Volume 286, number 1,2, 47–54 FEBS 09908 July 1991

ADONIS 001457939100618F

Shuffled domains in extracellular proteins

Peer Bork

Central Institute of Molecular Biology, Robert-Rössle-Str. 10, O-1115 Berlin, Germany

Received 22 April 1991; revised version received 14 May 1991

A comprehensive list of domains in extracellular mosaic proteins is presented. About 40 domains were distinguished by consensus patterns. A subsequent sequence database search recognized these domains in more than 200 extracellular proteins. The results point to a structural network, which may also represent the molecular basis for a complex coordination of various functions within the world of extracellular proteins.

Mosaic protein; Extracellular domain; Pattern recognition

1. INTRODUCTION

With the increasing amount of primary structures also numerous extracellular proteins could be characterized at the sequence level. Many of them are called mosaic proteins, because they result from exon shuffling [1] and therefore contain a set of different structural units, domains. A lot of homologies between those domains were detected so far (e.g. [2,3]). Since there is a correlation between such domains and some special types of surrounding introns [4] the exon shuffling can be assumed to be a 'fast tool' of evolution. Therefore, it is not a surprise that similar domains were found to be widespread among seemingly unrelated extracellular proteins. An estimate for the number of original exons says that only a few thousand exons could be sufficient to produce the current universe of proteins via exon shuffling [5]. At present, no conclusion is possible as to the conservation of function in these structural units. In many cases a common function of these domains in different proteins was proposed, but sometimes functions may have changed after shuffling. Often, the similarities between related domains are rather weak and fall into the so-called 'twilight zone' [6]. In order to identify domains, even if the homologies are very weak, and to obtain a unique description of such domains we have been using our property pattern approach [7] which is sensitive enough to find distant similarities between proteins [8]. Having described by this kind of consensus pattern about 40 domains (Fig. 1), a subsequent homology search in protein sequence databases identified a number of extracellular proteins, supplying an overview about the occurrence of shuffled domains in the most diverse biochemical pathways. For some of them, like the EGF domain in period clock protein from mouse or the VWA domain in the malaria thrombospondin related anonymous protein (TRAP), relationships to other domains are not reported so far. Surely, the list is far from being complete because many more domains exist. Nevertheless, we are able to present a comprehensive list (in terms of our domain description) of mosaic proteins which are composed of defined domains involved in a network of extracellular processes. The list is not restricted to well-defined (known) systems like coagulation (Fig. 2) and complement (Fig. 3). Globular domains may coexist within one molecule with nonglobular domains of uncertain length (Figs. 1,4). In spite of a rapid evolution, some of the domains can also be found in invertebrates (Fig. 5). The origin of the similar domains in single cell parasites or viruses (Fig. 6) remains uncertain, but a gene transfer is probable.

Correspondence address: P. Bork, Central Institute of Molecular Biology, Department of Biomathematics, Robert-Rössle-Str. 10, O-1115 Berlin, Germany

2. STRUCTURAL FEATURES

Extracellular domains are often characterized by specific cysteine patterns. The cysteines form disulfide bridges, stabilizing the folded structure. The connecting segments between these structural elements are very flexible in length as well as in amino acid composition and may have different binding specificities. A similar situation was found in domains without disulfide bridges like Fn3 and VA. Even if in equivalent domains not a single amino acid is absolutely conserved, there are nevertheless property patterns in all corresponding domains suggesting common elements of secondary structure. The more flexible regions between these segments complicate any multiple sequence alignment.

Some of the domains like EGF, SCR or Fn3, seem to be more widespread than others among the different extracellular complexes, but the more specimens of a do-

domains	symbol	aa	domains	symbol	aa
gamma-carboxyl glutamate domain	G	60	complement C1q/collagen C-term. rep.		135
kringle	K	120	fibrillar collagen C-term. domain		260
Kallekrein/F XI repeats		95	avidin like domain	Av	140
fibronectin-repeat type 1		40	immunglobulin domain	Ig	100
fibronectin-repeat type 2		60	proteogluc. tandem repeat	PT	100
fibronectin-repeat type 3		90	Kunitz type inhibitor domain	KI	100
EGF like domain		40	mannose 6 phosphate/IGF 2 rec. rep.	M6	150
serine protease domain	SPR	230	coagulation factor 5/8 A domain	F8A	330
type C lectin domain	LC	130	coagulation factor 5/8 C domain	F8C	150
short consensus repeat		65	N-terminal collagen 9/12 domain		250
LDL-receptor/MAC repeat		40	T+Y region in lin-12/glp-1		30
YWTD-repeat	Y	50	repeat of notch/lin-12/glp-1		30
TGF binding protein repeat		70	cytokine receptor domain	4C	110
specif. repeat Fi/MAC		80	cysteine rich receptor repeat	CRR	170
perforin/MAC repeat	PFM	250	laminin A/agrosin G-repeat	GR	190
specif. repeat C1r/C1s/uEGF	C1	115	keratinsulfate binding domain		var
thrombospondin repeat type 1		60	chondriotinsulfate binding domain		var
thrombospondin repeat type 3		60	S/T-rich (O-glycolysated?) domains	S	var
von Willebrand factor A rep.	VA	200	K/P rich repeats		var
von Willebrand factor B rep.		30	collagen like triple-helix		var
von Willebrand factor C rep.	VC	115	coiled coil region	CC	var
von Willebrand factor D rep.	VD	330	transmembrane region		var
fibrinogen B/tenascin segment	FB	250	cytosolic region	I	var
scavenger rec./factor I domain	R	110	?		var

Fig. 1. Domains considered in this study. They occur at least in two proteins with different domain assembly. Only the average length of a domain counted as amino acids (aa) is symbolized. Domains labeled as being extremely variable in length (var) need not be necessarily homologous to each other, but have often similar functions. White boxes stand for domains for which no homologous segments were found in proteins with different domain composition. Small spacer regions as well as signal peptides were neglected. The symbols used are the same as in the other figures.

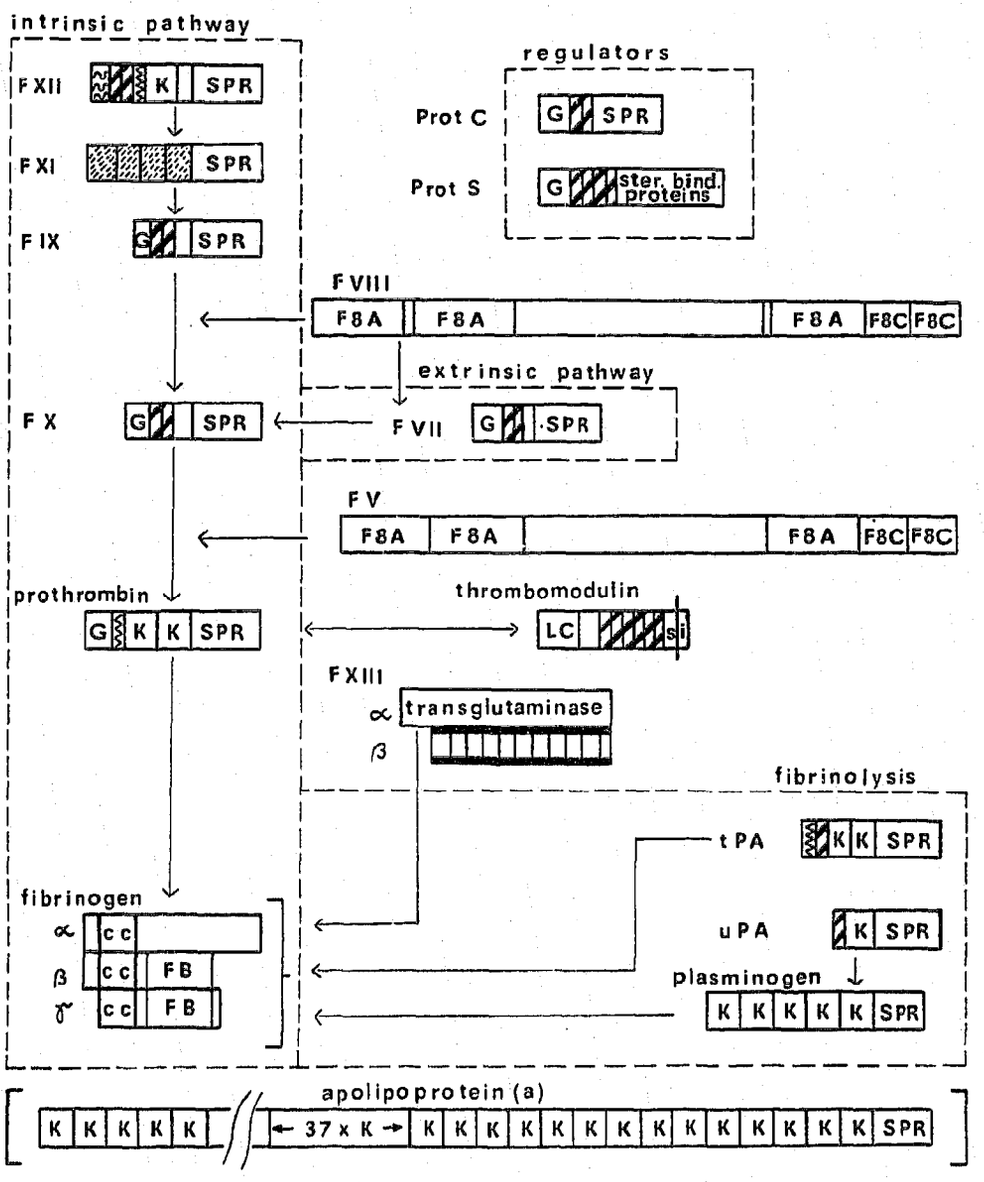

Fig. 2. Simplified flow chart of coagulation and fibrinolysis. Some regulatory proteins are also shown. Though the zymogen activation via serine proteases (SPR domains) is known for about 20 years, the molecular basis of the important regulatory mechanisms is not well understood. So far, defined functions could be related only to a few regulatory domains. For example, the N-terminal γ-carboxyglutamate domains (G) bind to membranes via calcium and the 'kringles' (K) are involved in fibrin binding (for reviews of domains in coagulation and references see e.g. [16–18]). Other domains like FB of fibrinogen which was recently also identified in invertebrates [19] seem to have more general regulatory functions. The triplicated type A domains in coagulation factors V and VIII can be also found in ceruloplasmin [20], but in coagulation the typical copper binding sites are lost. Apolipoprotein (a) of the low density lipoprotein (LDL) fraction was added to the proteins of the cascades because its domain assembly is similar to that of plasminogen [21]. This fact has led to experiments which have shown that apolipoprotein (a), a main risk factor in atherosclerosis, promotes coagulation [22]. The other component of LDL, apolipoprotein B100 was proposed to be distantly related to vitellogenin [23].

main family are known the lower is the degree of conserved features. For example, about 500 EGF (or so-called EGF-like) domains can be found in the protein sequence databases (only some of them are shown in Figs. 2–6), but on aligning them not a single disulfide bridge remains absolutely conserved and the number of amino acids between them can vary considerably. Thus several groups have proposed classifications according to different structural or functional features [9–12] and the overall function of the domain can only be stated as involvement in growth and differentiation of cells.

It is impossible to discuss all the domains or proteins shown in Figs. 1–6 in detail, so that the attention should be focussed on some more general points. For example, (i) many of the receptors as shown in Fig. 4 are involved in the transfer of important signals. These processes are highly regulated by a lot of different domains located in the extracellular parts of the receptors.

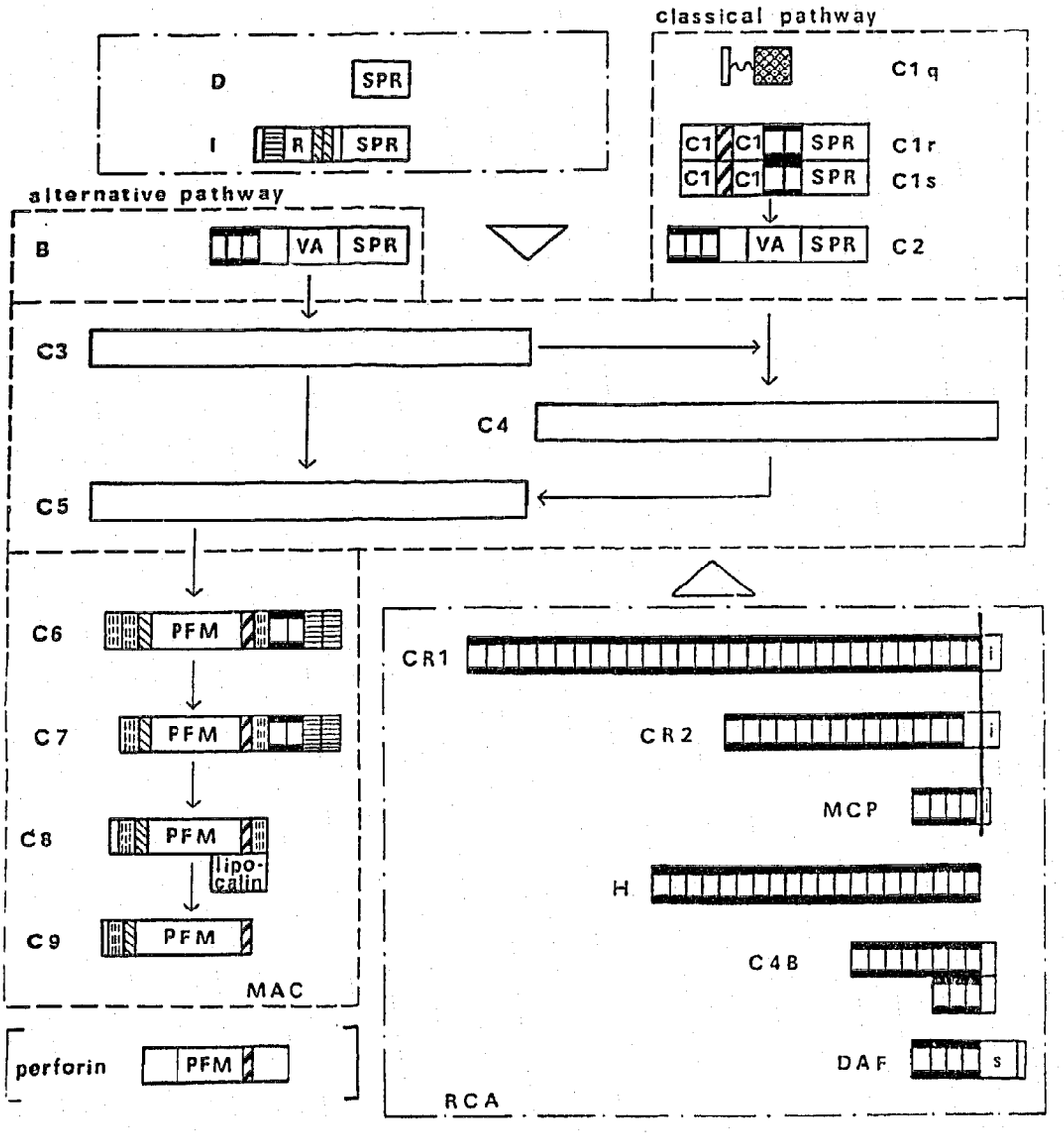

Fig. 3. Simplified flow chart of the cytolytic complement system, which plays an important role in inflammation and in host defense against infections. The negative regulators of complement activation (RCA) as well as the regulatory factors D and I are also shown (most of the proteins and domains are reviewed in [24,25]). Another cytolytic protein, perforin, is arranged below the terminal components of the membrane attachment complex (MAC) because of the similar (catalytic?) central domains [26]. Although C3, C4 and C5 playing a central role in complement control, become cleaved into active fragments, at present none of the liberated segments could be found in proteins with different domain assembly. Instead there are mutual homologies as well as sequence similarities to pregnancy zone protein and to α_2-macroglobulins [27].

All of these receptors have only one transmembrane region (excluding the signal peptide), which is independent of anchoring (i.e. of whether N- or C-terminals of the molecules are located in the cytosol). (ii) Another point of surprise is the occurrence of the immunoglobulin-like domains of the C2 subfamily [13] in mosaic proteins, suggesting that immunoglobulins are subject to the same evolutionary mechanisms as the other involved domains. (iii) Triple-helix forming, glycine-rich segments as in the collagens are associated with the most different globular domains and may be shuffled themselves. This points to a defined molecular tuning of mechanical functions.

3. A FUNCTIONAL NETWORK

Many extracellular enzymes get support (in transport, binding features, protection, regulation, etc.) from shuffled domains. At present various self-contained proteins are known to be surrounded by regulatory domains (or rather they are themselves domains). Examples are the well-known serine proteases of cascades, but also the biotin binding protein avidin, Kunitz type inhibitors, type IV collagenase and thyroid peroxidase. Sometimes subunits of completely different evolutionary origin are associated and work together as seen in coagulation factor XIII (the α subunit contains

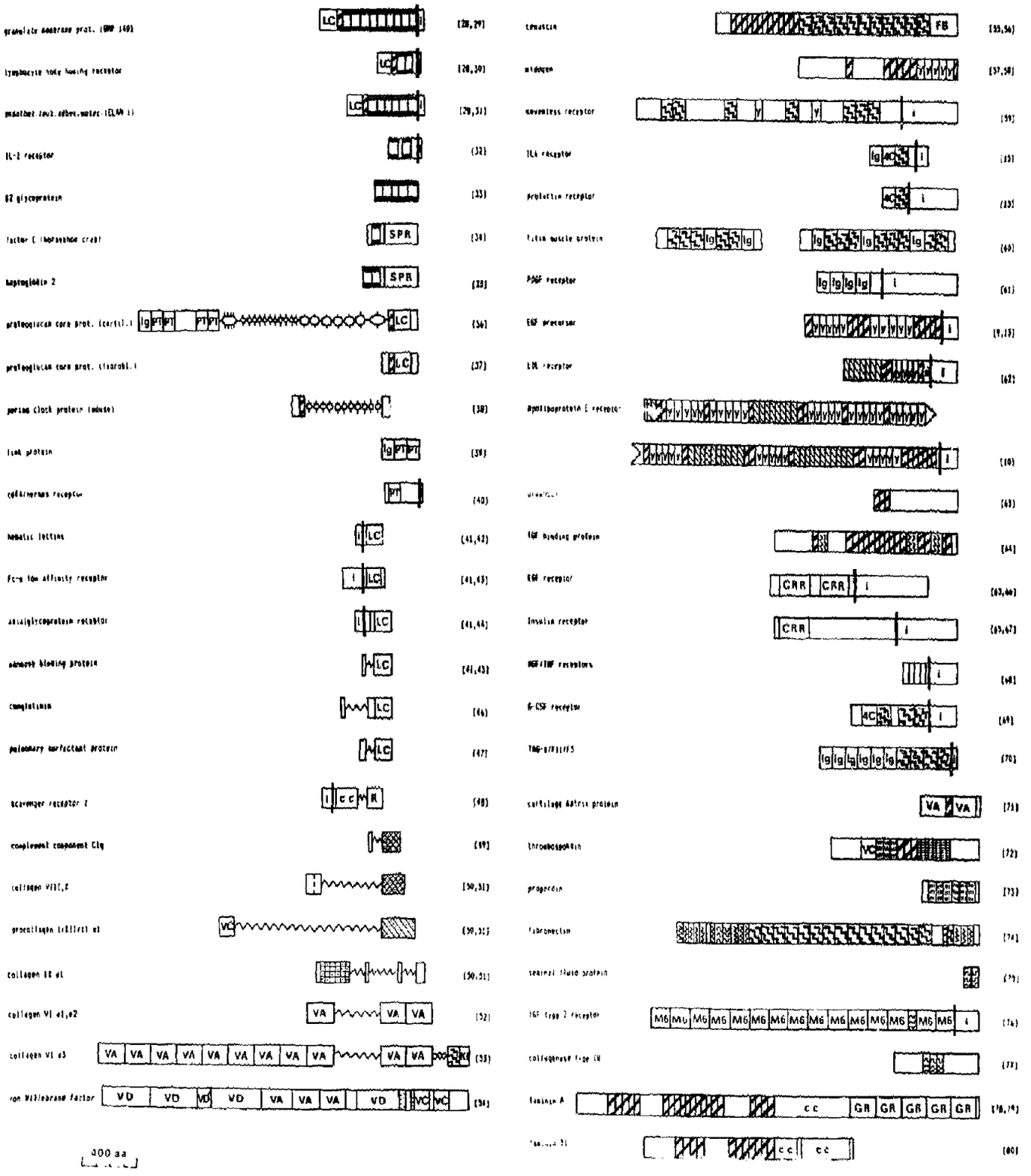

Fig. 4. A collection of proteins involved in adhesive processes. Should different splicing patterns be possible, then the longest one of the resulting sequences is shown. If there exist different proteins having the same domain composition only one representative is shown. Some of the mosaic proteins are neglected because of space limitations. For example, the VA domain was also found in some integrin α subunits (the so-called I-domain, for recent review see [81]), a domain common in scavenger receptor as well as in complement factor I could be recently detected in some more proteins [82] and a repeated segment of complement components C1r/C1s was recognized in different developmental regulated proteins [83].

a number of SCRs, whereas the β subunit is a transglutaminase) or complement component C8 (the α and β chains are similar to other proteins of membrane attachment complex, but the gamma subunit is related to the lipocalins, Fig. 3).

Even if the functions of many extracellular proteins are not well understood and only a few 3D structures of extracellular domains are available [14], it becomes clear that 'exon shuffling' was not only useful for increasing the speed of protein modification in evolution,

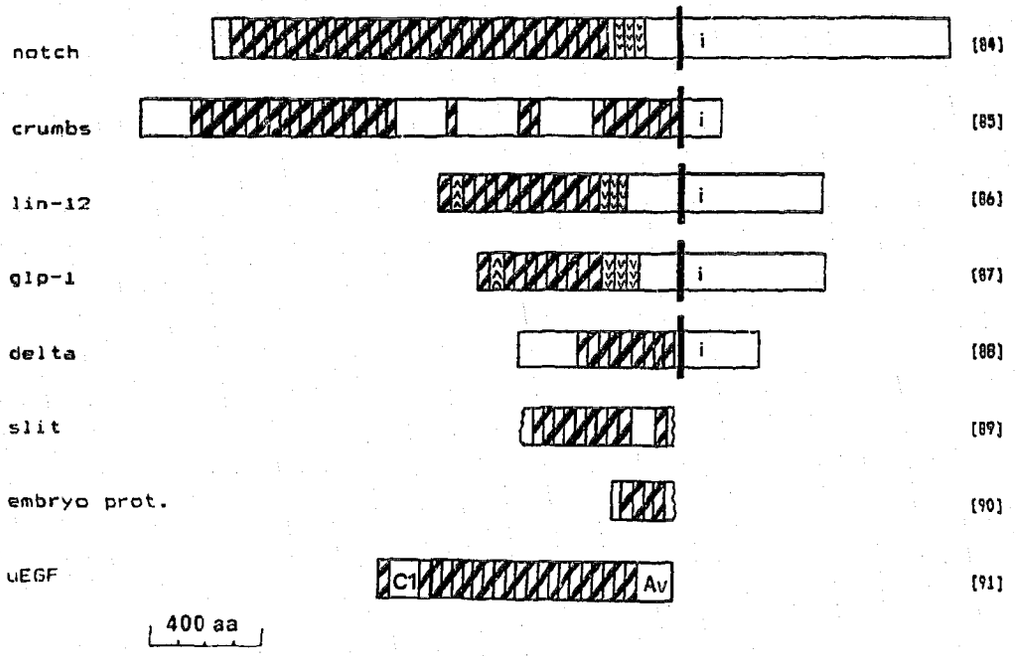

Fig. 5. A sample of invertebrate proteins containing EGF-like domains. An interesting feature of these proteins is, that the EGF-like domain often occurs in multicopies without interrupting introns. In contrast, the exon borders of sea urchin uEGF are located between the EGF domains as in vertebrates [92].

but it also supplied the molecular basis for a complex functional network allowing multiple regulation in and between nearly all tissues. It combines the most different functional complexes like, for example, the antibody framework, inflammation processes or the haemopoietic system, where, in turn, a number of involved domains could be recently classified [15]. Different splicing patterns can lead to the coexistence of protein variants that differ in length and lack some domains (mostly those of multicopies, i.e. where the same domain occurs many times within one molecule). The functional reasons for such multicopies (spacer function or increasing binding affinity?) remains to be solved. A comparative analysis of domains and their location within the proteins should prove valuable in clarifying functional aspects as well.

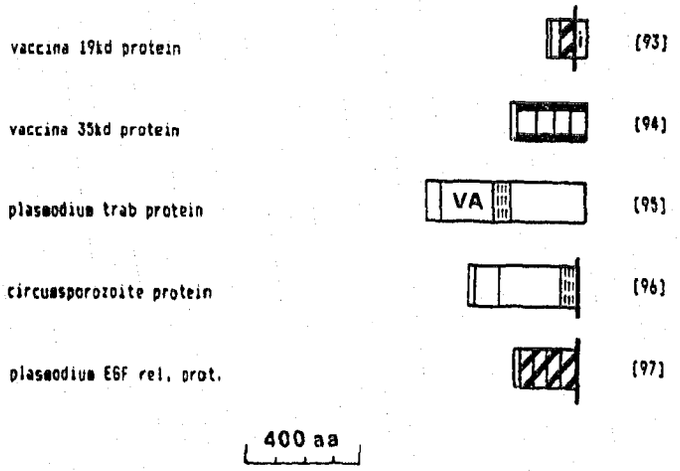

Fig. 6. Representation of vaccinia virus and malaria parasite proteins containing domains like those considered in this study. The limited number of recognition domains used by the various receptors and regulatory proteins of the host offers a chance for viruses and parasites to mimic ligands starting their infiltration process. Often viruses also stimulate the proliferation of neighbouring cells to support their own replication. Thus many of the oncogenes of retroviruses code for receptor tyrosine kinases or other growth factor receptors that are very similar to those of the host excepting parts of the extracellular domains [65]. Shown are only some vaccinia virus and malaria parasite proteins, for which no equivalent protein is known in the host.

Acknowledgements: The author is indebted to T.A. Rapoport, J.G. Reich and C. Sander for helpful suggestions as well as critical reading of the manuscript. I thank E. Wolf and G. Freudenberg for technical assistance.

REFERENCES

[1] Gilbert, W. (1978) Nature 271, 501.
[2] Doolittle, R.F. (1985) Trends Biochem. Sci. 10, 233–237.
[3] Patthy, L. (1988) J. Mol. Biol. 202, 689–696.
[4] Patthy, L. (1987) FEBS Lett. 214, 1–7.
[5] Dorit, R.L., Schoenbach, L. and Gilbert, W. (1990) Science 250, 1377–1382.
[6] Doolittle, R.F. (1985) Sci. Am. 253, 78–83.
[7] Bork, P. and Grunwald, C. (1990) Eur. J. Biochem. 191, 347–358.
[8] Bork, P. and Rohde, K. (1990) Biochem. Biophys. Res. Commun. 171, 1319–1325.
[9] Doolittle, R.F., Feng, D.F. and Johnson, M.S. (1984) Nature 307, 558–560.
[10] Herz, J., Hamann, U., Rogne, S., Myklebost, O., Gausepohl, H. and Stanley, K.K. (1988) EMBO J. 7, 4119–4127.

[11] Rees, D.J.G., Jones, I.M., Handford, P.A., Walter, S.J., Esnouf, M.P., Smith, K.J. and Brownlee, G.G. (1988) EMBO J. 7, 2053–2061.
[12] Apella, E., Weber, I.T. and Blasi, F. (1988) FEBS Lett. 231, 1–4.
[13] Williams, A.F. and Barclay, A.N. (1988) Annu. Rev. Immunol. 6, 381–405.
[14] Barton, M., Norman, D.G. and Campbell, I.D. (1991) Trends Biochem. Sci. 16, 13–17.
[15] Bazan, J.F. (1990) Immunol. Today 11, 350–354.
[16] Patthy, L. (1985) Cell 41, 657–663.
[17] Blake, C.C.F., Harlos, K. and Holland, S.K. (1987) Cold Spring Harb. Symp. Quant. Biol. 52, 925–931.
[18] Furie, B. and Furie, B.C. (1988) Cell 53, 505–518.
[19] Baker, N.E., Mlodzik, M. and Rubin, G.M. (1990) Science 250, 1370–1377.
[20] Church, W.R., Jernigan, R.L., Toole, J., Hewick, R.M., Knopf, J., Knutson, G.J., Meshein, M.E., Mann, K.G. and Fass, D.N. (1984) Proc. Natl. Acad. Sci. USA 81, 6934–6937.
[21] McLean, J.W., Tomlinson, J.E., Kuang, W.-J., Eaton, D.L., Chen, E.Y., Fless, G.M., Scanu, A.M. and Lawn, R.M. (1987) Nature 330, 132–137.
[22] Scott, J. (1989) Nature 341, 22–23.
[23] Baker, M.E. (1988) Biochem. J. 255, 1057–1060.
[24] Reid, K.B.M. and Day, A.J. (1989) Immunol. Today 10, 177–180.
[25] Müller-Eberhardt, H.J. (1988) Annu. Rev. Biochem. 57, 321–347.
[26] Stanley, K. and Luzio, P. (1988) Nature 334, 475–476.
[27] Sottrup-Jensen, L. (1989) J. Biol. Chem. 264, 11539–11542.
[28] Johnston, G.I., Cook, R.G. and McEver, R.P. (1989) Cell 56, 1033–1044.
[29] Springer, T.A. (1990) Nature 346, 425–434.
[30] Siegelman, M.H., Van de Rijn, M. and Weissmann, J.L. (1989) Science 243, 1165–1172.
[31] Bevilacqua, M.P., Stengelin, S., Gimlrone Jr., M.A. and Seed, P. (1989) Science 243, 1160–1165.
[32] Leonard, W.J., Depper, J.H., Kanesiha, M., Krönke, M., Pfeffer, N.J., Svedlik, P.B., Sullivan, M. and Greene, W.C. (1985) Science 230, 633–639.
[33] Lozier, J., Takahashi, N. and Putnam, F.W. (1984) Proc. Natl. Acad. Sci. USA 81, 3640–3644.
[34] Tokunaga, F., Miyata, T., Nakamura, T., Morita, T., Kuma, K., Miyata, T. and Iwanaga, S. (1987) Eur. J. Biochem. 167, 405–416.
[35] Kurosky, A., Barnett, D.R., Lee, T.-H., Touchstone, B., Hay, R.G., Arnott, M.S., Bowman, B.H. and Fitch, W. (1980) Proc. Natl. Acad. Sci. USA 77, 3388–3392.
[36] Doege, K., Sasaki, M., Horigan, E., Hassel, J. and Yamada, Y. (1988) J. Biol. Chem. 262, 17757–17767.
[37] Krusius, T., Gehlsen, K.R. and Ruoslahti, E. (1987) J. Biol. Chem. 262, 13121–13125.
[38] Shin, H.S., Bargiello, T.A., Clark, B.T., Jackson, F.R. and Young, M.W. (1986) Nature 317, 445–447.
[39] Bonnet, F., Perin, J., Lorenzo, F., Jolles, J. and Jolles, P. (1986) Biochim. Biophys. Acta 873, 152–155.
[40] Goldstein, L.A., Zhou, D.F.H., Picker, L.J., Minty, C.N., Bargatze, R.F., Ding, J.F. and Butcher, E.C. (1989) Cell 56, 1063–1072.
[41] Drickamer, K. (1988) J. Biol. Chem. 263, 9557–9560.
[42] Drickamer, K., Mamon, J.F., Binns, G. and Leung, J.O. (1984) J. Biol. Chem. 259, 770–778.
[43] Kikutani, H., Inui, S., Sato, R., Barsumian, E.L., Owaki, H., Yamasaki, K., Kaisho, T., Uchiboyashi, N., Hardy, R.R., Hirano, T., Tsunasawa, S., Sakiyama, F., Suemura, M. and Kishimoto, T. (1986) Cell 47, 657–665.
[44] Drickamer, K. and McCreavy, V. J. Biol. Chem. 262, 2582–2589.
[45] Drickamer, K., Dordal, M.S. and Reynold, L. (1986) J. Biol. Chem. 261, 6878–6887.
[46] Young, M.N. and Leon, M.A. (1987) Biochem. Biophys. Res. Commun. 143, 645–651.
[47] Patthy, L. (1987) Nature 325, 490.
[48] Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P. and Krieger, M. (1990) Nature 343, 531–535.
[49] Reid, K.B.M. and Day, A.J. (1990) Immunol. Today 11, 387–388.
[50] Kornblihtt, A.R. and Gutman, A. (1988) Biol. Rev. Cambr. Phil. Soc. 63, 465–507.
[51] Vuorio, E. and De Crombrugghe, B. (1990) Annu. Rev. Biochem. 59, 837–872.
[52] Chu, M.-L., Pan, T., Conway, D., Kuo, H.-J., Glanville, R.W., Timpl, R., Mann, K. and Deutzmann, R. (1989) EMBO J. 8, 1839–1846.
[53] Chu, M.-L., Zang, R.-Z., Pan, T., Stokes, D., Conway, D., Kuo, H.-J., Glanville, R.W., Mayer, U., Mann, K., Deutzmann, R. and Timpl, R. (1990) EMBO J. 9, 385–393.
[54] Titani, K. and Walsh, K.A. (1988) Trends Biochem. Sci. 13, 94–97.
[55] Jones, F.S., Hoffmann, S., Cunningham, B.A. and Edelman, G.M. (1989) Proc. Natl. Acad. Sci. USA 86, 1905–1909.
[56] Chiquet-Ehrismann, R. (1990) FASEB J. 4, 2598–2604.
[57] Mann, K., Deutzmann, R., Aumailley, M., Timpl, R., Raimondi, L., Yamada, Y., Pan, T., Conway, D. and Chu, L.-M. (1989) EMBO J. 8, 65–72.
[58] Engel, J. (1989) FEBS Lett. 251, 1–7.
[59] Norton, P.A., Hynes, R.O. and Rees, D.J.G. (1990) Cell 61, 15–16.
[60] Labeit, S., Barlow, D.P., Gautel, M., Gibson, T., Holt, J., Hsieh, C.-L., Francke, U., Leonard, K., Wardale, J., Whiting, A. and Trinick, J. (1990) Nature 345, 273–276.
[61] Claesson-Welsh, L., Eriksson, A., Westermark, B. and Heldin, C.-H. (1989) Proc. Natl. Acad. Sci. USA 86, 4917–4921.
[62] Südhoff (1985) Science 228, 815–822.
[63] Pennica, D., Kohr, W.J., Kuang, W.J., Glaister, D., Aggarwal, B.B., Chen, E.J. and Goeddel, D.V. (1987) Science 236, 83–88.
[64] Kanzaki, T., Olofsson, A., Moren, A., Wernstedt, C., Hellman, U., Miyazono, K., Claesson-Welsh, L. and Heldin, C.-H. (1990) Cell 61, 1051–1061.
[65] Yarden, Y. and Ullrich, A. (1988) Annu. Rev. Biochem. 57, 443–478.
[66] Pfeffer, S. and Ullrich, A. (1985) Nature 313, 184.
[67] Ullrich, A., Bell, J.R., Chen, E.-Y., Herrera, R., Petruzzelli, L.M., Dull, T.J., Gray, A., Coussens, L., Liao, Y.C., Tsubokawa, M., Mason, A., Seeburg, P.H., Grunfeld, C., Rosen, O.M. and Ramachandran, J. (1985) Nature 313, 756–761.
[68] Sprang, S.R. (1990) Trends Biochem. Sci. 15, 366–368.
[69] Fukunaga, R., Ishizaka-Ikeda, E., Seto, Y. and Nagata, Y. (1990) Cell 61, 341–350.
[70] Furley, A.J., Morton, S., Manalo, D., Laragogeos, D., Dodd, J. and Jessell, T.M. (1990) Cell 61, 157–170.
[71] Kiss, I., Deak, F., Holloway Jr., R.G., Delius, H., Mebust, K.A., Frimberger, E., Argraves, W.S., Tsonis, P.A., Winterbottom, N. and Goetnick, P.F. (1989) J. Biol. Chem. 264, 8126–8134.
[72] Lawler, J. and Hynes, R.O. (1986) J. Cell. Biol. 103, 1635–1648.
[73] Goundis, D. and Reid, K.B.M. (1988) Nature 335, 82–84.
[74] Kornblihtt, A.R., Umezawa, K., Vibe-Petersen, K. and Baralle, F.E. (1986) EMBO J. 4, 1755–1759.
[75] Seidah, N.G., Manjunath, P., Rochemont, J., Sairam, M.R. and Chretien, M. (1987) Biochem. J. 243, 195–203.
[76] Lobel, P., Dahms, N.M. and Kornfeld, S. (1988) J. Biol. Chem. 263, 2563–2570.
[77] Collier, I.E., Wilhelm, S.M., Eisen, A.Z., Marmer, B.L., Grant, G.A., Seltzer, J.L., Kronberger, A., He, C., Bauer, E.A. and Goldberg, G.I. (1988) J. Biol. Chem. 263, 6579–6587.
[78] Sasaki, M., Kleinman, H.K., Huber, H., Deutzmann, R. and Yamada, Y. (1988) J. Biol. Chem. 263, 16536–16544.

[79] Ehrig, K., Leivo, I., Argraves, W.S., Ruoslahti, E. and Engvall, E. (1990) Proc. Natl. Acad. Sci. USA 87, 3264–3268.
[80] Vuolteenaho, R., Chow, L.T. and Tryggvason, K. (1990) J. Biol. Chem. 265, 15611–15616.
[81] Larson, S.L. and Springer, T.A. (1990) Immunol. Rev. 114, 181–217.
[82] Freeman, M., Ashkenas, J., Rees, D.J.G., Kingsley, D.M., Copeland, N.G., Jenkins, N.A. and Krieger, M. (1990) Proc. Natl. Acad. Sci. USA 87, 8810–8814.
[83] Bork, P. (1991) FEBS Lett. (1991) 282, 9–12.
[84] Wharton, K.A., Yedvobnick, B., Finnerty, V.G. and Artavanis-Tsakonas, S. (1985) Cell 43, 567–581.
[85] Tepass, U., Theres, C. and Knust, E. (1990) Cell 61, 787–799.
[86] Greenwald, I. (1985) Cell 43, 583–590.
[87] *Yochem, J. and Greenwald, I. (1989) Cell 58, 553–563.*
[88] Vässin, H., Bremer, K.A., Knust, E. and Campos-Ortega, J.A. (1988) EMBO J. 6, 3431–3440.
[89] Rothberg, J.M., Hartley, D.A., Walter, Z. and Artavanis-Tsakonas, S. (1988) Cell 55, 1047–1059.
[90] Yang, Q., Angerer, L.M. and Angerer, R.C. (1989) Science 246, 806–808.
[91] Hursh, D.A., Andrews, M.E. and Raff, R.A. (1987) Science 237, 1487–1490.
[92] Delgadillo-Reynoso, M.G., Rollo, D.R., Hursh, D.A. and Raff, R.A. (1989) J. Mol. Evol. 29, 314–327.
[93] Brown, J.P., Twardzik, D.R., Marquard, H. and Todaro, G.J. (1985) Nature 313, 491–492.
[94] Kotwal, G.J. and Moss, B. (1988) Nature 335, 176–178.
[95] Robson, K.J.H., Hall, J.R.S., Jennings, M.W., Harris, T.J.R., Marsh, K., Tate, V.E. and Weatherhall, D.J. (1988) Nature 335, 79–82.
[96] Rich, K.A., George IV, F.W., Law, J.L. and Martin, W.J. (1990) Science 249, 1574–1577.
[97] Kaslow, D.C., Quaki, I.A., Syin, C., Raum, M.G., Keister, D.B., Coligan, J.E., McCutchan, T.F. and Miller, L.H. (1988) Nature 333, 74–76.