Polgar (1991) has demonstrated that the catalytic mechanism is significantly different from those of the serine endopeptidases of the chymotrypsin and subtilisin families. Because EC 3.4.21.26 hydrolyses oligopeptides but not proteins, unlike the prolyl endopeptidases from various species of bacteria that degrade immunoglobulin A, we propose that it should be renamed prolyl oligopeptidase.

It now appears that prolyl oligopeptidase can be placed in an evolutionary family of serine-type peptidases. The sequence of prolyl oligopeptidase shows a moderate similarity to that of dipeptidyl-peptidase IV (EC 3.4.14.5), but there is a still clearer relationship to acylaminoacyl-peptidase (EC 3.4.19.1). The strongest resemblance between the sequences is seen in the *C*-terminal 250 residues or so, which contain the identified catalytic residues of prolyl oligopeptidase (Fig. 1).

The statistical significance of the relationships between prolyl oligopeptidase and the other enzymes was tested with the RDF2 program from the FASTA package of Pearson & Lipman (1988). The KTUP value was set to 1, and 100 random uniform shuffles of the test sequences were performed. Values of six standard deviation units or more are considered significant in this test, and the following sequence pairs met this requirement (standard deviation values in parentheses): dipeptidyl-peptidase IV and yeast dipeptidyl-peptidase B (80.34); prolyl oligopeptidase and acylaminoacyl-peptidase (8.73); and dipeptidyl-peptidase B and acylaminoacyl-peptidase (8.21). It therefore follows that all four enzymes are members of a single homologous family of proteins.

The sequence of pig acylaminoacyl-peptidase (Mitta *et al.*, 1989) is closely similar to that of rat and it has been reported that the (unpublished) sequence of yeast dipeptidyl-peptidase A is related to that of dipeptidyl-peptidase B (Roberts *et al.*, 1989). Another homologous sequence is that of the human protein 3p21. The gene for 3p21 is located on the short arm of human chromosome 3, deletion of which is associated with small cell lung cancer (Naylor *et al.*, 1989). Protein 3p21 shows marked sequence similarity to the acylaminoacyl-peptidases; it is shorter at both *N*- and *C*-termini (Jones *et al.*, 1991), but contains the putative catalytic site residues we identify here. These proteins can therefore be considered additional members of the new family.

By analogy with prolyl oligopeptidase, we would predict that the catalytic residues in the other enzymes are the aligned Ser^{587} and His^{707} (acylaminoacyl-peptidase), Ser^{631} and His^{741} (rat dipeptidyl-peptidase IV), and Ser^{678} and His^{788} (yeast dipeptidyl-peptidase B). Other known serine-dependent peptidases contain a 'catalytic triad' of essential residues: serine, histidine and aspartic acid. There is no direct evidence as to the identity of any essential aspartic acid residue in the prolyl oligopeptidase family, but candidates are the conserved Asp^{529} and Asp^{642} residues (or Asp^{641}, in a slightly shifted alignment).

Like prolyl oligopeptidase, dipeptidyl-peptidase IV and acylaminoacyl-peptidase are inhibited by di-isopropylfluorophosphate, and acylaminoacyl-peptidase also is inactivated by a chloromethane (Kobayashi & Smith, 1987). The other enzymes do not show the marked thiol-dependence of prolyl oligopeptidase, however.

From the point of view of the catalytic activities of the component enzymes, this is one of the most disparate families of peptidases so far recognized, containing an endopeptidase that is confined to action on oligopeptides (prolyl oligopeptidase), an exopeptidase that cleaves dipeptides from the *N*-termini of polypeptides only when the *N*-terminus is free (dipeptidyl-peptidase IV), and an omega peptidase that preferentially cleaves *N*-terminal acetyl-aminoacyl residues from polypeptides (acylaminoacyl-peptidase). The enzymes therefore fall into three distinct sub-subsections of the Enzyme Nomenclature list (IUB, 1984). Prolyl oligopeptidase and dipeptidyl-peptidase IV show marked preference for the cleavage of prolyl bonds, but no such specificity is seen with acylaminoacyl-peptidase (McDonald & Barrett, 1986). Moreover, prolyl oligopeptidase and acylaminoacyl-peptidase are soluble, cytosolic enzymes, whereas dipeptidyl-peptidase IV is a heavily glycosylated enzyme that is predominantly membrane-associated in both rat and yeast, containing a transmembrane segment near the *N*-terminus (Ogata *et al.*, 1989).

Neil D. RAWLINGS,* László POLGAR† and
Alan J. BARRETT*

*Department of Biochemistry, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, U.K., and †Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, P.O. Box 7, H-1502 Budapest, Hungary

IUB (1984) Enzyme Nomenclature 1984: Recommendations of the Nomenclature Committee of the International Union of Biochemistry on the Nomenclature and Classification of Enzyme-Catalysed Reactions, Academic Press, New York
Jones, W. M., Scaloni, A., Bossa, F., Popowicz, A. M., Schneewind, O. & Manning, J. M. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 2194–2198
Kobayashi, K., Lin, L.-W., Yeadon, J. E., Klickstein, L. B. & Smith, J. A. (1989) J. Biol. Chem. **264**, 8892–8899
Kobayashi, K. & Smith, J. A. (1987) J. Biol. Chem. **262**, 11435–11445
McDonald, J. K. & Barrett, A. J. (1986) Mammalian Proteases: a Glossary and Bibliography. Volume 2: Exopeptidases, Academic Press, London
Mitta, M., Asada, K., Uchimura, Y., Kimizuka, F., Kato, I., Sakiyama, F. & Tsunasawa, S. (1989) J. Biochem. (Tokyo) **106**, 548–551
Naylor, S. L., Marshall, A., Hensel, C., Martinez, P. F., Holley, B. & Sakaguchi, A. Y. (1989) Genomics **4**, 355–361
Ogata, S., Misumi, Y. & Ikehara, Y. (1989) J. Biol. Chem. **264**, 3596–3601
Pearson, W. R. & Lipman, D. J. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2444–2448
Polgar, L. (1991) Eur. J. Biochem. **197**, 441–447
Rennex, D., Hemmings, B. A., Hofsteenge, J. & Stone, S. R. (1991) Biochemistry **30**, 2195–2203
Roberts, C. J., Pohlig, G., Rothman, J. H. & Stevens, T. H. (1989) J. Cell Biol. **108**, 1363–1373
Stone, S. R., Rennex, D., Wikstrom, P., Shaw, E. & Hofsteenge, J. (1991) Biochem. J. **276**, 837–840

Received 9 July 1991

More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-α-trypsin inhibitor

von Willebrand factor is an adhesive glycoprotein involved in various binding interactions (for recent review see [1]). The subunits of this oligomeric complex are large in size and have a typical mosaic structure containing a number of repeated domains. As a result of the 'exon shuffling' process [2] such domains may occur in apparently unrelated proteins involved in very different extracellular pathways [3–6]. Indeed, different types of domains observed in von Willebrand factor were also

identified in other proteins. In particular, the triplicated type A domain (VWA) was found to be widespread in adhesive proteins and receptors. VWA could be detected in the complement components B and C2 [7], twice in cartilage matrix protein [8], in some integrin α subunits (the so called I-domains; for review see e.g. [9]) and in the different chains of collagen VI [10–12] in up to 11 copies (α_3 chain).

Segments corresponding to the VWA domain in the different proteins share only a limited sequence similarity; no amino acid is completely conserved (Fig. 1). The pairwise amino acid identities remain within the so-called twilight zone [13], where current alignment procedures fail. But low-level sequence resemblance becomes more significant as more members are added to the alignment [14]. The multiple alignment carried out by PULIGN [15] revealed some reliable regions. The most stringent motifs (1–2–3 and 6–7 in Fig. 1) were selected for a property pattern search [16] to identify distant related domains in other extracellular proteins as described in [17]. Screening of MIPSX (MIPSX is a merged database, containing 32800 sequences in release 17) indicates exactly four candidates which appear clearly separated from a random background (Fig. 2). The findings are corroborated by the fact that they also match the remaining motifs (Fig. 1).

The presence of a VWA domain in thrombospondin-related anonymous protein (TRAP) from the malaria parasite *Plasmodium falciparium* [18] is suggested by the known mosaic structure of that protein and by the exact location of the domain next to an identified thrombospondin type 1 repeat, which, in

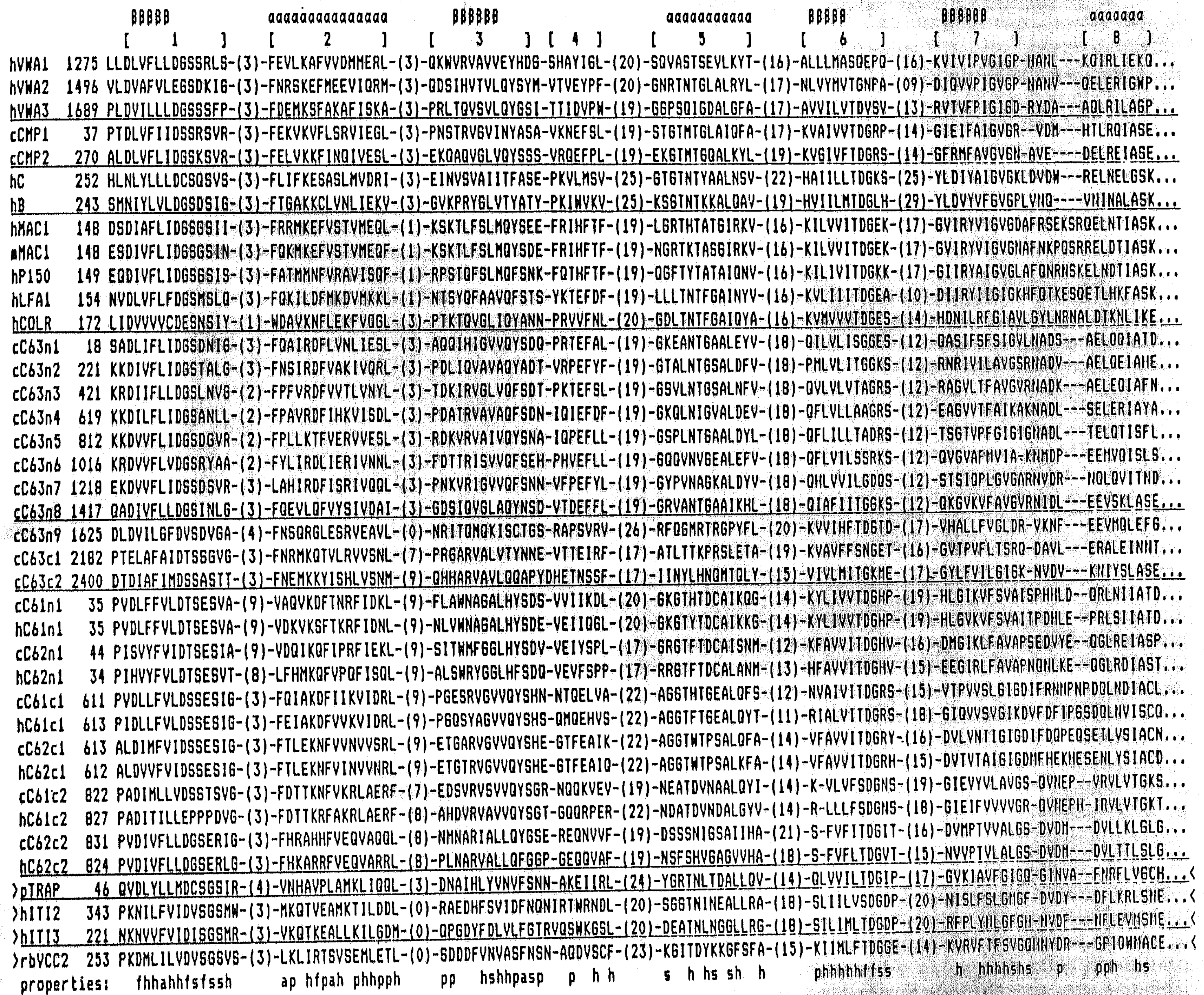

Fig. 1. Multiple alignment of the VWA family

Sequences that show higher similarities to one another are grouped between vertical lines. Only the most conserved regions are shown. These motifs are numbered. Predicted secondary structures (α, β) within the motifs correspond to typical hydrophobicity patterns. The number of amino acids between the motifs is indicated. The suggested new members are marked by arrows. The consensus line summarizes characteristic features of conserved positions: hydrophobic (h), polar (p), small (s), aromatic (a), extremely conserved and suggested to be involved in function (f). Sequences were taken from MIPSX: hVWA, human von Willebrand factor; cCMP, chicken cartilage matrix protein; hC and hB, human complement factors C and B; hMAC1 (mMAC1), α chain of human (mouse) leukocyte adhesion molecule MAC-1; hP150, α chain of human integrin p150,95; hLFA1, α chain of human integrin LFA-1; hCOLR, α_2 chain of collagen receptor; cC63, chicken collagen VI α_3 chain; cC61 (cC62), chicken collagen VI α_1 (α_2) chain; hC61 (hC62), human collagen VI α_1 (α_2) chain; fTRAP, thrombospondin-related anonymous protein from *Plasmodium falciparium*; hITI2 (hITI3), human inter-α-trypsin inhibitor α_2 (α_3) subunit; rbVCC, rabbit voltage-dependent dihydropyridine-sensitive calcium channel α_2 subunit.

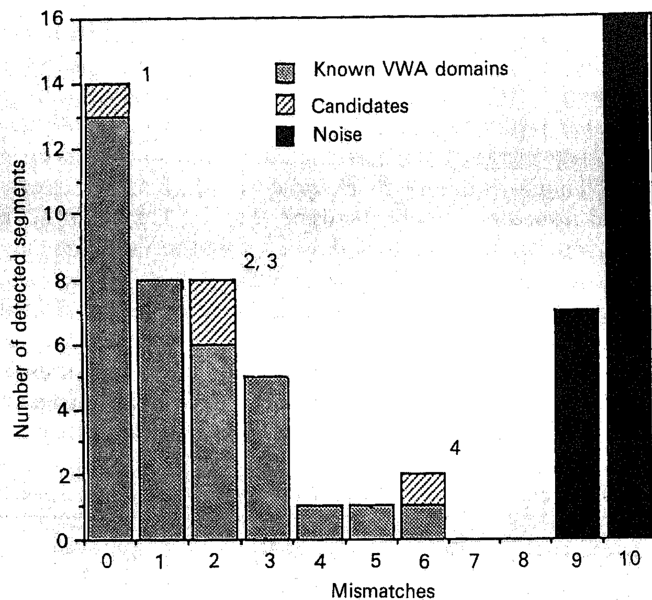

Fig. 2. Pattern-matching statistics

The property-pattern approach extracts from each position of the multiple alignment of known VWA domains a set of common steric and physicochemical amino acid properties (to each amino acid a defined set has been assigned [16]). A sequence database search has been carried out with the derived consensus property patterns. The y-axis represents the number of detected sequence segments matching both patterns (1–2–3 and 6–7), while the x-axis shows the respective number of mismatches. The number of mismatches corresponds to the number of amino acid properties deviating from the consensus property pattern. Sequences of the learning set fall into the range of up to six mismatches without any misclassification. Above the threshold of nine mismatches many misclassified (not related) sequences are recorded and considered as 'noise'. Since there is no overlap between the two curves (members of the learning set and 'noise') we regard the unknown sequences among the 'correct positives' as belonging to the family described by the consensus patterns. The four candidates for possible relationships: (1) the α_2 subunit of a dihydropyridine-sensitive voltage-dependent calcium channel [27], (2) and (3) the noninhibitory α_2 and α_3 subunits of a inter-α-trypsin Kunitz type inhibitor [24,29], and (4) thrombospondin-related anonymous protein [18].

turn, can be found in properdin [19], malaria circumsporozoite protein (e.g. [20]) and in some terminal complement components [21]. The biological function of TRAP is unknown, but the thrombospondin type 1 therein has been suggested to play a role in parasite–cell interactions [18]. A VWA domain could aid such a function by specific binding.

The physiological function of the whole inter-α-trypsin inhibitor complex (for recent reviews see [22,23]) and in particular its two noninhibitory subunits α_2 and α_3 (which share about 40% sequence identity) is uncertain [24]. The precursor of the precursor of the α_1 subunit containing the two catalytic Kunitz-type inhibitor domains has a typical mosaic structure and releases α_1-microglobulin after cleavage. Also, the precursors of the α_2 and α_3 subunits are cleaved, but the proposed VWA domains remain constituents of the mature inter-α-trypsin inhibitor. Kunitz type inhibitors seem generally to be involved in a network of shuffled domains as seen in the case of amyloid precursor protein (e.g. [25]). A Kunitz type inhibitor domain also coexists with VWA repeats in the α_3 chain of collagen VI [12]. Hence a VWA domain may recognize proteins which then complex with the inhibitor.

Voltage dependent dihydropyridine-sensitive calcium channels belong to a group of membrane-spanning proteins that allow the controlled entry of extracellular calcium into the cell. The α_2 subunit of the dihydropyridine-sensitive calcium channel complex does not act directly as a channel-forming protein [26]. Three putative transmembrane regions have been predicted, and the suggested VWA-like domain is located in the N-terminal extracellular part of the protein [27]. The α_2 subunit interacts with the other components of the complex and the VWA-like domain may also have a binding function. Interestingly, the sequence similarity to the proposed VWA unit in the α_2 subunit of inter-α-trypsin inhibitor is as high as 23% identical residues, only slightly below a value that suggests similar folding topology [28].

Even if experimental support is needed, the observed sequence similarities within conserved regions point to a structural homology and represent a further step in classification of domains in extracellular mosaic proteins. A comparative analysis of shuffled domains like VWA may also contribute to a better understanding of their functional relations.

Peer BORK and Klaus ROHDE

Central Institute of Molecular Biology,
Department of Biomathematics, Robert-Rössle-Strasse 10,
O-1115 Berlin-Buch, Federal Republic of Germany

1. Wagner, D. D. (1990) Annu. Rev. Cell. Biol. **6**, 217–246
2. Gilbert, W. (1978) Nature (London) **271**, 501
3. Doolittle, R. F. (1985) Trends Biochem. Sci. **10**, 233–237
4. Baron, M., Norman, D. G. & Campbell, I. D. (1991) Trends Biochem. Sci. **16**, 13–17
5. Patthy, L. (1991) Curr. Opin. Struct. Biol. **1**, 351–361
6. Bork, P. (1991) FEBS Lett. **286**, 47–54
7. Shelton-Inloes, B. B., Titani, K. & Sadler, J. E. (1986) Biochemistry **25**, 3164–3171
8. Kiss, I., Deak, F., Holloway, R. G., Jr., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N. & Goetnick, P. F. (1989) J. Biol. Chem. **264**, 8126–8134
9. Larson, S. L. & Springer, T. A. (1990) Immunol. Rev. **114**, 181–217
10. Koller, E., Winterhalter, H. & Trueb, B. (1989) EMBO J. **8**, 1073–1077
11. Chu, M.-L., Pan, T., Conway, D., Kuo, H.-J., Glanville, R. W., Timpl, R., Mann, K. & Deutzmann, R. (1989) EMBO J. **8**, 1939–1946
12. Chu, M.-L., Zhang, R.-Z., Pan, T., Stokes, D., Conway, D., Kuo, H.-J., Glanville, R. W., Mayer, U., Mann, K., Deutzmann, R. & Timpl, R. (1990) EMBO J. **9**, 385–393
13. Doolittle, R. F. (1985) Sci. Am. **253**, 74–83
14. Doolittle, R. F. (1989) in Prediction of Protein Structure and the Principles of Protein Conformation (Fasman, G. D., ed.), pp. 599–623, Plenum, New York
15. Bork, P. & Rohde, K. (1990) Biochem. Biophys. Res. Commun. **171**, 1316–1325
16. Bork, P. & Grunwald, C. (1990) Eur. J. Biochem. **191**, 347–358
17. Bork, P. (1991) FEBS Lett. **282**, 9–12
18. Robson, K. J. H., Hal, J. R. S., Jennings, M. W., Harris, T. J. R., Marsh, K., Tate, V. E. & Weatherall, D. J. (1988) Nature (London) **335**, 79–82
19. Goundis, D. & Reid, K. B. M. (1988) Nature (London) **335**, 82–84
20. Rich, K. A., George, F. W., IV, Law, J. L. & Martin, W. J. (1990) Science **249**, 1574–1577
21. Reid, K. B. M. & Day, A. J. (1989) Immunol. Today **10**, 177–180
22. Odum, L. (1990) Int. J. Biochem. **22**, 925–930
23. Salier, J.-P. (1990) Trends Biochem. Sci. **15**, 435–439
24. Gebhard, W., Schreitmüller, T., Hochstrasser, K. & Wachter, E. (1989) Eur. J. Biochem. **181**, 571–576
25. Tanzi, R. E., McClatchey, A. I., Lamperti, E. D., Villa-Komaroff, L., Gusella, J. F. & Neve, R. L. (1988) Nature (London) **331**, 528–530
26. Catterall, W. A. (1991) Cell **64**, 871–874
27. Ellis, S. B., Williams, M. E., Ways, N. R., Brenner, R., Sharp, A. H., Leung, A. T., Campell, K. P., McKenna, E., Koch, W. J., Hui, A., Schwartz, A. & Harpold, M. M. (1988) Science **241**, 1661–1664
28. Sander, C. & Schneider, R. (1991) Proteins **9**, 56–68
29. Gebhard, W., Schreitmüller, T., Hochstrasser, K. & Wachter, E. (1988) FEBS Lett. **229**, 63–67

Received 17 April 1991