PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Structure and Function of Protein Modules

Iain D. Campbell and Martin Baron

Phil. Trans. R. Soc. Lond. B 1991 **332**, 165-170
doi: 10.1098/rstb.1991.0045

The structure and function of protein modules

IAIN D. CAMPBELL AND MARTIN BARON

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

SUMMARY

Analysis of protein sequences shows that many proteins in multicellular organisms have evolved by a process of exon shuffling, deletion and duplication. These exons often correspond to autonomously folding protein modules. Many extracellular enzymes have this modular structure; for example, serine proteases involved in blood-clotting, fibrinolysis and complement. The main role of these modules is to confer specificity by protein–protein interactions. Lack of structural information about such proteins has required a new strategy for studying the structure and function of protein modules. The strategy involves the production of individual modules by protein expression techniques, determination of their structure by high resolution nuclear magnetic resonance and definition of functional patches on the modules by site-directed mutagenesis and biological assays. The structures of the growth factor module, the fibronectin type 1 module and the complement module are briefly described. The possible functional roles of modules in various proteins, including the enzymes factor IX and tissue plasminogen activator, are discussed.

1. INTRODUCTION

There is, currently, an explosive growth in amino acid and nucleic acid sequence information. One of the striking features of these sequences is that, on analysis, certain patterns of amino acids, consensus sequences, appear over and over again. For example, a consensus sequence first recognized in epidermal growth factor (EGF) can be identified over 300 times in a wide variety of proteins. In many cases the domains or modules defined by the consensus sequences correspond to single exons and it is extremely likely that they have evolved and multiplied by a process of exon shuffling, insertion and duplication (Doolittle 1989; Patthy 1985, 1987). Modules are usually small, in the range 40–100 amino acids, and often fold autonomously (Baron *et al.* 1991).

The main biological role of protein modules seems to be for specific protein–protein interactions. They appear, for example, in cell surface molecules, such as receptors (Bazan 1990) and cell adhesion molecules (Bevilacqua *et al.* 1989; Springer 1990) and extracellular matrix proteins, such as fibronectin (Yamada 1989). They also appear in various serine proteases associated with blood clotting, fibrinolysis and complement (Neurath 1989); some members of the serum protease class of protein are shown in figure 1 and more information on the individual modules is given in the figure caption. The modular proteins found so far have been mainly extracellular although intracellular ones are also being found (Labeit *et al.* 1990).

In this paper, recent studies on the structure and function of some protein modules will be described with emphasis on proteins associated with blood clotting and fibrinolysis.

2. MODULAR PROTEINS AND THE STRUCTURES OF SOME MODULES

In general, proteins containing modules have been rather difficult to crystallize. This may be because they are usually glycosylated and possibly because of extensive flexibility in a protein consisting of a mosaic of modules (see figure 1). In addition, many of these proteins have only been available in rather small amounts and so were difficult to investigate by physical techniques. In recent years alternative approaches have become possible because of the relative ease with which one can produce proteins by the expression of heterologous genes using recombinant DNA. It is thus possible to produce fragments of modular proteins and determine their structure. This depends on the individual modules folding autonomously, but this often appears to occur. This approach has been used, for example, to investigate the structure of a pair of immunoglobulin modules from the cell-surface glycoprotein CD4 by X-ray diffraction methods (Wang *et al.* 1990; Ryu *et al.* 1990). As will be discussed here, recent advances in nuclear magnetic resonance (NMR) technology have also resulted in the determination of solution structures of a number of protein modules.

A strategy for studying protein modules has recently been outlined (Baron *et al.* 1991). The strategy can be summarized as follows. A module is identified in a sequence database; its gene is synthesized and expressed in a host cell such as yeast; the module is purified and its structure determined by NMR; this consensus structure can be used to model related modules; functional patches on the modules are recognized by sequence comparison, assay and site-specific mutagenesis; a model of the intact protein can be built up from low-resolution information from

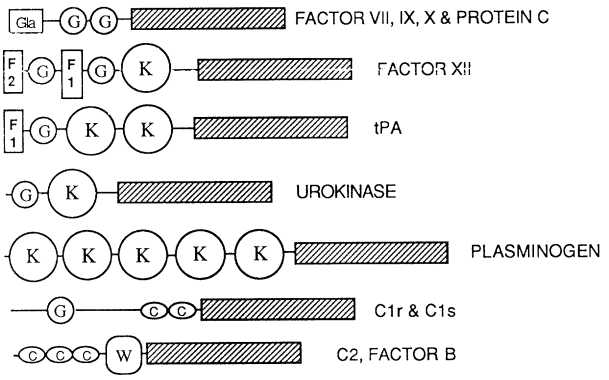

Figure 1. A diagram of the mosaic nature of some serine proteases found in serum. Factors VII, IX, X, XII and protein C are involved in blood clotting (Mann *et al.* 1988); tPA, urokinase and plasminogen are involved in fibrinolysis (Furie & Furie 1988); C1r, C1s, C2 and factor B are involved in complement (Law & Reid 1988).

The module Gla contains γ-carboxyglutamate and its structure has been determined by X-ray crystallography (Soriano-Garcia *et al.* 1989). The structure of the kringle module (K) has been determined both by diffraction (Soriano-Garcia *et al.* 1989, Harlos *et al.* 1987) and by NMR (Atkinson & Williams 1990). The structures of both the C module, also known as the short consensus repeat (SCR) and the complement control protein (CCP) module, and the type 1 module of fibronectin (F1) have been determined recently in this laboratory (Barlow *et al.* 1991; Baron *et al.* 1990). The G module is like epidermal growth factor (EGF) and its structure has been determined by NMR (see, for example, Campbell *et al.* 1990). The C7 module is found in various complement components such as C6 and C7 (Reid & Day 1990). The serine protease module (shaded box) is like trypsin (Neurath 1989) and the W module is found in the von Willebrand factor which is a multimeric plasma protein which associates with factor VIII (Mann *et al.* 1988; Reid & Day 1990).

microscopy plus information about how modules fit together, derived from studies of module pairs.

The technology associated with the determination of the structure of proteins in solution by using high resolution [^{1}H]-NMR is relatively recent (Wüthrich 1989; Bax 1989) but it is particularly convenient for this kind of study where relatively small polypeptides are involved. The current limit in protein size is around 150 amino acids, but this allows the determination of the structure of most modules and many module pairs. It is now possible to determine such structures in a relatively short time without the requirement of crystals or isomorphous heavy atom derivatives.

Figure 2 shows, schematically, structures of three modules derived in our laboratory. These were produced by using a yeast expression system that secreted the module into the cell medium (Baron *et al.* 1990). This was followed by analyses of high resolution NMR data and computation of families of structures consistent with the NMR restraints.

The G structure shown in figure 2 was derived from studies on human EGF (Cooke *et al.* 1987) and transforming growth factor alpha (TGF-α) (Tappin *et al.* 1989; Campbell *et al.* 1990). This module has three disulphide bridges and consists of a major double stranded antiparallel β-sheet with the three disulphides radiating up from one face of this sheet, attaching a rather ill-defined N-terminal strand and loop and a well-defined C-terminal region. (Lack of definition in NMR determined structures arises either from flexibility of the protein fragment or experimental difficulties; these two possibilities can often be distinguished by relaxation experiments.)

The dominant structural feature of the F1 module, which is about 40 residues long, is again antiparallel β-sheet (Baron *et al.* 1990). Two such sheets fold over a core of hydrophobic residues which are conserved in all members of the F1 module family. One of the two disulphide bonds joins these two sheet regions, the other stabilizes two adjacent strands of the main sheet.

The structure of the C module, which is about 60 amino acids long, has only been determined very recently (Barlow *et al.* 1991). This is again mainly formed from extended strands and β-sheet that fold around conserved hydrophobic residues.

It is worth emphasizing that, so far, evidence suggests that each member of a module family has the same

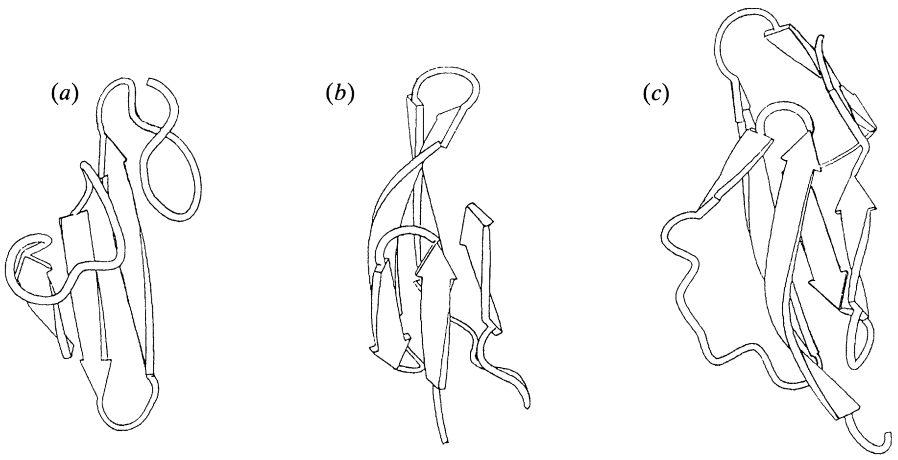

Figure 2. Schematic representations of the structures of three of the modules in figure 1. (*a*) shows the structure of a G module (Cooke *et al.* 1987); (*b*) shows the structure of an F1 module (Baron *et al.* 1990), and (*c*) shows the structure of a C module (Barlow *et al.* 1991). These structures were derived with NMR data collected from a solution containing modules produced by recombinant DNA techniques.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

global fold but that different modules in a family can have different functional roles. The ability of proteins with a similar basic structure to bind a wide variety of molecules is well known for the immunoglobulin superfamily (Williams & Barclay 1988). It is becoming clear, however, that a similar multifunctional role is exhibited by other modules. Various very different functional patches have been identified on G modules; for example, a relatively large patch appears to be necessary on growth factors that bind to the EGF receptor (Campbell *et al.* 1990); an RGD sequence has been found to be involved in binding of cells to the basement membrane protein nidogen (Mann *et al.* 1989); a calcium-binding patch has been found in the N-terminal region of G modules from a wide variety of proteins including factor IX (Handford *et al.* 1990; Persson *et al.* 1989) and, in tissue plasminogen activator (tPA), yet another patch on a different part of the G structure has been implicated in the clearance of tPA from plasma by a liver receptor (Mark Edwards, British Biotechnology, private communication). The factor IX and tPA examples will be discussed in greater detail below.

It is of interest to consider how these different modules might fit together. The N- and C-termini of the C module are at different ends of the molecule and one might expect that this would facilitate a 'beads on a string' kind of assembly. The usual exon boundaries of the F1 and G modules, however, appear to leave uncompleted β-sheet structures at their C- and N-termini, respectively. It is thus tempting to speculate that these two modules might link up to complete a β-sheet structure (see discussion below on tPA and figure 4*b*). In contrast to the three modules in figure 2, the N- and C-termini of the kringle module are close together in space. This would be expected to facilitate the formation of a coiled structure.

3. THE ROLE OF SOME MODULES IN BLOOD CLOTTING AND FIBRINOLYSIS

Vascular injury leads to the formation of an insoluble clot formed from fibrin and platelets (Mann *et al.* 1988; Furie & Furie 1988). This is brought about by a series of proteolytic enzyme reactions. This cascade of zymogen activations leads to the formation of thrombin and the subsequent cross-linking of a fibrin network. The cascade is tightly controlled by the presence of various regulatory proteins, e.g. factor VIII, and anticoagulant factors, e.g. protein C (Esmon 1989). In addition to clot formation, dissolution of clots is an important pathway and this, too, involves various zymogen activations leading to an attack on fibrin by plasmin (Furie & Furie 1988; Haber *et al.* 1989).

The role of modules will be briefly discussed in two examples here. One is the role of a G module and calcium ions in a complex associated with the formation of a blood clot, the other is the possible role of the F1-G module pair in tissue plasminogen activator, an enzyme which activates plasminogen.

(*a*) *Modules in factor IX*

Factor IX and factor X are similar serine protease clotting factors (see figure 1). They are made up of a Gla module which contains modified glutamic acid residues, γ-carboxyglutamate, two G modules and a serine protease domain which is homologous to trypsin. The first G module usually contains either a β-hydroxyaspartate or a β-hydroxyasparagine. Factors IX and X form part of the coagulation cascade; the role of activated factor IX (IXa) being to activate factor X by cleaving its polypeptide chain. The chains remain attached after cleavage by means of a disulphide bridge.

Molecular defects in factor IX can lead to haemophilia B (Furie & Furie 1988; Lozier *et al.* 1990). Various experiments on fragments and different components have indicated that factors IXa and X form a complex with activated factor VIII (Mann *et al.* 1988). Factor VIII is a soluble plasma protein which binds with high affinity to certain phospholipid membranes and acts as a cofactor in the activation of factor X. Calcium ions are also required in this complex, as shown schematically in figure 3. Figure 3 is based on previously published schemes for the complex involving factor Xa, factor V and prothrombin (Harlos *et al.* 1987; Mann *et al.* 1988). This complex and the one shown in figure 3 are expected to be similar except that prothrombin contains kringle modules.

It is well known that the γ-carboxyglutamate residues in the Gla module bind calcium. Calcium ions appear not only to be associated with the Gla module, however, but also with the first G modules of factors IX and X. A consensus sequence D/N, D/N, D*/N*, Y/F (* denotes β-hydroxylation) has been recognized within this module (Rees *et al.* 1988), not only in factors IX and X but also in various other proteins including *notch* from *Drosophila* (Fehon *et al.* 1990). After obtaining the structure of a G module, Cooke *et al.* (1987) proposed that this sequence could form a calcium binding site, on one face of the G module. The putative calcium binding residues are shown in bold in figure 4*a*, together with other 'structural' consensus residues of the G module shown in normal lettering. The importance of these residues is also implied by known point mutations in the first G module of factor

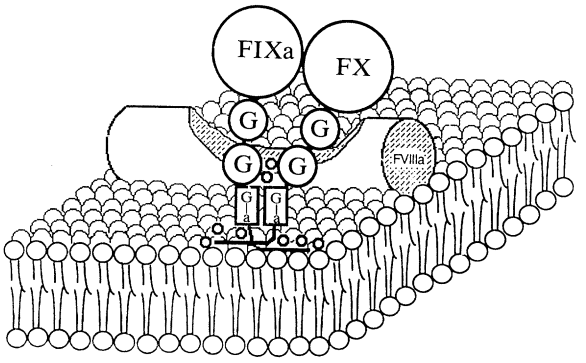

Figure 3. A schematic view of the coagulation enzyme complex involving factors VIIIa, IXa and X, calcium ions and phospholipid. Note the calcium ions (○) associated with the two G modules of factors IXa and X as well as those associated with the Gla modules and phospholipid.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

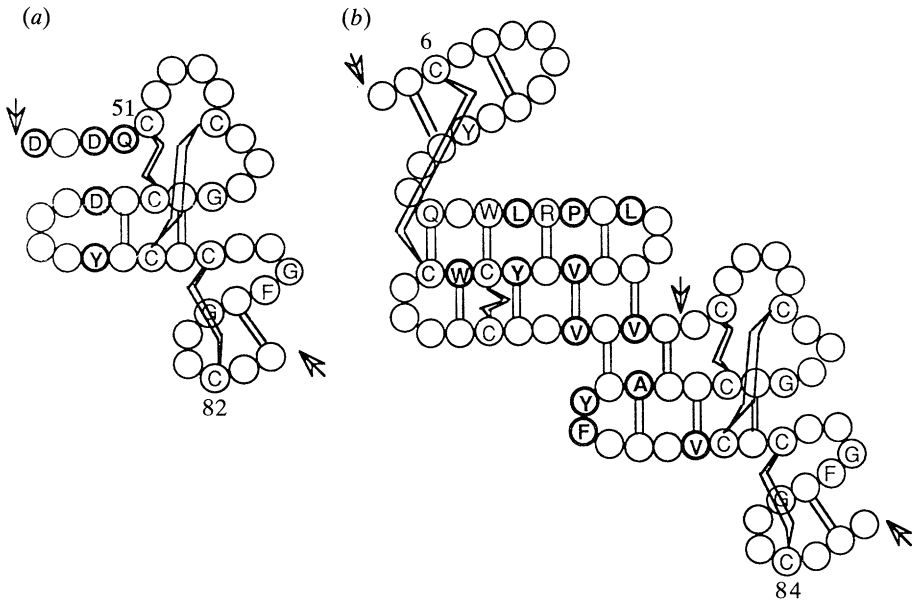

Figure 4. Structural representations of modules, showing the residue numbering in the intact proteins, the disulphide bonds, and possible H-bonding patterns. The consensus sequences are marked in normal lettering and bold letters indicate possible functional residues. The arrows show the exon boundaries of the F1 and G modules. (*a*) Corresponds to the first G module in factors IX and X, with the proposed calcium binding residues in bold. (*b*) Shows a possible structure of the F1-G module pair in tPA; a hydrophobic patch, formed by side-chains on one face of this model, is identified in bold.

IX which lead to haemophilia B (e.g. Lozier *et al.* 1990).

To explore in more detail the nature of this calcium binding site and its relationship to the complex in figure 3, we set out, in a collaboration with George Brownlee and his colleagues to define the role of the first G module of factor IX. This module was produced by recombinant methods and peptide synthesis. It was shown then that the NMR resonances of the tyrosine (Y69) in the consensus binding sequence were perturbed by calcium. This allowed an apparent K_d of 250 μM to be measured (Handford *et al.* 1990). In an independent study on the first G module produced from bovine factor X by proteolysis Persson *et al.* (1989) showed that this, too, bound calcium.

The next step was to obtain more precise information about residues involved in calcium binding by using site-directed mutagenesis. The effect of changed amino acids on calcium binding efficiency to the individual G modules was compared with the effect of such changes on the clotting efficiency of the intact protein (Handford *et al.* 1991). All the consensus binding residues have now been changed. A clear correlation between apparent K_d for calcium binding to the recombinant module and clotting efficiency of the intact protein was observed. For example, when the relatively conservative changes D47 → E and D64 → N were made, the calcium binding to the module was greatly reduced. These two mutations are known to cause haemophilia B and reduce clotting activity by a factor of about 100.

Another feature of investigations of module function is that there are large numbers of available module sequences, thus making sequence comparison a powerful tool (e.g. the prediction of a receptor binding patch on EGF, Campbell *et al.* (1990)). A comparison of putative calcium binding sequences on other G modules (e.g. on *notch*, Fehon *et al.* (1990)) indicated that the position equivalent to D64 could be changed into N. In such modules, however, the residue equivalent to Q50 was usually E. This observation together with site-directed mutagenesis studies on recombinant G modules led to the realization that Q50 is important for the formation of the calcium binding site. This residue is therefore added to the consensus calcium binding patch shown in figure 4*a*.

These studies on this functional patch on a G module of factor IX indicate that the formation of the correct calcium binding site on the G module is very important in the complex shown in figure 3. Because the first G module on factor X has a very similar calcium binding site to the one on factor IX, it is not inconceivable that the two sites interact with each other in some way.

(*b*) *Modules in tissue plasminogen activator*

Tissue plasminogen activator (tPA) is involved in the removal of fibrin from the vascular system. The key enzyme in fibrinolysis is plasmin but the zymogen plasminogen must first be activated by a plasminogen activator. Of several such activators known, tPA appears to be the most specific for clot-associated plasminogen. The most likely mechanism for this specificity is a complex formed from plasminogen, fibrin and tPA (see, for example, Harris (1987); Haber *et al.* (1989); Higgins & Bennett (1990)).

Since tPA has considerable potential therapeutic and commercial value as a thrombolytic agent it has been extensively studied and many protein variants have been produced by recombinant methods. The

main aims of these studies have been to improve fibrinolysis specificity, to increase the rate of thrombolysis and to increase the lifetime of the circulating protein (the *in vivo* half-life is only a few minutes).

tPA is composed of an F1 module, a G module, 2 K modules and a serine protease domain. There are also four potential N-glycosylation sites. More than 50 variant tPAs have been produced with module additions and deletions. The results have not always been clear cut and none of these variants has been shown to be superior to wild-type tPA in animal models (Higgins & Bennett 1990). It is generally agreed, however, that the F1 module and the second kringle are involved in binding to fibrin and that the G module is involved in clearance by a liver receptor.

To our knowledge no crystals of tPA suitable for X-ray studies have been produced, although much effort has gone into this by various groups. Before our studies, structures were available for the K module and the serine protease domain. Now that we also have knowledge of the consensus structures of F1 (Baron *et al.* 1990) and G (Campbell *et al.* 1990), structural information is available for all the components of the tPA mosaic. This does not, of course, give information about the overall shape of the molecule directly, since there are many possible ways the different components could fit together. It should, however, be possible to combine information about the individual module structures with information obtained from lower resolution methods, such as microscopy and scattering to refine models of the intact structure. Another way of obtaining this information might be to determine the structures of all the possible module pairs, e.g. F1-G, G-K and so on. In the case of tPA we have recently expressed an F1-G module pair for structural studies.

Until experimental information is available, one can only speculate about how any given pair of modules might fit together. However, as discussed in §2 above, the F1 and G modules appear to be constructed such that they have incomplete β-sheets at their C- and N-termini, respectively. In fact, the β-strands of the F1 module are rather like the G module in reverse (figure 4). It is therefore easy to visualize how the two modules might come together to complete these sheets as shown in figure 4*b*. A similar kind of rigid connection between two immunoglobulin modules has recently been reported in a study, by X-ray diffraction, of a recombinant fragment of CD4 (Wang *et al.* 1990). The kind of structure proposed in figure 4*b* would have one predominantly hydrophobic face and one predominantly hydrophilic face. One possibility is that the hydrophobic face could pack against the kringle modules, thus leading to a relatively compact structure for tPA. We expect to be able to produce a structure of the F1-G pair shortly and to combine this information with scattering data being collected by Colin Blake's group in Oxford to obtain a more complete picture of the solution structure of tPA.

4. CONCLUSIONS

The strategy outlined here, which involves the production of individual modules, the determination of their structure by NMR and the definition of functional patches by mutagenesis and assay seems to be extremely powerful. The method has, so far, been applied successfully in this laboratory to various G modules, an F1 module and a C module. Numerous other modules and module pairs are currently under investigation and a rich harvest of information about modular proteins and their role in protein–protein interactions is expected soon.

This is a contribution from the Oxford Centre for Molecular Science which is supported by SERC and MRC. We also thank ICI Pharmaceuticals and British Biotechnology for their financial and technical support. The work owes much to many colleagues in Oxford, including Paul Barlow, Jonathan Boyd, Paul Driscoll, Tim Dudgeon, Tim Harvey, Uli Hommel and David Norman, who provided figure 2. We are particularly indebted to the George Brownlee group, especially Penny Handford and Mark Mayhew, with whom we collaborate on the work on FIX modules.

REFERENCES

Atkinson, A. & Williams, R. J. P. 1990 Solution structure of the kringle 4 domain from human plasminogen by ^{1}H NMR and distance geometry. *J. molec. Biol.* **212**, 541–552.

Barlow, P., Norman, D. G., Baron, M. & Campbell, I. D. 1991 The secondary structure of a complement control protein module by two-dimensional ^{1}H NMR. *Biochemistry* **30**, 997–1004.

Baron, M., Norman, D. & Campbell, I. D. 1991 Protein modules *TIBS*. **16**, 13–17.

Baron, M., Norman, D. G., Willis, A. & Campbell, I. D. 1990 Structure of the fibronectin type 1 module. *Nature, Lond.* **345**, 642–646.

Bax, A. 1989 2D NMR and protein structure. *A. Rev. Biochem.* **58**, 223–256.

Bazan, J. F. 1990 Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. nat. Acad. Sci., U.S.A.* **87**, 6934–6938.

Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A. & Seed, B. 1989 Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins. *Science, Wash.* **243**, 1160–1165.

Campbell, I. D., Cooke, R. M., Baron, M., Harvey, T. S. & Tappin, M. J. 1989 The solution structures of EGF and TGF-α. *Prog. grow. Fact. Res.* **1**, 13–22.

Campbell, I. D., Baron, M., Cooke, R. M., Dudgeon, T. J., Fallon, A., Harvey, T. S. & Tappin, M. J. 1990 Structure function relationships in EGF and TGF-α. *Biochem. Pharmac.* **40**, 35–40.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H. & Sheard, B. 1987 The solution structure of human epidermal growth factor. *Nature, Lond.* **327**, 339–341.

Doolittle, R. F. 1989 Similar amino-acids revisited. *TIBS* **14**, 244–245.

Dudgeon, T. J., Baron, M., Cooke, R. M., Campbell, I. D., Edwards, R. M. & Fallon, A. 1990 Structure and function of hEGF: receptor binding and NMR. *FEBS Lett.* **261**, 392–396.

Esmon, C. T. 1989 The roles of protein C and thrombomodulin in the regulation of blood coagulation. *J. biol. Chem.* **264**, 4743–4746.

Fehon, R. G., Kooh, P. J., Rebay, I., Regan, C. L., Xu, T., Muskavitch, M. A. T. & Artavanis-Tsakonas, S. 1990 Molecular interactions between the protein products of the

neutrogenic loci *notch* and *delta*, two EGF-homologous genes in *Drosophila*. *Cell* **61**, 523–534.
Furie, B. & Furie, B. C. 1988 The molecular basis of blood coagulation. *Cell* **53**, 505–517.
Haber, E., Quertermous, T., Matsueda, G. R. & Runge, M. S. 1989 Innovative approaches to plasminogen activator therapy. *Science, Wash.* **243**, 51–56.
Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G. & Campbell, I. D. 1990 The first EGF-like domain of human factor IX has a high affinity Ca^{++} binding site. *EMBO J.* **9**, 475–480.
Handford, P. A., Mayhew, M., Baron, M., Winship, P. R., Campbell, I. D. & Brownlee, G. G. *Nature, Lond.* (In the press.)
Harris, T. J. R. 1987 Second generation plasminogen activators. *Protein Engineering* **1**, 449–458.
Harlos, K., Holland, S. K., Boys, C. W. G., Burgess, A. I., Esnouf, M. P. & Blake, C. C. F. 1987 Vitamin K-dependent blood coagulation proteins form heterodimers. *Nature, Lond.* **330**, 82–84.
Higgins, D. L. & Bennett, W. F. 1990 Tissue plasminogen activator: the biochemistry and pharmacology of variants produced by mutagenesis. *A. Rev. Pharmac. Toxicol.* **30**, 91–121.
Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, Francke, U., Leonard, K., Wardale, J., Whiting, A. & Trinick, J. 1990 A regular pattern of two types of 100-residue motif in the sequence of titin. *Nature, Lond.* **345**, 273–276.
Law, S. K. A. & Reid, K. B. M. 1988 *Complement* Oxford: IRL Press.
Lozier, J. N., Monroe, D. M., Stanfield-Oakley, S., Lin, S.-W., Smith, K. J., Roberts, H. R. & High, K. A. 1990 Factor IX New London: substitution of proline for glutamine at position 50 causes several hemophilia B. *Blood* **75**, 1097–1104.
Mann, K., Deutzmann, Aumailley, M., Timpl, R., Raimondi, L., Yamada, Y., Pan, T., Conway, D. & Chu, M.-L. 1989 Amino-acid sequence of mouse nidogen, a multidomain basement membrane protein with binding activity for laminin, collagen IV and cells. *EMBO J.* **8**, 65–72.
Mann, K. G., Jenny, R. J. & Krishnaswamy, S. 1988 Cofactor proteins in the assembly and expression of blood clotting enzyme complexes. *A. Rev. Biochem.* **57**, 915–956.
Neurath, H. 1989 Proteolytic processing and physiological regulation. *TIBS* **14**, 268–271.
Patthy, L. 1985 Evolution of the proteases of the blood clotting and fibrinolysis by assembly from modules. *Cell* **41**, 657–663.
Patthy, L. 1987 Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* **214**, 1–7.
Persson, E., Selander, M., Linse, S., Drakenberg, T., Ohlin, A. & Stenflo, J. 1989 Calcium binding to the isolated β-hydroxyaspartic acid-containing epidermal growth factor-like domain of bovine factor X. *J. biol. Chem.* **264**, 16897–16904.
Rees, D. J. G., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J. & Brownlee, G. G. 1988 The role of β-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX. *EMBO J.* **7**, 2053–2061.
Reid, K. B. M. and Day, A. J. 1989 Structure function relationships of the complement system. *Immun. Today* **10**, 177–180.
Ryu, S. E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N., Axel, R., Sweet, R. W. & Hendrickson, W. A. 1990 Crystal structure of HIV-binding recombinant fragment of human CD4. *Nature, Lond.* **348**, 419–426.
Soriano-Garcia, M., Park, C. H., Tulinsky, A., Ravichandran, K. G. & Skrzypzcak-Jankun, E. 1989 Structure of Ca^{++} prothrombin fragment 1 including the conformation of the gla domain. *Biochemistry* **28**, 6805–6810.
Springer, T. A. 1990 Adhesion molecules of the immune system. *Nature, Lond.* **346**, 425–434.
Tappin, M. J., Cooke, R. M., Fitton, J. & Campbell, I. D. 1989 A high resolution ^1H NMR study of hTGF-α: structure and pH dependent conformational interconversion. *Eur. J. Biochem* **179**, 629–637.
Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L. & Harrison, S. C. 1990 Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature, Lond.* **348**, 411–418.
Williams, A. F. & Barclay, A. N. 1988 The immunoglobulin superfamily – domains for cell surface recognition. *A. Rev. Immunol.* **6**, 381–405.
Wüthrich, K. 1989 Protein structure determination in solution by NMR. *Science, Wash.* **243**, 45–50.
Yamada, K. M. 1989 Fibronectins: structure functions and receptors. *Curr. Opin. Cell Biol.* **1**, 956–963.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES