View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Volume 204, number 2 FEBS 3926 August 1986

Exon structure of the collagen-binding domain of human fibronectin

Raymond J. Owens and Francisco E. Baralle

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, England

Received 15 May 1986

Fibronectin consists of a series of three internal homology repeats (types I, II, and III [(1983) Proc. Natl. Acad. Sci. USA 80, 137–141]). The type III units are each coded for by two exons with the exception of the alternatively spliced extra domain type III [(1984) EMBO J. 3, 2511–2516]. To investigate the gene organisation of the type I and II repeats, genomic clones covering the collagen-binding domain of human fibronectin have been isolated and characterised. The results show that each of the type I and II homology units in this region corresponds to an exon. Taken together with the previous data concerning the type III units, the data lend support to a gene fusion model of fibronectin evolution.

(Human) *Fibronectin* *Gene structure* *Collagen binding*

1. INTRODUCTION

Fibronectin is a ubiquitous extracellular glycoprotein that is involved in a variety of cell contact processes [1]. Two major forms of the molecule have been distinguished: a plasma form, which is a soluble disulphide-linked heterodimer, and a cell surface-associated (or cellular) form, which consists of dimers and multimers. Each subunit of fibronectin has a molecular mass of 230–250 kDa and contains specific sites for binding to cells and a range of macromolecules, including collagens, fibrin and heparin. These binding sites have been mapped to a series of protease-resistant domains. Partial amino acid sequencing of these fragments has shown that fibronectin consists of three types of homologous repeats (types I, II and III; [2]). The type I and II internal repeats consist of approx. 45 and 60 amino acids, respectively, and are stabilised by intrachain disulphide bridges, whilst the type III units are 90 residues long and have no disulphide bonds. The binding domains of fibronectin are characterised by a particular arrangement of internal repeats, implying a functional association between specific homology units and binding sites. For example, the amino- and carboxy-terminal fibrin-binding fragments comprise only type I repeats [2,3], whilst the 30–40 kDa collagen-binding domain consists of the only two type II units in fibronectin, in addition to a number of type I units [4].

There is a single fibronectin gene per haploid genome in man [5] and rodents [6]. Subunit variants of fibronectin are generated by complex alternative splicing of the primary transcript [7,8]. The repeating structure of the fibronectin polypeptide raises the question of whether this is reflected in the organisation of the fibronectin gene. Recently it has been established, both for human and rat fibronectins, that the type III units are each encoded by two exons [9–11]. The only exception appears to be a unique type III repeat (referred to as the extra domain, ED) that undergoes alternative splicing and is coded for by a single exon. There is some evidence from a study of the chicken fibronectin gene that the type I units may correspond to single exons [12]. However, no information concerning the exon structure of the type II repeats has been reported.

In this paper we describe the isolation of λ

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50 © 1986 Federation of European Biochemical Societies

genomic clones covering approximately the N-terminal third of the human fibronectin gene. Sequences corresponding to the collagen-binding domain have been partially characterised. The results show that the type I and type II units are each coded for by single exons. We discuss the implications of this for the evolution of the fibronectin gene and the structure of the collagen-binding site of the molecule.

2. MATERIALS AND METHODS

2.1. *Materials*

A partial *Mbo*I human genomic library in λ EMBL3 was generously provided by Dr Colin Sharpe [13] from this laboratory. The isolation of the following human fibronectin cDNA clones has been described [14], pFH6, 16, 51 and 134. The plasmid pFH16/II was prepared by subcloning a *Pvu*II/*Rsa*I fragment of pFH16 into pAT153/*Pvu*II/8 [15] and corresponds to the two type II homology units of fibronectin. Probes were either restriction fragments 3′-end-labelled with the Klenow fragment of DNA polymerase I [16] and eluted from acrylamide gels or eluted from agarose gels and nick-translated [17].

2.2. *Screening of genomic library*

Approx. 4×10^5 recombinants were screened with pFH6 labelled by nick-translation [18]. Hybridisation positive signals were re-screened separately with pFH6, 51, and 16/II. From nine purified plaques, two independent λ clones were obtained (λFN1 and λFN7).

2.3. *Subcloning*

Total *HindIII digests of* λFN1 and -7 were cloned separately into pAT153/*Pvu*II/8 vector and two different clones (pλFN Hind 1.3 and 5.2) which hybridised to pFH16/II and one (pλFH Hind 5.5) which hybridised to pFH51 but not pFH16/II, were chosen for further analysis.

2.4. *DNA sequencing*

End-labelled fragments were sequenced according to the method of Maxam and Gilbert [16].

3. RESULTS

3.1. *Isolation and characterisation of genomic clones for fibronectin*

A series of cDNA clones corresponding to the N-terminal 40% of human fibronectin has been isolated [14]. One of these clones, pFH6, which covers ~1.3 kb from the 5′-end of the coding sequence, was used to isolate two closely overlapping genomic clones for fibronectin, named λFN1 (13.5 kb) and λFN7 (13.0 kb). Dot hybridisation experiments indicated that together the clones extended from approximately position 600 bp to 1800 bp in the fibronectin cDNAs [14]. They thus encompassed the sequences coding for collagen-binding domains of the protein. Restriction enzyme mapping in conjunction with Southern blot-

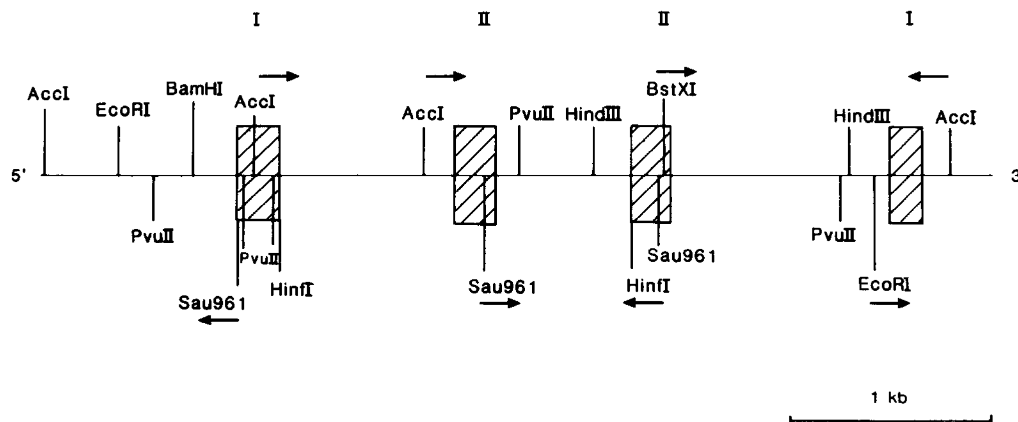

Fig.1. Restriction enzyme map of part of the human fibronectin gene showing the position of the type II and flanking type I exons. The strategy for sequencing is shown with arrows. Only the *Sau*961 and *Hin*fI sites in the exons are shown.

ting showed that the type II homology units and immediately adjacent type I units were contained in a series of consecutive *Hin*dIII fragments. These were subcloned and the resulting recombinant plasmids were mapped in detail; part of this restriction map is shown in fig.1. In total the inserts of these subclones covered 12 kb of the fibronectin gene and contained the exon sequences of both type II homology units together with five type I units.

3.2. *Exon/intron organisation*

The nucleotide sequence of the exon/intron boundaries of the two type IIs and two flanking type I repeats was determined as outlined in fig.1. Exon sequences were identical to those reported for fibronectin cDNAs [14]. The results showed that the type I and II units are encoded by one exon each. The 5′-type I exon comprises some non-homologous sequence in addition to the type I homology. The exon/intron junctions of the 3′-type I and two type IIs correspond exactly to the limits of these units inferred by the alignment of the protein sequence (fig.2). All of the splice junctions obey the GT/AG rule and all show some resemblance to the consensus donor ($^{A}_{C}$AG/GT$^{A}_{G}$AGT) and acceptor ([Py]$_6$XCAG/G$^{G}_{C}$) sequences [19].

Although the boundaries of three other type I units further 5′ and 3′ of the type II repeats were

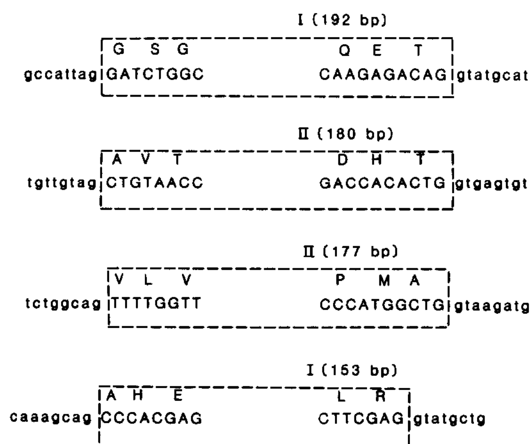

Fig.2. Sequence of the exon/intron boundaries of the type II exons and flanking type I exons. Intron sequences are in lower case letters.

not analysed, restriction mapping and Southern blotting experiments using cDNA probes specific for these exons, indicated that each was encoded by a single exon (not shown).

4. DISCUSSION

In this report the exon structure of the collagen-binding domain of fibronectin has been described for the first time. Each of the two type II homology units, unique to this region, are defined by a complete exon. The type I units in this domain also represent single exons. It seems likely that all the type I units of fibronectin are each coded by an exon. Indirect evidence in support of this comes from correlating the overall structure of the human fibronectin polypeptide deduced from cDNAs [14] with the size of the chicken fibronectin gene. Hirano et al. [12] showed, by R-loop mapping, that the chicken gene is approx. 50 kb long and contains at least 48 exons. Of these, 33 may be accounted for by the type II and III repeats ([2 type II × 1] + [15 type III × 2] + [ED × 1]) leaving approx. 15 exons. Since there are 12 type I units in the fibronectin polypeptide, each most probably represents a single exon. The size of the exons (~130 bp) in the type I homology rich areas of chicken fibronectin [12] is consistent with this assignment. Thus the modular organisation of the fibronectin polypeptide is reflected in the organisation of the fibronectin gene.

It has been suggested that the three types of homology in fibronectin originated from three independent genes [2]. If so, the contemporary fibronectin sequence may be the result of the fusion and amplification of the three primitive genes, with a subsequent process of divergence of the original repeats during evolution. These events must have occurred well before the divergence of vertebrates since the level of sequence conservation between any two homology units of different species is much greater than within a species. The exact association between the internal repeats and exons lends general support to this model of fibronectin evolution. The finding that the type III homologies are each coded for by two exons shows that the position of the intron between these component exons has been remarkably conserved. Alternatively it seems possible that fibronectin may have originated from four rather than three

primordial genes. In any case the ED segment seems to be the result of the subsequent fusion of two type III exons.

Indirect evidence supporting a gene fusion model of fibronectin evolution comes from an examination of the structure of some contemporary proteins. Tissue plasminogen activator contains a homologous type I repeat [20] which is encoded by a single exon [21]. It has also recently been shown that human factor XIIa contains a type I unit together with a type II homology [22,23]. The type II repeat in factor XIIa shows approx. 50% homology with the type II repeats of fibronectin and is also very similar to two type II sequences found in a bovine seminal plasma protein [24]. From the results presented here we would predict that the type II repeats of factor XIIa and the bovine protein each correspond to a complete exon. Thus two of the three building blocks of fibronectin have been identified in other proteins, suggesting that they arose independently.

Finally it is interesting to consider the relationship between the exon structure of the type I and II homology units reported here and the functional domain which they form. The unique association of the type II repeats with the collagen-binding domain strongly suggests that they may be involved in the binding activity. We have recently analysed this binding site in human fibronectin using a bacterial expression system (Owens and Baralle, in preparation). The results indicate that although both type II units and adjacent type Is are probably necessary for full binding activity, the actual binding specificity for collagen lies within the second type II and adjacent type I repeats. Thus the binding site is not contained in an exon but is generated by the splicing together of adjacent exons. Furthermore, it seems that the two type II units, though structurally similar at both gene and protein level, are not functionally equivalent. A similar argument may be made for the type I units which feature in both fibrin and collagen-binding domains and the type IIIs of the cell-attachment domain of which only one contains the Arg-Gly-Asp-Ser cell-binding signal [11]. The exon/intron boundaries in the fibronectin gene clearly define structural units in the protein but do not necessarily define the functional units. This is consistent with the pattern observed in other genes (e.g. serine proteases [25]).

ACKNOWLEDGEMENTS

We are grateful to Dr Colin Sharpe for providing the human gene library and Dr Gianfranco Sebastio for help with the initial isolation of the clones. R.J.O. is the recipient of an MRC Training Fellowship in recombinant DNA technology.

REFERENCES

[1] Yamada, K.M. (1983) Annu. Rev. Biochem. 52, 761–799.
[2] Petersen, T., 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.
[3] Garcia-Pardo, A., Pearlstein, E. and Frangione, B. (1983) J. Biol. Chem. 258, 12670–12674.
[4] Skorstengaard, K., Thøgersen, H. and Petersen, T. (1984) Eur. J. Biochem. 140, 235–243.
[5] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1983) Proc. Natl. Acad. Sci. USA 80, 3218–3222.
[6] Tamkun, J.W., Schwarzbauer, J.E. and Hynes, R.O. (1984) Proc. Natl. Acad. Sci. USA 81, 5140–5144.
[7] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1984) Nucleic Acids Res. 12, 5853–5868.
[8] Schwarzbauer, J.E., Tamkun, J.W., Lemischka, I.R. and Hynes, R.O. (1983) Cell 35, 421–431.
[9] Vibe-Pedersen, K., Kornblihtt, A.R. and Baralle, F.E. (1984) EMBO J. 3, 2511–2516.
[10] Odermatt, E., Tamkun, J.W. and Hynes, R.O. (1985) Proc. Natl. Acad. Sci. USA 82, 6571–6575.
[11] Oldberg, A. and Ruoslahti, E. (1986) J. Biol. Chem. 261, 2113–2116.
[12] Hirano, H., Yamada, Y., Sullivan, M., De Crombrugghe, B., Pastan, I. and Yamada, K.M. (1983) Proc. Natl. Acad. Sci. USA 80, 46–50.
[13] Sharpe, C.R. (1985) D. Phil. Thesis, University of Oxford.
[14] Kornblihtt, A.R., Umezawa, K., Vibe-Pedersen, K. and Baralle, F.E. (1985) EMBO J. 4, 1755–1759.
[15] Anson, D.S., Choo, K.H., Rees, D.J.G., Giannelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1984) EMBO J. 3, 1053–1060.
[16] Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–569.
[17] Rigby, P.W., Dieckmann, H., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[18] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[19] Cech, T.R. (1983) Cell 34, 713–716.

[20] Banyai, L., Varadi, A. and Patthy, L. (1983) FEBS Lett. 163, 37–41.
[21] Ny, T., Elgh, F. and Lund, B. (1984) Proc. Natl. Acad. Sci. USA 81, 5355–5359.
[22] McMullen, B.A. and Fujikawa, K. (1985) J. Biol. Chem. 260, 5328–5341.
[23] Cool, D.E., Edgell, C.-J., Louie, G.V., Zoller, M.J., Brayer, G.D. and MacGillivray, R.T. (1985) J. Biol. Chem. 260, 13666–13676.
[24] Esch, F.S., Ling, N.C., Bohlen, P., Ying, S.Y. and Guillemin, R. (1983) Biochem. Biophys. Res. Commun. 113, 861–867.
[25] Rogers, J. (1985) Nature 315, 458–459.