Kringles: modules specialized for protein binding

Homology of the gelatin-binding region of fibronectin with the kringle structures of proteases

L. Patthy, M. Trexler, Z. Váli, L. Bányai and A. Váradi

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, XI. Karolina ut 29, 1113 Budapest, Hungary

Received 3 April 1984

Prothrombin, plasminogen, urokinase- and tissue-type plasminogen activators contain homologous structures known as kringles. The kringles correspond to autonomous structural and folding domains which mediate the binding of these multidomain proteins to other proteins. During evolution the different kringles retained the same gross architecture, the kringle-fold, yet diverged to bind different proteins. We show that the amino acid sequences of the type II structures of the gelatin-binding region of fibronectin are homologous with those of the protease-kringles. Prediction of secondary structures revealed a remarkable agreement in the positions of predicted β-sheets, suggesting that the folding of kringles and type II structures may also be similar. As a corollary of this finding, the disulphide-bridge pattern of type II structures is shown to be homologous to that in kringles. It is noteworthy that protease-kringles and fibronectin type II structures have similar functions inasmuch as they mediate the binding of multidomain proteins to other proteins. It is proposed that the kringles of proteases and type II structures of fibronectin evolved from a common ancestral protein binding module.

Prothrombin *Plasminogen* *Plasminogen activator* *Fibronectin* *Haptoglobin* *Kringle* *Protein evolution* *Protein folding*

1. INTRODUCTION

The amino-terminal non-protease extensions of prothrombin (PT), plasminogen (PL), urokinase- (u-PA) and tissue-type plasminogen activator (t-PA) contain homologous triple-loop, 3-disulphide-bridge kringle structures [1–5]. The kringles are essential for the biological specificity and control of the action of PT and PL because they are involved in interactions that regulate PT-thrombin conversion or control PL activation and plasmin-catalyzed fibrinolysis [6,7]. Thus the kringles participate in the association of factor V_a with PT [8] and in the binding of PL to fibrin, α_2-antiplasmin [9]. Although the exact function of the kringles of u-PAs and t-PAs is not yet known, their homology with the PT- and PL-kringles argues that they may also be involved in binding interactions essential for the regulation of PL activation.

Physicochemical studies and refolding experiments have shown that the kringles correspond to autonomous structural and folding domains of these proteins [10–13]. Kringles are also independent functional domains as evidenced by the observation that kringles isolated by limited proteolysis retain their original binding function [3,8,14–16]. Although the different kringles have the same triple-loop architecture their binding specificity diverged to bind different proteins or low-molecular-mass ligands [8,14–17]. The remarkable autonomy and functional variability of kringles suggests that these 'miniproteins' may serve as versatile protein-binding modules of different proteins.

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

Table 1

Alignment of the sequences of type II structures of bovine fibronectin (FN), kringles of human prothrombin (PT), human plasminogen (PL), human urokinase- and tissue-type plasminogen activators (u-PA and t-PA) and the α-chain of human haptoglobin (HP) using the one-letter code and the kringle-numbering of residues introduced previously [3,33]

(A)

```
FN 1                          TAVTQTYGGNSNGEPCVLPF
FN 2                          TVLVQTRGGNSNGALCHFPF
PT A    CAEGLGTNYRGNVSITRS          GIECQL WRSRYPHKPEI    NSTTHPGAD
PT S    CVPDRGQQYQGRLAVTTH          GLPC LAWASAQAKALSK    HQDF"SAVQ
PL 1    CKTGDGKNYRGTMSKTKN          GITCQ KWSSTSPHRPRF    SPATHPSEG
PL 2    CMHCSGENYDGKISKTMS          GLECQ AWDSQSPHAHGY    IPSKFPNKN
PL 3    CLKGTGENYRGNVAVTVS          GHTCQ HWSAQTPHTHNR    TPENFPCKN
PL 4    CYHGDGQSYRGTSSTTTT          GKKCQ SWSSMTPHRHQK    TPENYPNAG
PL 5    CMFGDGKGYRGKRATTVT          GTPCQ DWAAQEPHRHSIF   TPETNPRAG
u-PA    CYEGNGHFYRGK ASTDTM         GRPC LPWNSATVLQQTY    HAHRSDALQLG
t-PA 1  CYEDQGISYRGTWSTAES         GAEC TNWNSSALAQKPY    SGRRPDAIRLG
t-PA 2  CYFGNGSAYRGTHSLTES         GASC LPWNSMILIGKVY    TAQNPSAQALG
HP              VDSGNDVTDIA         DDGC PKPPE IAHGYVE HSVRYQC
        1       10                  20                30      40
```

(B)

```
FN 1    TYNGKTFYSC TTEGRQDGHLWCSTTSNYEQDQKY SFC TDH
FN 2    LYNNHNYTDC TSEGRRDNMKWCGTTQNYDADQKF GFC PM
PT A    LQE NF CRNPDSSITG PWCYTT DPTARR QEC STPVC
PT S    LVE NF CRNPDGDEEG VWCYVA GKPGDF GYC DLNYC
PL 1    LEE NY CRNPDNDPQG PWCYTT DPEKRY DYC DILEC
PL 2    LKK NY CRNPD RELRPWCFTT DPNKRW ELC DIPRC
PL 3    LDE NY CRNPDGKR A PWCHTT NSQVRW EYC KIPSC
PL 4    LTM NY CRNPDADK G PWCFTT DPSVRW EYC NLKKC
PL 5    LEK NY CRNPDGNVGG PWCYTT NPRKLY DYC DVPQC
u-PA    LGK HNY CRNPDNRR RPWCYVQVG LKPLV QEC MVHDC
t-PA 1  LGN HNY CRNPD RDSKPWCYVFKA GKYSS EFC STPAC
t-PA 2  LGK HNY CRNPDG DAKPWCHVLKN RRLTW EYC DVPSC
HP      KNYYKLR TEG DG V YTLNN EKQWTNKAVGDKLPEC
        50      60      70      79
```

The underlined segments are predicted as β-sheets by the method in [26,27]

We show here that type II structures [18,19] of the gelatin-binding region of fibronectin (FN) are also homologous with kringles, indicating that kringle-like protein-binding modules are not restricted to proteases.

2. MATERIALS AND METHODS

The amino acid sequences of human PT [2], human PL [3], human u-PA [4], human t-PA [5], bovine FN [18,19] and human HP [20] were taken from the literature. The optimum alignment of the sequences was found using a scoring system in which identities are equal to +10, identity of cysteines are equal to +20 and each gap was equal to −25 [21,22]. Terminal gaps were not included in this calculation. The similarity of sequences was determined as percent identity calculated from the aligned regions only. Similarity of the aligned segments was also determined using the log odds matrix for 250 PAMs [23]. Genealogy of sequences was constructed on the basis of percent difference and alignment scores [24], the position and number of gaps and the position of disulphide bridges [25].

The potentials for formation of α-helical regions, β-sheet regions and β-turns were calculated as in [26,27].

Human plasma FN was purified by affinity chromatography on gelatin–Sepharose [28] and the gelatin-binding fragment was digested with pepsin, trypsin and chymotrypsin and the peptides were isolated by gel filtration and high-voltage paper electrophoresis at pH 1.9 and 6.5. Following isolation and amino acid analysis, the peptides containing disulphide bridges were oxidized with performic acid [30] and the component peptides were isolated and sequenced. Details of the purification procedures and the sequence work will be described elsewhere.

3. RESULTS AND DISCUSSION

3.1. *Comparison of amino acid sequences*

The alignment of the sequences of the kringles of PT, PL, u-PA and t-PA, the type II structures of FN and the α-chain of HP is shown in table 1.

Since the protease-kringles compared differ in binding specificity but possess the same triple-loop structure the residues involved in the different binding functions may show great variability whereas the residues essential for the kringle-fold should be unchanged in all or most of the kringles. As pointed out earlier using a smaller data base [13], the residues conserved in protease-kringles are clustered around the Cys_{22}–Cys_{62} and Cys_{50}–Cys_{74} bridges, underlining the importance of these disulphide bonds in the kringle-fold. NMR studies on kringles have demonstrated an interaction between the conservative Leu_{45}, Trp_{25} and Trp_{61} residues in the kringle-fold [31] lending further support to the view that the conserved regions are important for the kringle-structure. In contrast with these structurally conserved regions, the segment containing residues 26–44 is highly variable, suggesting that this region is not part of the core of the kringle architecture. Studies on the binding sites of kringles have revealed that the residues determining the binding specificity of PL-kringle 4 are located at positions 56, 70 and 71 [32,33], i.e., in positions that are not part of the structurally conserved regions [13].

When the sequences of FN 1 and FN 2 (type II structures) are compared with the kringle sequences the average identity for the aligned regions is 22% for FN 1 and 24% for FN 2; the α-chain of HP, which was shown by others to be related to kringles [20] shows only an average of 17% identity with kringles. It is noteworthy that the regions of type II structures most similar to kringles coincide with the structurally conserved regions of kringles. Significantly, the 4 cysteines of type II structures are present in the same positions as in the kringle sequences.

The alignment of the sequences necessitated the introduction of several gaps. It was shown previously [25] that the position and number of gaps are useful traits to construct evolutionary trees, since changes in length of proteins are fixed less frequently than amino acid replacements. Moreover, it is a common observation that the position of gaps changes less frequently than the size of the gap, since insertions and deletions occur in regions which correspond to external loops which are more tolerant to alterations in size [34]. The homology of protease-kringles and type II structures of FN is also indicated by the fact that gaps introduced for the alignment of FN 1 and FN 2 with kringles coincide with gaps that are dictated by alignment of the different protease-

kringles and HP (table 1). With respect to gap events, the major difference between kringles and FN type II structures is the absence of residues 1–12, 26–44 and 77–79 from the latter.

3.2. *Comparison of predicted structures*

NMR studies have shown that the closely related protease-kringles have similar 3-dimensional structure, the kringle-fold [31,35], therefore prediction of secondary structures should reflect this similarity. By this criterion kringles do not appear to contain α-helices since helices were not predicted in homologous positions of the different kringles (not shown). Similarly, α-helices were not predicted in homologous positions of the closely related FN 1 and FN 2 structures. In agreement with this conclusion, physicochemical methods detected only β-structures but not α-helices in the gelatin-binding fragment of FN [36]; circular dichroism spectra of plasminogen (5 kringles accounting for about half of the molecule) also indicated the presence of β-structure but no significant amount of α-helix [37].

As shown in table 1, β-sheets are predicted for all kringles and both FN type II structures around Cys_{22}, Cys_{62} and Cys_{74}, i.e., around 3 of the 4 cysteines common to kringles and FN 1 and FN 2. β-turns are predicted for the majority of kringles at positions 9, 16–18, 40, 50–54 and 59, the highest β-turn potential observed at residue 52. The importance of the β-turn occurring at this position probably explains the conservation of this segment in kringles. β-turns are predicted for both FN 1 and FN 2 at positions 17–18 (highest β-turn potential at 18α), 52, 56 and 65.

The remarkable agreement in the positions of the predicted β-sheets and β-turns suggests that the folding of FN 1 and FN 2 may be similar to that of kringles. We have found that the disulphide-bridges of the FN 2 structure of human FN connect Cys_{22} to Cys_{62} and Cys_{50} to Cys_{74} (fig.1), a pattern of disulphide bonding homologous to that found in kringles [1,3]. The observation that the homologous cysteines are paired in an identical manner shows that the homologous β-sheets of type II structures and kringles must be arranged similarly in the folded proteins.

The similarity of the pattern of β-sheets and disulphide-bridges may seem surprising in view of the fact that 3 regions of kringles are deleted from FN 1 and FN 2. The gap events affecting the

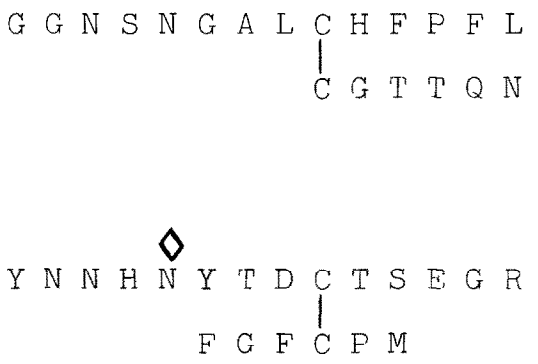

Fig.1. Disulphide-bridge pattern of the FN 2 structure of human FN. (◇) Oligosaccharide substituent.

amino-terminal and carboxy-terminal regions (residues 1–12 and 77–79) also affected Cys_1 and Cys_{79} that form a third disulphide-bridge in protease-kringles [1,3]. The probable explanation for the observation that this change did not profoundly affect the folding pattern is that this region of kringles is apparently not an essential part of the kringle-fold. We have shown that this disulphide-bridge is not essential for the biological activity or structural integrity of the kringle-fold [13,31] or for proper refolding of reduced-denatured kringle [38]. Another major difference between protease-kringles and type II structures is the absence of residues 26–44 from FN. As pointed out above, this segment is not part of the structurally conserved region of kringles and is notable for the lack of β-sheet structures (table 1). NMR studies have shown that Trp_{25} and Leu_{45} (conserved residues that delimit the variable region) interact in the kringle-fold [31] forcing this variable region to fold back on itself (probably at the β-turn predicted at position 40) and to bulge out of the kringle core. The gap events affecting the length of this segment also suggest that this region corresponds to a loop protruding from the kringle-fold. It seems therefore probable that the absence of this loop from FN 1 and FN 2 is compatible with the preservation of the basic features of the kringle-fold.

3.3. *Comparison of functions*

All kringle-structures identified so far occur in the amino-terminal non-protease extensions of

trypsin-type proteases (PT, PL, u-PA, t-PA) that are involved in complex biological processes such as blood coagulation, fibrinolysis and tissue remodelling [1–5]. It was shown above that the kringles are autonomous structural and functional modules that mediate the binding of these proteases or their zymogens to other proteins and thus ensure the regulation and biological specificity of their action.

Fibronectin is a high-molecular-mass extracellular protein that plays an important role in cell–cell interaction, cell adhesion, cell migration, wound healing and phagocytosis. The structural basis of the ability of fibronectin to mediate the interaction of cells with other cells and with a variety of extracellular macromolecules is that it is composed of several structural and functional modules that bind to cell surfaces, fibrin, collagen, actin, heparin, DNA and bacteria (review [39–42]). The large size of fibronectin and the structural and functional independence of the domains involved in the diverse binding functions suggest that this protein evolved from several smaller proteins. Protein sequence data on fibronectin have revealed the existence of several small repeats that belong to 3 different homology groups [18,19]; the type II structures are present in a region that has high affinity for collagen.

Comparison of the functions of FN with those of PT, PL, u-PA and t-PA thus shows that a common feature of these proteins is that both FN and the proteases contain autonomous structural-functional modules that serve to bind these multidomain proteins to other proteins or macroscopic structures.

3.4. *Genealogy of kringle-like structures*

The genealogies of the sequences constructed on the basis of percent difference and log odds matrices, consideration of gap and disulphide-bridge patterns and prediction of secondary structures gave consistent results. The evolutionary tree is shown in fig.2.

The fact that type II structures of FN are more similar to kringles of PT, PL, u-PA, t-PA than to the α-chain of HP indicates that type II structures branched off from the serine-protease lineage after the divergence of HP from other proteases. Since FN does not appear to have a protease-homologue

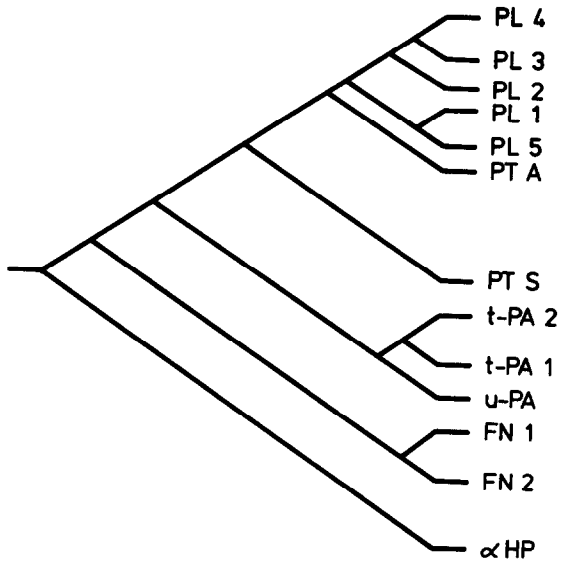

Fig.2. Genealogy of the kringle-like structures of prothrombin (PT), plasminogen (PL), urokinase- and tissue-type plasminogen activators (u-PA and t-PA), fibronectin (FN) and haptoglobin (HP).

region this would suggest that type II structures were introduced into the FN molecule by exchange of genetic materials. It is noteworthy in this respect that we have shown that a structure homologous to type I structures of FN is also present in the amino-terminal region of t-PA and we proposed that gene fusion or exon-reshuffling was the mechanism whereby these otherwise unrelated proteins acquired the same domain [43]. Such a mechanism would seem even more likely in view of the recent finding that type I structural repeat does indeed correspond to an exon of the FN gene [44]. A similar genetic 'autonomy' of the kringle regions of proteases is suggested by the fact that the A kringle of PT is much more similar to the kringles of PL than expected on the basis of the distance of the protease parts, raising the possibility that exchange of genetic materials occurred long after PT and PL diverged [20].

In view of the homology of the sequences, the homology of the predicted secondary structures and similarity of functions we propose that the protease-kringles and type II structures of FN are derived from a common ancestral protein-binding module.

REFERENCES

[1] Magnusson, S., Sottrup-Jensen, L., Petersen, T.E. and Claeys, H. (1975) in: Proteases and Biological Control (Reich, E. et al. eds) vol.2, pp.123–149, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

[2] Walz, D.A., Hewett-Emmett, D. and Seegers, W.H. (1977) Proc. Natl. Acad. Sci. USA 74, 1969–1972.

[3] Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T.E. and Magnusson, S. (1978) in: Progress in Chemical Fibrinolysis and Thrombolysis (Davidson, J.F. et al. eds) vol.3, pp.191–209, Raven, New York.

[4] Günzler, W.A., Steffens, G.J., Ötting, F., Kim, S.-M.A., Frankus, E. and Flohé, L. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 1155–1165.

[5] Pennica, D., Holmes, W.E., Kohr, W.J., Harkins, R.N., Vehar, G.A., Ward, C.A., Bennett, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.L., Goeddel, D.V. and Collen, D. (1983) Nature 301, 214–221.

[6] Jackson, C.M. and Nemerson, Y. (1980) Annu. Rev. Biochem. 49, 765–811.

[7] Wiman, B. and Collen, D. (1978) Nature 272, 549–550.

[8] Esmon, C.T. and Jackson, C.M. (1974) J. Biol. Chem. 249, 7791–7797.

[9] Wiman, B., Lijnen, H.R. and Collen, D. (1979) Biochim. Biophys. Acta 579, 142–154.

[10] Bloom, J.W. and Mann, K.G. (1979) Biochemistry 18, 1957–1961.

[11] Ploplis, V.A., Strickland, D.K. and Castellino, F.J. (1981) Biochemistry 20, 15–21.

[12] Castellino, F.J., Ploplis, V.A., Powell, J.R. and Strickland, D.K. (1981) J. Biol. Chem. 256, 4778–4782.

[13] Trexler, M. and Patthy, L. (1983) Proc. Natl. Acad. Sci. USA 80, 2457–2461.

[14] Lerch, P.G., Rickli, E.E., Lergier, W. and Gillessen, D. (1980) Eur. J. Biochem. 107, 7–13.

[15] Váradi, A. and Patthy, L. (1981) Biochem. Biophys. Res. Commun. 103, 97–102.

[16] Seegers, W.H., Novoa, E., Walz, D.A., Andary, T.J. and Hassouna, H.I. (1976) Thromb. Res. 8, 83–97.

[17] Markus, G., Priore, R.L. and Wissler, F.C. (1979) J. Biol. Chem. 254, 1211–1216.

[18] Skorstengaard, K., Thogersen, H.C., Vibe-Pedersen, K., Petersen, T.E. and Magnusson, S. (1982) Eur. J. Biochem. 128, 605–623.

[19] Petersen, T.E., Thøgersen, H.C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. and Magnusson, S. (1983) Proc. Natl. Acad. Sci. USA 80, 137–141.

[20] Kurosky, A., Barnett, D.R., Lee, T.-H., Touchstone, B., Hay, R.E., Arnott, M.S., Bowman, B.H. and Fitch, W.M. (1980) Proc. Natl. Acad. Sci. USA 77, 3388–3392.

[21] Jue, R.A., Woodbury, N.W. and Doolittle, R.F. (1980) J. Mol. Evol. 15, 129–148.

[22] Doolittle, R.F. (1981) Science 214, 149–159.

[23] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed) pp.345–352, National Biomedical Research Foundation, Silver Spring, MD.

[24] Schwartz, R.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed) pp.353–358, National Biomedical Research Foundation, Silver Spring, MD.

[25] De Haën, C., Neurath, H. and Teller, D.C. (1975) J. Mol. Biol. 92, 225–259.

[26] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.

[27] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.

[28] Engvall, E. and Ruoslahti, E. (1977) Int. J. Cancer 20, 1–5.

[29] Sekiguchi, K. and Hakomori, S.-I. (1980) Proc. Natl. Acad. Sci. USA 77, 2661–2665.

[30] Hirs, C.H.W. (1956) J. Biol. Chem. 219, 611–621.

[31] Trexler, M., Bányai, L., Patthy, L., Pluck, N.D. and Williams, R.J.P. (1983) FEBS Lett. 154, 311–318.

[32] Hochschwender, S.M. and Laursen, R.A. (1981) J. Biol. Chem. 256, 11172–11176.

[33] Trexler, M., Váli, Z. and Patthy, L. (1982) J. Biol. Chem. 257, 7401–7406.

[34] Greer, J. (1981) J. Mol. Biol. 153, 1027–1042.

[35] Llinas, M., De Marco, A., Hochschwender, S.M. and Laursen, R.A. (1983) Eur. J. Biochem. 135, 379–391.

[36] Venyaminov, S.Yu., Metsis, M.L., Chernousov, M.A. and Koteliansky, V.E. (1983) Eur. J. Biochem. 135, 485–489.

[37] Sjöholm, I., Wiman, B. and Wallén, P. (1973) Eur. J. Biochem. 39, 471–479.

[38] Trexler, M. and Patthy, L. (1984) Biochim. Biophys. Acta, in press.

[39] Hynes, R.O. and Yamada, K.M. (1982) J. Cell Biol. 95, 369–377.

[40] Ruoslahti, E., Pierschbacher, M., Hayman, E.G. and Engvall, E. (1982) Trends Biochem. Sci. 7, 188–190.

[41] Vartio, T. and Vaheri, A. (1983) Trends Biochem. Sci. 8, 442–444.

[42] Yamada, K.M. (1983) Annu. Rev. Biochem. 52, 761–799.

[43] Bányai, L., Váradi, A. and Patthy, L. (1983) FEBS Lett. 163, 37–41.

[44] Hirano, H., Yamada, Y., Sullivan, M., De Crombrugghe, B., Pastan, I. and Yamada, K.M. (1983) Proc. Natl. Acad. Sci. USA 80, 46–50.