factors are generally activated in wounds, these functions of the 1st EGF-like domains of coagulation factors may contribute to wound healing and immunological reactions.

There are three other vitamin-K dependent coagulation-related proteins: protein C, protein S, and protein Z.32 They also consist of a Gla domain, an EGF-like domain, and a peptidase domain, similar to FVII, FIX and FX. However, they negatively regulate coagulation. Interestingly, protein C and protein S have an NXXXXY sequence instead of DXXXXY. Protein Z has a DXXXXY motif but is not a cbEGF.

**Other proteins with the DXXXXY motif**

---

**Citation:** Hidai C. EGF-like domains with a C-x-D-x(4)-Y-x-C motif. *Open Access J Trans Med Res.* 2018;2(2):67–71. DOI: 10.15406/oajtmr.2018.02.00039
Other than Del-1 and vitamin-K dependent coagulation factors, there are several other proteins with the DXXXXY motif in EGF-like domains (Table 1). There are 31 proteins with the DXXXXY motif in the UniPro database and they can be categorized into 5 groups: ECM-associated proteins, vitamin-K dependent coagulation factors, Notch-related proteins, Lipoprotein transport-related proteins, and other membrane proteins. All contain between one to 12 DXXXXY motifs. Their associated domains are often involved in binding other molecules. The four amino acids residues between the aspartic acid and tyrosine in DXXXXY share several characteristics (Table 2). Many of the residues are aliphatic, non-polar, neutral amino acids and thus the DXXXXY motif should be hydrophobic.  

### Table 1 List of proteins with the DXXXXY motif in EGF-like domains

| #   | UniProtKB # | ECM associated proteins                                      | No of DXXXXY motifs | Associated domains           |
|-----|-------------|--------------------------------------------------------------|---------------------|------------------------------|
| 1   | O43854      | Dell                                                         | 1                   | Discoidin                    |
| 2   | O75093      | Slit homolog 1 protein                                       | 1                   |                               |
| 3   | O94813      | Slit homolog 2 protein                                       | 1                   | Laminin, LRR                 |
| 4   | O75094      | Slit homolog 3 protein                                       | 1                   | Laminin, LRR                 |
| 5   | P35556      | Fibillin-2                                                  | 1                   | TB                           |
| 6   | Q75N90      | Fibillin-3                                                  | 1                   | TB                           |
| 7   | Q8TER0      | Sushi, nidogen and EGF-like domain-containing protein        | 3                   | Sushi, Follistatin           |
| 8   | Q4LDE5      | Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein | 2 | Sushi |

### Vitamin K dependent coagulation factors

| #   | UniProtKB # | ECM associated proteins                                      | No of DXXXXY motifs | Associated domains           |
|-----|-------------|--------------------------------------------------------------|---------------------|------------------------------|
| 9   | P08709      | Coagulation factor V1I                                        | 1                   | Gla                          |
| 10  | P00742      | Coagulation factor X                                         | 1                   | Gla                          |
| 11  | P00740      | Coagulation factor IX                                        | 1                   | Gla                          |
| 12  | P22891      | Vitamin K-dependent protein Z                                 | 1                   | Gla                          |

### Notch related proteins

| #   | UniProtKB # | ECM associated proteins                                      | No of DXXXXY motifs | Associated domains           |
|-----|-------------|--------------------------------------------------------------|---------------------|------------------------------|
| 13  | P46331      | Notch 1                                                     | 12                  | Ank                          |
| 14  | Q04721      | Notch 2                                                     | 10                  | Ank                          |
| 15  | Q9UM47      | Notch 3                                                     | 3                   | Ank                          |
| 16  | Q99466      | Notch 4                                                     | 2                   | Ank                          |
| 17  | Q00548      | Delta-like protein 1                                         | 1                   | DSL                          |
| 18  | Q9NY7       | Delta-like protein 3                                         | 1                   | DSL                          |
| 19  | Q9NR61      | Delta-like protein 4                                         | 1                   | DSL                          |
| 20  | P78504      | Protein Jagged-1                                            | 3                   | DSL                          |
| 21  | Q9Y219      | Protein Jagged-2                                            | 1                   | DSL                          |
| 22  | Q8NFT8      | Delta and Notch-like epidermal growth factor related receptor | 2 | Follistatin |

### Lipoprotein transport related proteins

| #   | UniProtKB # | ECM associated proteins                                      | No of DXXXXY motifs | Associated domains           |
|-----|-------------|--------------------------------------------------------------|---------------------|------------------------------|
| 23  | P01130      | Low-density lipoprotein receptor                              | 1                   | LDLR-I                       |
| 24  | P98155      | Very low-density lipoprotein receptor                        | 1                   | LDLR-I                       |
| 25  | Q9NZ2R2     | Low-density lipoprotein receptor -related protein 1B         | 1                   | LDLR-I                       |
| 26  | O60494      | Cubilin                                                     | 1                   | CUB                          |

### Other transmembrane proteins

| #   | UniProtKB # | ECM associated proteins                                      | No of DXXXXY motifs | Associated domains           |
|-----|-------------|--------------------------------------------------------------|---------------------|------------------------------|
| 27  | P82279      | Protein crumbs homolog 1                                      | 4                   | Laminin G                    |
| 28  | Q51J48      | Protein crumbs homolog 2                                      | 5                   | Laminin G                    |
| 29  | Q9UBL1      | Neurexin-1-alpha                                            | 1                   | Laminin G                    |
| 30  | Q9NYQ6      | Caderhin EGF LAG seven-pass G-type receptor 1               | 1                   | Laminin G                    |
| 31  | Q727M0      | Multiple epidermal growth factor-like domains protein B      | 1                   | CUB, PSI                      |

### Table 2 Frequency of amino acids residues at the four positions between D and Y

Citation: Hidai C. EGF-like domains with a C-x-D-x(Y)-x-C motif. Open Access J Trans Med Res. 2018;2(2):67–71. DOI: 10.15406/oajtmr.2018.02.00039
EGF-like domains with a C-x-D-(4)-Y-x-C motif

The three most frequent amino acids are presented: AA, amino acid.

Notch-1 contains 36 EGF-like domains. Human and Xenopus Notch-1 have 12 DXXXXY motifs and mouse, zebrafish, and drosophila Notch-1 have 11 DXXXXY motifs mainly in a middle section of their extracellular domain. All EGF-like domains with the DXXXXY motif are hEGF, and 11 of 12 DXXXXY motifs are cbEGF in human Notch-1. Several proteins that are Notch ligands also have DXXXXY motifs and both Notchs, and their ligand proteins play critical roles in development, cancer, and angiogenesis. It is well documented that EGF-like domains 11-12 are necessary and sufficient for ligand binding.\(^\text{15,18,40}\) Additionally, a study in Drosophila suggests that EGF-like domains 24-29 are important in ligand-dependent receptor activation.\(^\text{41}\) However, the roles of the DXXXXX motifs in Notch-related proteins are poorly understood.

LDLR is essential for the uptake of LDL by hepatocytes via endocytosis. Mutations of LDLR cause the autosomal dominant form of familial hypercholesterolemia (FH) and autosomal recessive hypercholesterolemia (ARH).\(^\text{1,3-5}\) LDLR has an N-terminus LDL binding domain followed by three EGF-like domains that are all cbEGF. The 1st EGF-like domain has the amino acid sequence GTNECLDNNGGCSHVCNDLKIGYECLC, an atypical cbEGF sequence, and contains a DXXXXY motif. The 1st EGF-like domain of LDLR plays an important role in pathogenesis of FH. Wild type LDLR is recycled and spared from degradation after endocytosis.\(^\text{43}\) A mutant LDLR protein lacking the EGF-like domains is degraded in lysosomes after endocytosis, thereby resulting in a decrease in the levels of LDLR. The decrease of LDLR decreases the uptake of LDL by hepatocytes and increases the LDL concentration in serum. Proprotein convertase subtilisin/kexin type 9 (PCSK9) was originally recognized as a key molecule in the regulation of serum LDL levels.\(^\text{44-46}\) PCSK9 is a member of the superfamily of cysteine proteases (EGF-like proteases), which includes the proprotein convertases (PCs) and furin.\(^\text{47}\) The three most frequent amino acids are presented. AA, amino acid.

The authors declare no conflict of interest.

References

1. Gray A, Dull TJ, Ulrich A. Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor. Nature. 1983;303(5919):722–725.
2. Scott J, Urdea M, Quiroga M, et al. Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. Science. 1983;221(4607):236–240.
3. Esparis-Ogando A, Montero JC, Arribas J, et al. Targeting the EGF/HER Ligand-Receptor System in Cancer. Curr Pharm Des. 2016;22(39):5887–5898.
4. Lindsey S, Langhans SA. Epidermal growth factor signaling in transformed cells. Int Rev Cell Mol Biol. 2015;314:1–41.
5. Abifadel M, Rabes JP, Jambart S, et al. The molecular basis of familial hypercholesterolemia in Lebanon: spectrum of LDLR mutations and role of PCSK9 as a modifier gene. Hum Mutat. 2009;30(7):E682–691.
6. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutat. 1992;1(6):445–466.
7. Zeyer KA, Reinhardt DP. Engineered mutations in fibrillin-1 leading to Marfan syndrome act on the protein, cellular and organismal levels. Mutat Res Rev Mutat Res. 2015;765:7–18.
8. Lillicrap D. FIX It in One Go: Enhanced Factor IX Gene Therapy for Hemophilia B. Cell. 2017;171(7):1478–1480.
9. Otto JC, Dewitt DL, Smith WL. N-glycosylation of prostaglandin endoperoxide synthases-1 and -2 and their orientations in the endoplasmic reticulum. J Biol Chem. 1993;268(24):18234–18242.
10. Wouters MA, Rigoutsos I, Chu CK, et al. Evolution of distinct EGF domains with specific functions. Protein Sci. 2005;14(4):1091–1103.
11. Rand MD, Lindblom A, Carlson J, et al. Calcium binding to tandem repeats of EGF–like modules. Expression and characterization of the EGF–like modules of human Notch–1 implicated in receptor–ligand interactions. Protein Sci. 1997;6(10):2059–2071.
12. Bensch B, Hernandez JF, Marion D, et al. Solution structure of the epidermal growth factor (EGF)–like module of human complement protease C1r, an atypical member of the EGF family. Biochemistry. 1998;37(5):1204–1214.
13. Handford PA, Baron M, Mayhew M, et al. The first EGF–like domain from human factor IX contains a high–affinity calcium binding site. J Biol Chem. 1993;268(24):18234–18242.
EGF-like domains with a C-x-D-x(4)-Y-x-C motif

Lichtman MA, E Beutler, TJ Kipps, et al. Williams Hematology. 7th ed. McGrawHill: New York, USA, 2006;30(3):1856.

Stenflo J, Stenberg Y, Muranyi A. Calcium–binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions. Biochim Biophys Acta. 2000;1477(1–2):51–63.

Vadivel K, Bajaj SP. Structural biology of factor VIIa/tissue factor initiated coagulation. Front Biosci (Landmark Ed). 2012;17:2476–2494.

Sunnerhagen M, Olah GA, Stenflo J, et al. The relative orientation of Gla and EGF domains in coagulation factor X is altered by Ca2+ binding to the first EGF domain. A combined NMR–small angle X-ray scattering study. Biochemistry. 1996;35(36):11547–11559.

Sunnerhagen M, Forsen S, Hoffren AM, et al. Structure of the Car2(+)–free Gla domain sheds light on membrane binding of blood coagulation proteinases. Nat Struct Biol. 1995;2(6):504–509.

Ishikawa T, Kitano H, Mamiya A, et al. The first EGF domain of coagulation factor IX attenuates cell adhesion and induces apoptosis. Biochimie. 2016;3(3).

Hidai C, Fujisawa Y, Kokubun S, et al. EGF domain of coagulation factor IX is conducive to exposure of phosphatidylserine. Cell Biol Int. 2017;41(4):374–383.

Kyte J, Dolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–132.

Luca VC, Kim BC, Ge C, et al. Notch–Jagged complex structure implicates a catch bond in tuning ligand sensitivity. Science. 2017;355(6331):1320–1324.

Sharma A, Rangarajan A, Dighe RR. Antibodies against the extracellular domain of human Notch1 receptor reveal the critical role of epidermal–growth–factor–like repeats 25–26 in ligand binding and receptor activation. Biochem J. 2013;449(2):519–530.

Dvir H, Shah M, Girardi E, et al. Atomic structure of the autosomal recessive hypercholesterolemia phosphotyrosine–binding domain in complex with the LDL–receptor tail. Proc Natl Acad Sci USA. 2012;109(18):6916–6921.

Van Der Westhuizen DR, Stein ML, Henderson HE, et al. Deletion of two growth–factor repeats from the low–density–lipoprotein receptor accelerates its degradation. Biochem J. 1991;277(Pt 3):677–682.

Lagace TA, Curtis DE, Garuti R, et al. Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice. J Clin Invest. 2006;116(1):2995–3005.

Bingham B, Shen R, Kotnis S, et al. Proapoptotic effects of NARC 1 (=PCSK9), the gene encoding a novel serine proteinase. Cytometry A. 2006;69(11):1123–1131.

Kwon HJ, Lagace TA, Mcnutt MC, et al. Molecular basis for LDL receptor recognition by PCSK9. Proc Natl Acad Sci USA. 2008;105(6):1820–1825.

Wicinski M, Zak J, Malinowski B, et al. PCSK9 signaling pathways and their potential importance in clinical practice. EPMA J. 2017;8(4):391–402.

Aquila G, Pannella M, Morelli MB, et al. The role of Notch pathway in cardiovascular diseases. Glob Cardiol Sci Pract. 2013;2013(4):364–371.

Solanki A, Bhatt LK, Johnston TP. Evolving targets for the treatment of atherosclerosis. Pharmacol Ther. 2018;S0163-7258(18):30021–30024.