Epidermal growth factor-like modules

Iain D. Campbell and Peer Bork

University of Oxford, Oxford, UK and European Molecular Biology Laboratory, Heidelberg, Germany

Numerous copies of domains, or modules, with similarity recognizable from sequence analysis, are being found in a wide variety of different proteins. New information on the structure and distribution of modules in the epidermal growth factor superfamily is presented.

Current Opinions in Structural Biology 1993, **3**:385–392

Introduction

The idea that many proteins are constructed from various modules, each identifiable by a consensus sequence that is sometimes related to the exon structure of the gene, is becoming increasingly familiar [1–3]. The epidermal growth factor (EGF)-like module was one of the earliest examples of this. It was first characterized and sequenced in 1972 within EGF itself (see [4] and references therein). About 10 years later, with the sequencing of its precursor and the discovery of sequence similarities in regions of some blood-clotting factors, it became apparent that the module is widely distributed [5–7]. It is now known to be present in a wide variety of proteins including those associated with blood coagulation, fibrinolysis, neural development and cell adhesion (see Fig. 1).

Despite the relatively long history of this small module, which usually comprises ~45 residues and three disulphide bridges, considerable interest continues to surround its biological role, which mainly seems to be the mediation of protein–protein interactions. The past year has yielded much new information in several areas: the module's distribution in biological systems, from analysis

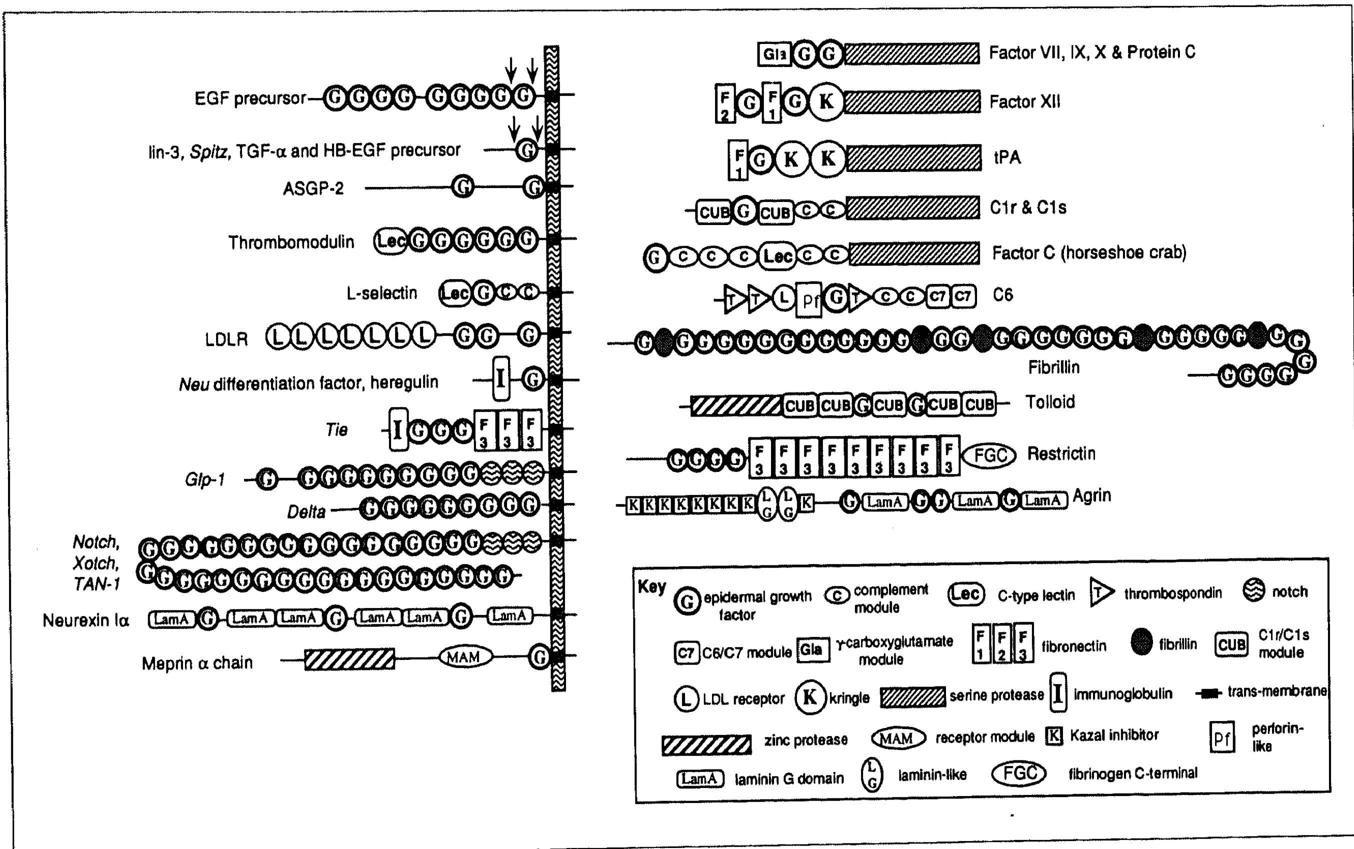

Fig. 1. An illustration of some of the diverse mosaic proteins that have been found to contain EGF modules, including most of those mentioned in the text. Signal sequences and cytoplasmic carboxy-terminal domains are not represented in this diagram. The arrows indicate possible cleavage sites for release of growth factors. The nomenclature of the modules is arbitrary. The many different abbreviations used by different authors represent a challenge for agreement on a standard nomenclature.

Abbreviations

3D—three-dimensional; **EGF**—epidermal growth factor; **HB**—heparin-binding; **TGF**—tissue growth factor.

© Current Biology Ltd ISSN 0959-440X

of the growing sequence database; its three-dimensional (3D) structure, from NMR and X-ray studies; and its various functional roles, by analysis of the effects of gene mutations. The aim of this brief review is to discuss these new findings; the reader is referred to several recent reviews for discussion of some of the better known features of EGF modules [8–12].

Some new EGF modules

As illustrated in Fig. 1, the EGF module is found in many different kinds of mosaic protein in association with many different kinds of protein module. During the review period, several new mosaic proteins containing the EGF module have been sequenced. It has recently been discovered, for the first time, in association with an immunoglobulin superfamily module in the *neu* differentiation factor [13•] and with both immunoglobulin and fibronectin type III modules in an endothelial cell-surface tyrosine kinase receptor product of the *tie* gene [14•]. Several new examples of EGF-like molecules acting as growth factors have also been found recently, including *lin-3*, which encodes an inductive signal for vulval development in *Caenorhabditis elegans* [15], the *spitz* gene, which participates in axis formation and neurogenesis of *Drosophila* [16], and the heparin-binding (HB) EGF-like growth factor [17]. Many EGF-like growth factors have a membrane-bound precursor; in the case of HB growth factor, it also acts as a receptor for diphtheria toxin (see Fig. 1) [18]. Three other new receptor-like molecules contain an EGF module which is located at the carboxyl terminus of the extracellular part adjacent to the transmembrane region: meprin, a membrane-bound oligomeric metalloendopeptidase [19•]; ASGP2, the membrane-associated component of a cell-surface sialomucin complex observed in rat mammary adenocarcinoma [20•]; and heregulin, which acts as a specific activator of the proto-oncogenic tyrosine kinase $p185^{erbB2}$ [21•]. Analysis of the latter's sequence revealed an amino-terminal V-type immunoglobulin-like domain (Fig. 1). A secreted molecule, *gil* (giant lens), with a carboxy-terminal EGF module has been recently shown to be involved in cell determination and axon pathfinding in the visual system of *Drosophila* [22•].

EGF modules often occur in multiple copies, particularly in matrix proteins. An example of this is provided by restrictin, a large multifunctional, tenascin-like molecule which is implicated in neural cell attachment [23]. Whereas restrictin and other matrix proteins contain consecutive copies of the EGF module, the architecture of the synaptic neurexin family [24•] suggests the duplication of larger units composed of one EGF and two so-called G domains, which were first identified in laminin A (Fig. 1).

Comparative sequence analysis of EGF modules

The examples given above document the steadily increasing number of EGF-like modules in functionally diverse proteins. One problem is that the more members of a sequence family that are known, the more difficult it becomes to separate this family from non-related proteins. The six cysteines characteristic of the EGF-like module are no longer sufficient for unique identification in sequence database searches. During the past few years, it has become apparent that, among identified EGF-like modules, the number of amino acids between the six cysteines can vary considerably; not a single region between successive cysteines is invariant in length. Based on a comprehensive comparison of many EGF-like modules (Fig. 2) and some structural constraints (see below), we propose the following consensus sequence:

$$\mathrm{xxxx}\mathbf{C}\mathrm{x}_{2-7}\mathbf{C}\mathrm{x}_{1-4}\mathrm{(G/A)x}\mathbf{C}\mathrm{x}_{1-13}\mathrm{ttax}\mathbf{C}\mathrm{x}\mathbf{C}\mathrm{xxGax}_{1-6}\mathrm{Gxx}\mathbf{C}\mathrm{x}$$

using the one-letter amino acids code and where a denotes aromatic, t non-hydrophobic, and x any amino acid. Large deviations can occur nevertheless, including the insertion of as many as 20 amino acids. Using a 'fuzzy' description of this consensus and allowing for such deviations, more than 600 versions of the EGF module can be detected in current sequence databases. Taking into account redundancy (orthologous genes in different species, different splicing variants, and other highly similar domains) still leaves more than 300 different EGF-like modules in about 70 distinct proteins. This is a sizeable number considering that not even 3000 different animal sequences (species redundancies aside) have been sequenced so far [25]. The EGF modules has been detected mostly in animal proteins, but it is also present in several virus proteins. In addition, a few EGF modules have been reported in protozoa such as *Plasmodium palcifarum*, and one recent report describes similarities to the plant toxin purothionin [26]. One should be cautious when interpreting the latter similarities, as one or two of the cysteines are missing in the respective groups of plant toxins. Most EGF modules have been found in extracellular proteins, but, surprisingly, they have also been described in several intracellular peroxidases and cyclooxidases (for a recent summary, see [27]).

The average pairwise sequence identity between EGF modules is just over 30%. A dendrogram based on sequence similarities (Fig. 2) reveals some information about the evolution of the module. For example, EGF modules of proteins with similar modular architecture (Fig. 1) tend to cluster together. This could reflect functional analogies but might also indicate recent gene-duplication events. The grouping of multiple copies belonging to a single protein has also been observed for other modules, and suggests a preference for further duplications if a module already exists in several consecutive copies. Even if the dendrogram shown in Fig. 2 is a fairly simple representation of the relationships among EGF-like modules, it does indicate the sequences that are highly diver-

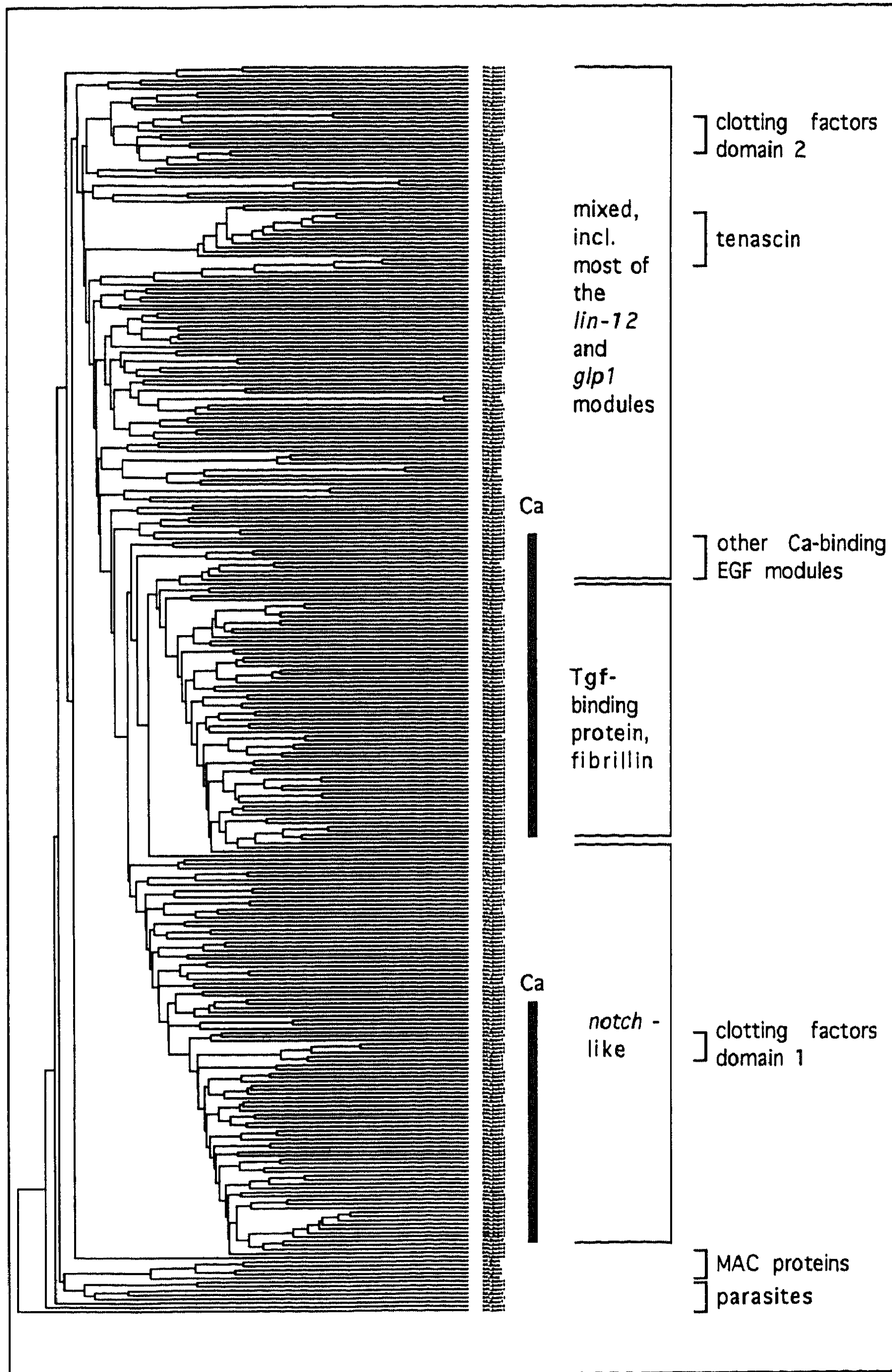

Fig. 2. Dendrogram of 302 EGF modules. To reduce redundancy, only those EGF modules with a pairwise sequence similarity of less than 80% have been considered. As the pairwise sequence identities are rather low, the dendrogram is not very stable and can only be considered as a rough indication of evolutionary relationships. Even so, EGF modules with the consensus sequence for calcium binding (see text) cluster together in two large groups (thick vertical lines). Some interesting subgroups are labelled at the right, e.g. the divergence of the two EGF modules in several clotting factors. For some highly divergent sequences like those in parasites and even some of the MAC (membrane attachment complex) proteins, the classification of membership in the EGF family should be reexamined.

gent and which might not be typical EGF modules. For example, all so-called EGF-like repeats in laminins appear to be outliers (data not shown) and the sequence similarities to EGF might not imply a similar 3D structure. The repeats in laminins are longer than the average EGF module and contain two additional cysteines. In a database search we have identified only one other protein containing the same repeats, the basal lamina component agrin [28,29]. Interestingly, agrin contains four copies of a typical EGF module containing six cysteines (only these have been identified in the database search) and two laminin-like repeats containing eight cysteines (Fig. 1) [28,29]. From the known structures of EGF modules, it can be concluded that the introduction of an additional disulfide bridge (as observed in the laminin-like repeats) is likely to introduce significant changes in topology. A final decision about family membership requires the 3D structure of this laminin module and superposition on known EGF structures.

Three-dimensional structures

Good structural information about the EGF module has been surprisingly difficult to obtain. Until this year, no crystals suitable for diffraction studies had been produced and many of the early NMR structures were of rather poor quality. New structural information from NMR studies of EGF, as well as both NMR and X-ray studies of EGF modules from the blood clotting factors IX and X have, however, been obtained recently, giving new structural insight into the module. Some of the structural information obtained from these studies is

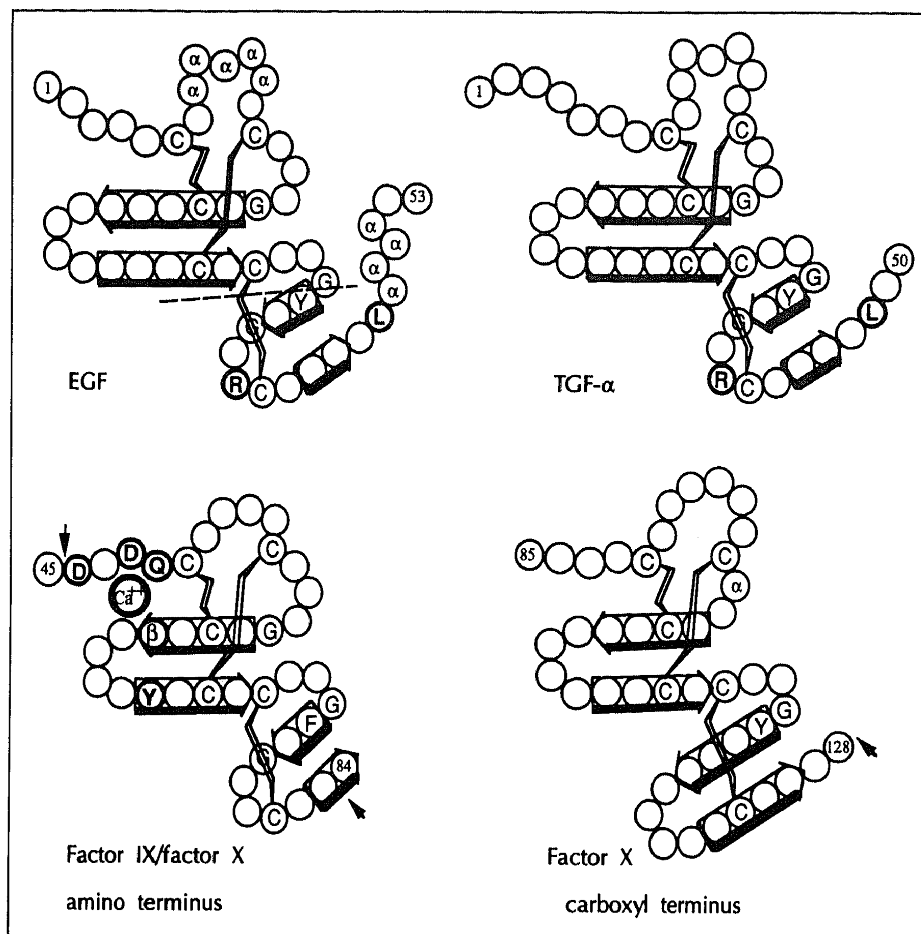

Fig. 3. A schematic representation of structural information now available about the EGF family, as gleaned from some recent structural work (K Padmanabhan *et al.*, personal communcation) [30••,31••,32•,33,34,35•,36•,37••]. A better representation of the 3D structures is obtained by 'folding' approximately along the dotted line. The approximate extents of the β-sheets are shown and amino acids found to adopt a helical conformation in some structures are indicated by an α. The arrows indicate the exon structure of the gene in factor X and the amino acid numbering shown is for factor X; the numbering for factor IX differs by 1. Some residues which have been clearly identified as important for function are indicated in bold.

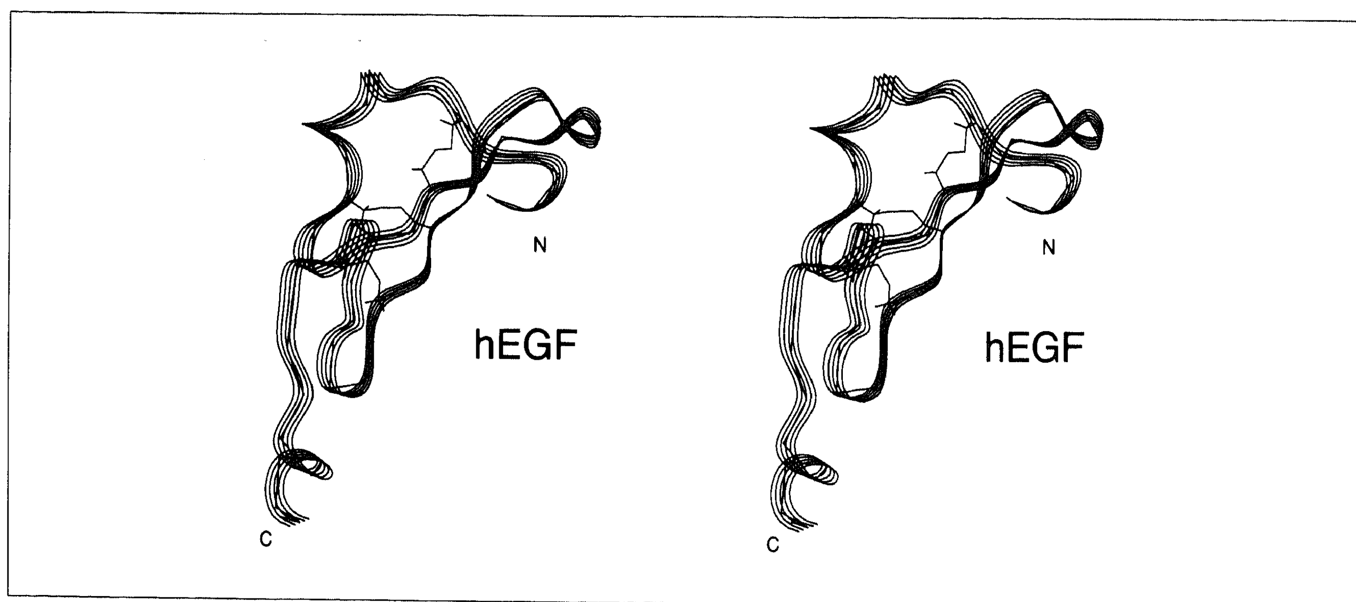

Fig. 4. Stereo ribbon representation of human EGF [31••]. The C_{α} trace as well as the side chains of the six cysteines involved in disulfide bridges are also displayed.

illustrated schematically in Fig. 3 and as a stereo diagram in Fig. 4.

Two relatively high-resolution NMR structures, one of mouse [30••], the other of human EGF [31••], have appeared. Both these structures were derived from more than 13 restraints per residue, which is a significant improvement over previously published structures. The information deduced from these studies can be summarized with reference to Fig. 3. The main structural feature is a two-stranded β-sheet. Three disulphide bridges radiate from one surface of this sheet to connect to a relatively loose amino-terminal strand and loop and to a carboxy-terminal domain which contains another short two-stranded sheet. An important feature of the 3D structure, not illustrated in the diagram in Fig. 3, is the close contact between the loop immediately after the second cysteine and the loop just before the sixth cysteine. This contact is defined by the observation of numerous inter-loop nuclear Overhauser effects; for example,

human EGF possesses an H-bond between Leu15 and Arg41. EGF possesses no very clear hydrophobic core, and seems to possess considerable conformational flexibility, especially at the amino and carboxyl termini. Flexibility is also suggested by the results of a lower-resolution NMR study of mouse EGF, which indicated the existence of slightly different conformations at different pH values [32•]. In the structural study of human EGF [31••], a tendency for α-helix formation was observed both in the carboxy-terminal tail, which contains two tryptophans, and in the loop between the first two cysteine residues (see Fig. 3). This study also compared the observed structure with one previously obtained for tissue growth factor (TGF)-α [33]; the overall structure and the structure of a surface which might bind the EGF receptor were found to be very similar in the two molecules (see also below).

The blood clotting factors IX and X each contain two EGF modules (Fig. 1). The first has, in addition to the usual EGF consensus sequence, a consensus sequence D/N-D/N-D*/N*-Y/F, which leads to β-hydroxylation of the residue marked * (see Fig. 3). Extensive NMR studies of the first EGF module from both factors IX and X have been carried out recently using material produced by proteolytic fragmentation, peptide synthesis and a yeast expression system. For factor IX, early predictions that the module would bind calcium were confirmed by monitoring calcium binding by NMR; key residues involved in the binding site were identified by mutating the specific residues D, D, Q and β marked in Fig. 3 (the post-translational modification at the β site is also discussed below) [34]. The structure of the factor IX module was shown to be similar to the fold already observed in EGF [35•]. The same module from factor X has been investigated both with and without bound calcium [36•,37••]. Although bound calcium could not be observed directly by NMR, it was concluded that the calcium ligands were the side chains of Gln49, β63 and possibly Asp46 together with the backbone carbonyls of Gly47 and Gly64. Because the structures of the apo and calcium-bound forms are very similar, the effect of calcium binding must be strictly local. Similar results have been observed with the calcium-bound factor IX module (M Robin *et al.*, unpublished data).

Numerous attempts to crystallize EGF and EGF modules have largely been unsuccessful but a factor X molecule without the Gla domain (the γ-carboxyglutamate-containing domain; see Fig. 1) was crystallized recently and its structure resolved to 2.5 Å (K Padmanabhan *et al.*, personal communication). No electron density was seen from the first EGF module, presumably because it is disordered. The second EGF module was resolved. The loops in this module, especially in the carboxy-terminal domain, are significantly different in size to those in those EGF modules whose structures have been determined by NMR. The structure is nonetheless again very similar, as illustrated in Fig 3. Close contact between the amino- and carboxy-terminal domains is again observed and is typified by the observation of an H-bond between Asp97 and Cys124. An interesting feature of the X-ray structure is that the second EGF module appears to make close contact with the catalytic serine-protease domain.

Functional role of EGF modules

There has been a long-term interest in defining the surface of EGF that might be involved in binding to the EGF receptor, the goal being the development of small-molecule agonists or antagonists. So far, the goal has been elusive. There is general agreement that Arg41 and Leu47 in EGF are essential for binding, yet these two residues are at least 17 Å apart in the NMR structure [31••]. Efforts to define the surface by site-directed mutagenesis continue. Koide *et al.* [38] have shown that mutations in the major β-sheet (at Ala30 and Asn32) cause a significant decrease in binding. In another study, using an EGF/endotoxin A chimeric protein, Shiah *et al.* [39] suggested that positions 19 and 34 were also important to receptor binding. Thus it appears that quite a large surface of EGF might be involved in binding to the receptor. One possibility is the formation of a 2:1 receptor:EGF complex in which different parts of EGF binding to the two receptor molecules. Richter *et al.* [40] have proposed that TGF-α might undergo a relatively large conformational change on binding to the receptor. As the overall EGF structures are well conserved even with rather different sequences, and as the interdomain interface is well defined in solution, this kind of rearrangement seems unlikely.

Now that a considerable amount of information is available about the first EGF module of factors IX and X, it is fruitful to ask questions about their role in the formation of a blood clotting complex which involves factors VIII, IX and X as well as calcium and phospholipid. There is a strong correlation between the ability of various mutant single EGF modules to bind calcium and the clotting activity of intact factor IX with the same mutations. For example, in the 'Alabama' mutation in factor IX, where Asp47 is changed to glycine in a haemophilia B patient, the intact molecule shows reduced clotting activity and the K_d for calcium binding to the isolated mutant module is lowered by about 30-fold [41]. In the study of calcium bound to the EGF module from factor X, it was pointed out that the calcium is not buried but is exposed to the solvent on one side [37••]. It is probable that either the neighbouring Gla domain or another protein contributes other calcium ligands. Many of the EGF-module family have sequence similarity to the calcium-binding modules from factors X and IX and are found in a wide range of proteins (see, for example, [37••] for a recent review). It seems certain that calcium is important for providing the correct protein–protein interface in such modules.

An interesting feature of EGF modules in various clotting and fibrinolysis proteins is their post-translational modification. For example, factors VII, IX, X, protein Z and tPA have been all found to have an O-linked sugar attached to serine or threonine at the position immediately preceding the second cysteine [42]. Unusual O-linked sugars have also been observed at a similar position in one of the three EGF modules in thrombospondin [43]. As mentioned above, asparagine or aspartate residues are known to become hydroxylated at the β position in EGF modules with a consensus tyrosine or phenylalanine in the

main β-sheet (see, for example, [9] and the factor X amino-terminal EGF module shown in Fig. 3). It has recently been observed that substitution of the tyrosine with valine in factor IX still leads to some β-hydroxylation (P Handford, personal communication). Recent calcium-binding studies on isolated amino-terminal EGF modules from both factors IX and X, with and without β-hydroxylation, have shown that the post-translational modification has relatively little effect on calcium-binding affinity (M Mayhew, M Selander-Sunnerhagen, personal communication). Thus the biological role of the glycosylation and hydroxylation of these EGF modules remains uncertain.

Several recent studies have sought to identify particular EGF modules with a biological function in a variety of proteins by the study of various mutant forms of the parent gene. As this database grows and becomes more precise, it seems likely that most of these results will be readily interpreted in terms of the known structure of EGF modules and their ability to mediate protein–protein interactions via, for example, calcium.

Thrombomodulin is an endothelial cell thrombin receptor which contains six EGF modules. The fifth and sixth modules have been identified as important for thrombin binding by module deletion and peptide-competition experiments [44]. In *Drosophila*, various genes corresponding to cell-surface molecules containing multiple EGF modules have been found. These include *Notch* (36 EGF modules) [45], *Delta* (nine modules) and *Serrate* (14 modules). (Homologues of *Notch* have been found in both humans, TAN-1 [46], and *Xenopus*, *Xotch*, see Fig. 1.) Deletion experiments with *Notch* identified the 11th and 12th EGF modules as important for mediating interaction between *Notch* and *Delta* [45], and single amino acid substitutions in EGF modules of *Notch* and *Delta* were found to modify *Drosophila* development and affect cell adhesion *in vitro* [47].

Members of the selectin family of proteins are found on lymphocyte cell surfaces and bind to glycoproteins on endothelial cells. The three main types of selections — P, E and L — are distinguished by different numbers of C modules (found in many complement proteins; see Fig. 1). The main interaction with the glycoprotein is probably mediated by the C-type lectin domain but there is growing evidence that the EGF module might also be important [48].

In the nematode *C. elegans*, studies have identified a *glp-1* gene which is implicated in cell–cell interactions [49]. This gene has 10 EGF modules as well as other modules similar to those found in *Notch*. Loss-of-function mutants of *glp-1* have been studied and several missense mutations have been mapped to particular EGF modules [49].

Marfan syndrome is a relatively common inherited disease of connective tissue. The defective protein has recently been identified as fibrillin, a large glycoprotein which contains at least 34 EGF modules [50]. Several mutations leading to Marfan syndrome have been found in the EGF modules, including mutations in the consensus cysteine residues and the substitution of an arginine or proline in the main β-sheet [12,51]. These kinds of mutation would be expected to disrupt the structure of the EGF module.

Conclusions

EGF is an interesting example of the growing number of known modules or 'superfamilies'. The EGF module seems to provide a convenient structural scaffold for various functions, sometimes perhaps as a spacer unit on cell-surface proteins but mainly associated with specific protein–protein interactions. No doubt it will be found in numerous new proteins in the future. The idea that an identified consensus sequence will have a consensus structure that can be modelled relatively well has been strengthened by recent structural studies on different members of the EGF family. Our knowledge of the EGF structure now also allows one to make fairly confident predictions about the likely effects of particular mutations on function. Increasing information about how module deletions and residue changes within EGF modules affect function is beginning to show how amino acids can be placed on a module surface to perform a variety of tasks and how EGF modules might, along with other modules types, contribute to the overall function of the proteins in which it occurs.

Acknowledgements

ID Campbell is a member of the Oxford Centre for Molecular Sciences, which is supported by the Science and Engineering Research Council and Medical Research Council. P Bork is supported by the Deutsche Forschung Gemeinschaft. We thank various groups for information provided before publication including those of GG Brownlee, J Stenflo and A Tulinsky.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doolittle RF: **Reconstructing History with Amino Acid Sequences.** *Protein Sci* 1992, 1:191–200.

2. Patthy L: **Modular Exchange Principles in Proteins.** *Curr Opin Struct Biol*, 1991, 1:351–361.

3. Bork P: **Mobile Modules and Motifs.** *Curr Opin Struct Biol* 1992, 2:413–421.

4. Savage CR Jr, Inagami T, Cohen SJ: **The Primary Structure of Epidermal Growth Factor.** *J Biol Chem* 1972, 247:7612–7621.

5. Gray A, Dull TJ, Ullrich A: **Nucleotide Sequence of Epidermal Growth Factor cDNA Predicts a 128,000-Molecular Weight Protein Precursor.** *Nature* 1983, 303:722–725.

6. Scott J, Urdea M, Quiroga M, Sanchez-Pescador R, Fong N, Selby M, Rutter WJ, Bell GI: **Structure of a Mouse Submaxil-**

lary Messenger RNA Encoding Epidermal Growth Factor and Seven Related Proteins. *Science* 1983, 221:236–240.

7. DOOLITTLE RF, FENG DF, JOHNSON MS: **Computer-Based Characterization of Epidermal Growth Factor Precursor.** *Nature* 1984, 307:558–560.

8. DAVIS CG: **The Many Faces of Epidermal Growth Factor Repeats.** *New Biol* 1990, 2:410–419.

9. STENFLO J: **Structure–Function Relationships of Epidermal Growth Factor Modules in Vitamin K-Dependent Clotting Factors.** *Blood* 1991, 78:1637–1651.

10. BORK P: **Shuffled Domains in Extracellular Proteins.** *FEBS Lett* 1991, 268:47–54.

11. YAMADA Y, KLEINMAN HK: **Functional Domains of Cell Adhesion Molecules.** *Curr Opin Cell Biol* 1992, 4:819–823.

12. MOSHER DF, SOTTILE J, WU C, MCDONALD JA: **Assembly of Extracellular Matrix.** *Curr Opin Cell Biol* 1992, 4:810–818.

13. • WEN D, PELES E, CUPPLES R, SUGGS SV, BACUS SS, LUO Y, TRAIL G, HU S, SILBIGER SM, LEVY RB, KOSKI RA, LU HS, YARDEN Y: ***Neu* Differentiation Factor: a Transmembrane Glycoprotein Containing an EGF Domain and an Immunoglobulin Homology Unit.** *Cell* 1992, 69:559–572.

For the first time, an EGF module is found together with an immunoglobulin-like domain within one protein.

14. • PARTANEN J, ARMSTRONG E, MAKELA TP, KORHONEN J, SANDBERG M, RENKONEN R, KNUUTILA S, HUEBNER K, ALITALO K: **A Novel Endothelial Cell Surface Receptor Tyrosine Kinase with Extracellular Epidermal Growth Factor Homology Domains.** *Mol Cell Biol* 1992, 12:1698–1707.

The occurrence of EGF modules together with Ig-like domains and fibronectin type III repeats defines a new type of receptor protein tyrosine kinase.

15. HILL RJ, STERNBERG PW: **The Gene *lin-3* Encodes an Inductive Signal for Vulval Development in *C. elegans*.** *Nature* 1992, 358:470–476.

16. RUTLEDGE BJ, ZHANG K, BIER E, JAN YN, PERRIMON N: **The *Drosophila spitz* Gene Encodes a Putative EGF-Like Growth Factor Involved in Dorsal-Ventral Axis Formation and Neurogenesis.** *Genes Dev* 1992, 6:1503–1517.

17. HIGASHIYAMA S, LAU K, BESNER GE, ABRAHAM JA, KLAGSBRUN M: **Structure of Heparin-Binding EGF-Like Growth Factor.** *J Biol Chem* 1992, 267:6205–6212.

18. NAGLICH JG, METHERALL JE, RUSSELL DW, EIDELS L: **Expression Cloning of a Diptheria Toxin Receptor: Identity with a Heparin-Binding EGF-Like Growth Factor Precursor.** *Cell* 1992, 69:1051–1061.

19. • JIANG W, GORBEA CM, FLANNERY AV, BEYNON RJ, GRANT GA, BOND JS: **The α Subunit of Meprin A.** *J Biol Chem* 1992, 267:9185–9193.

Meprin A is a membrane-bound oligomeric zinc protease comprising two similar subunits, α and β. Several developmentally regulated proteins are known to contain a zinc protease which coexists with CUB domains and EGF modules (Fig. 1). In meprin, the CUB domain is missing and the EGF module is located adjacent to the transmembrane region, as in many other receptor-like proteins containing EGF modules.

20. • SHENG Z, WU K, CARRAWAY KL, FREGIENS N: **Molecular Cloning of the Transmembrane Component of the 13762 Mammary Adenocarcinoma Sialomucin Complex.** *J Biol Chem* 1992, 267:16341–16346.

The membrane-associated glycoprotein ASGP2 is part of a sialomucin complex which is thought to protect carcinomas from immune destruction. The two EGF modules might be involved in the ability of the complex to modulate EGF-receptor phosphorylation activity. The two other extracellular domains of the new mosaic protein ASGP2 remain to be classified.

21. • HOLMES WE, SLIWKOWSKI MX, AKITA R, HENZEL WJ, LEE J, PARK JW, YANSURA D, ABADI N, RAAB H, LEWIS GD *ET AL.*: **Identification of Heregulin, a Specific Activator of $p185^{erbB2}$.** *Science* 1992, 256:1205–1210.

Heregulins have been shown to form a group of highly similar proteins which might be natural ligands for $p185^{erbB2}$, a proto-ongogenic receptor protein tyrosine kinase. Heregulin α induces specifically phosphorylation of $p185^{erbB2}$. Heregulins are shown to contain an extracellular EGF module adjacent to a transmembrane region.

22. • KRETSCHMAR D, BRUNNER A, WIERSDORF V, PFLUGFELDER GO, HEISENBERG M, SCHNEUWLY S: ***giant lens*, a Gene Involved in Cell Determination and Axon Guidance in the Visual System of *Drosophila melanogoster*.** *EMBO J* 1992, 11:2531–2539.

The gene product *gil* is shown to be a new type of signalling molecule involved in developmental processes in the visual system of *Drosophila*. The carboxy-terminal EGF module follows an ~350-residue region for which no homology to any other protein has yet been found. It is speculated, by analogy, that *gil* might be the processed portion of a membrane-associated form.

23. NÖRENBERG U, WILLE H, WOLFF JM, FRANK R, RATHJEN FG: **The Chicken Neural Extracellular Matrix Molecule Restrictin: Similarity with EGF-, Fibronectin Type III-, and Fibrinogen-Like Motifs.** *Neuron* 1992, 8:849–863.

24. • USHKARYOV YA, PETRENKO AG, GEPPERT M, SÜDHOF TC: **Neurexins: Synaptic Cell Surface Proteins Related to the α-Latrotoxin Receptor and Laminin.** *Science* 1992, 257:50–56.

Two similar genes encoding neurexins use alternative promotors and variable spliced exons, to generate more than 100 different transcripts. Because of this polymorphism, their neural location and their partial sequence similarity to proteins associated with neurogenesis (e.g. agrin, laminin A and slit; see also Fig. 1), it is suggested that neurexins act as cell-recognition molecules in the nerve terminal.

25. BORK P, DOOLITTLE RF: **Proposed Acquisition of an Animal Protein Domain by Bacteria.** *Proc Natl Acad Sci USA* 1992, 89:8990–8994.

26. OKA T, MURATA Y, NAKANISHI T, YOSHIZUMI H, HAYASHIDA H, OHTSUKI Y, TOYOSHIMA K, HAKURA A: **Similarity, in Molecular Structure and Function, between the Plant Toxin Purothionin and the Mammalian Pore-Forming Proteins.** *Mol Biol Evol* 1992, 9:707–715.

27. TOH H, URADE Y, TANABE T: **Molecular Evolution of Enzymes Involved in the Arachidonic Acid Cascade.** *Mediators of Inflammation* 1992, 1:223–233.

28. RUPP F, PAYAN DG, MAGILL-SOLC C, COWAN DM, SCHELLER RH: **Structure of Rat Agrin.** *Neuron* 1991, 6:811–823.

29. TSIM KWK, RUEGG MA, ESCHER G, KRÖGER S, MCMAHAN UJ: **cDNA that Encodes Active Agrin.** *Neuron* 1992, 8:677–689.

30. •• MONTELIONE G, WÜTHRICH K, BURGESS AW, NICE EC, WAGNER G, GIBSON KD, SHERAGA HA: **Solution Structure of Murine Epidermal Growth Factor Determined by NMR Spectroscopy and Refined by Energy Minimization with Restraints.** *Biochemistry* 1992, 31:236–249.

A high-resolution NMR structure of mouse EGF.

31. •• HOMMEL U, HARVEY TS, DRISCOLL PC, CAMPBELL ID: **Human Epidermal Growth Factor: High Resolution Solution Structure and Comparison with Human Transforming Growth Factor α.** *J Mol Biol* 1992, 227:271–282.

A high-resolution NMR structure of human EGF and a comparison with TGF-α.

32. • KOHDE D, INAGAKI F: **Three-Dimensional Structure Nuclear Magnetic Resonance Structures of Mouse Epidermal Growth Factor in Acidic and Physiological pH Solutions.** *Biochemistry* 1992, 31:11928–11939.

Shows that there may be slight changes in conformation of EGF at different pH values.

33. HARVEY TS, WILKINSON AJ, TAPPIN MJ, COOKE RM, CAMPBELL ID: **The Solution Structure of Human Transforming Growth Factor-α.** *Eur J Biochem* 1991, 198:555–562.

34. Handford PA, Mayhew M, Baron M, Winship P, Brownlee GG, Campbell ID: Key Residues Involved in Calcium Binding Motifs in EGF-Like Domains. *Nature* 1991, 351:164–167.

35. • Baron M, Norman DG, Harvey TS, Handford PA, Mayhew M, Brownlee GG, Campbell ID: The Three-Dimensional Structure of the First EGF-Like Module of Factor IX: Comparison with EGF and TGF-α. *Protein Sci* 1992, 1:81–90.
An NMR structure of an EGF module from factor IX.

36. • Ullner M, Selander M, Persson E, Stenflo J, Dakenberg T, Telleman O: Three-Dimensional Structure of the Apo Form of the N-Terminal EGF-Like Module of Blood Coagulation Factor X as Determined by NMR Spectroscopy and Simulated Folding. *Biochemistry* 1992, 31:5974–5983.
Another NMR structure of an EGF module from factor X, this time without calcium.

37. •• Selander-Sunnerhagen M, Ullner M, Persson E, Teleman O, Stenflo J, Drakenberg T: How an Epidermal Growth Factor (EGF)-Like Domain Binds Calcium. *J Biol Chem* 1992, 267:19642–19649.
An NMR structure of a calcium-bound EGF module from factor X. This paper contains a good review of calcium-binding EGF modules.

38. Koide H, Muto Y, Kasai H, Kohri K, Hoshi K, Takahashi S, Takusari H, Sasaki T, Tsukumo K, Miyake T *et al*: Recognition of an Antiparallel β-Sheet of Human Epidermal Growth Factor by its Receptor. *FEBS Lett* 1992, 302:39–42.

39. Shiah H-S, Chen T-Y, Chang C-M, Chow JT, Kung H-J, Hwang J: *Pseudomonas* Exotoxin A-Epidermal Growth Factor Mutant Chimaeric Protein as an Indicator for Identifying Amino-Acid Residues Important in EGF–Receptor Interaction. *J Biol Chem* 1992, 267:24034–24040.

40. Richter A, Conlan JW, Ward ME, Chamberlin SG, Alexander P, Richards NGJ, Davies DE: Multidomain Binding of Transforming Growth Factor α to the Epidermal Growth Factor Receptor. *Biochemistry* 1992, 31:9546–9554.

41. Mayhew M, Handford P, Baron M, Tse AGD, Campbell ID, Brownlee GG: Ligand Requirements for Ca^{2+} Binding to EGF-Like Domains. *Protein Eng* 1992, 5:489–494.

42. Nishimura H, Takao T, Hase S, Shimonishi Y, Iwanaga S: Human Factor IX Has a Tetrasaccharide O-Glycosidically Linked to Serine 61 through the Fucose Residue. *J Biol Chem* 1992, 267:17520–17525.

43. Nishimura H, Yamashita S, Zeng Z, Walz DA, Iwanaga S: Evidence for the Existence of O-Linked Sugar Chains Consisting of Glucose and Xylose in Bovine Thrombospondin. *J Biochem* 1992, 111:460–464.

44. Tsiang M, Lentz SR, Sadler JE: Functional Domains of Membrane-Bound Human Thrombomodulin. *J Biol Chem* 1992, 267:6164–6170.

45. Rebay I, Fleming RJ, Fehon RG, Cherbas L, Cherbas P, Artavanis-Tsakonas S: Specific EGF Repeats of ***notch*** Mediate Interactions with ***delta*** and ***serrate***: Implications for ***notch*** Has a Multifunctional Receptor. *Cell* 1991, 67:687–699.

46. Ellisen LW, Bird J, West DC, Lee Soreng A, Reynolds TC, Smith SD, Sklar J: ***TAN-1***, the Human Homolog of the ***Drosophila notch*** Gene Is Broken by Chromosomal Translocations in T-Lymphoblastic Neoplasms. *Cell* 1991, 66:649–661.

47. Lieber T, Wesley CS, Alcamo E, Hassel B, Krane JF, Camposortega JA, Young MW: Single Amino-Acid Substitutions in EGF-Like Elements of ***notch*** and ***delta*** Modify ***Drosophila*** Development and Affect Cell Adhesion ***in vitro***. *Neuron* 1992, 9:847–859.

48. Imai Y, Lasky LA, Rosen SD: Further Characterization of the Interaction Between L-Selectin and its Endothelial Ligands. *Glycobiology* 1992, 2:373–381.

49. Kodoyianni V, Maine EM, Kimble J: Molecular Basis of Loss-of-Function Mutations in the ***glp-1*** Gene of ***Caeorhabitis elegans***. *Mol Biol Cell* 1992, 3:1199–1213.

50. Maslen CL, Corson GM, Maddox BK, Glanville RW, Sakai LY: Partial Sequence of a Candidate for the Marfan Syndrome. *Nature* 1991, 352:334–337.

51. Dietz HC, Pyeritz RE, Puffenberger EG, Kenzdior RJ, Corson GM, Maslen CL, Sakai LY, Francomano CA, Cutting GR: Marfan Phenotype Variability in a Family Aggregating a Missense Mutation in the EGF-Like Motif of the Fibrillin Gene. *J Clin Invest* 1992, 89:1674–1680.

ID Campbell, Department of Biochemistry, University of Oxford, South Park Road, Oxford OX1 3QU, UK.

P Bork, European Molecular Biology Laboratory, Meyerhofstraße 1, 6900 Heidelberg, Germany.