Protein Science (1993), *2*, 1185–1187. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Fibronectin type III modules in the receptor phosphatase CD45 and tapeworm antigens

PEER BORK[1] AND RUSSELL F. DOOLITTLE[2]

[1] European Molecular Biology Laboratory, Heidelberg, Germany
[2] Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093-0634

(RECEIVED March 17, 1993; REVISED MANUSCRIPT RECEIVED April 5, 1993)

The fibronectin type III module (Fn3) is one of the most widespread domains in modern mosaic proteins, being found in both extracellular and intracellular proteins as well in the extracellular portions of numerous transmembrane receptor proteins (Bork, 1992; Bork & Doolittle, 1992; Doolittle, 1992). At this point, more than 300 occurrences have been stored in current sequence databases, and almost every week a new sequence report of an Fn3 module-containing protein appears in the literature. Recently, we screened several sequence databases for homologous Fn3 segments in an effort to characterize conserved features and to chart the course of their evolution (Bork & Doolittle, 1992). During that study, we uncovered a number of putative Fn3 domains that, to our knowledge, had not yet been described (Fig. 1; Table 1). These included (1) six domains in several tapeworm proteins, (2) a third module in the fruit fly receptor protein tyrosine phosphatase DPTP (Streuli et al., 1989), and (3) domains in two other tyrosine phosphatases, one known as PTPζ (Krueger & Saito, 1992) and the other the leukocyte common antigen referred to as LCA or CD45 (Saito et al. [1992] and references therein). The last-named are of particular importance because they were thought to be unique among the receptor phosphatases with large extracellular portions in lacking Fn3 units (Charbonneau & Tonks, 1992; Krueger & Saito, 1992). Indeed, we find hints of two additional Fn3 modules in CD45, but the degree of similarity is not sufficient to meet our criteria for significance (Table 1).

The uncovering of additional Fn3 modules in known mosaic proteins is revealing on several counts. First, restrictions can be imposed on possible binding sites. Second, it allows reasonable predictions to be made about the three-dimensional structure of these particular proteins. For example, in CD45, the major glycoprotein of all types of leucocytes and a focus of much current research, the presence of at least one Fn3 module (and perhaps two others) allows the delineation of domains and suggests, by analogy to other receptor proteins involved in signal transduction, possible dimerization sites. In this regard, the three-dimensional structures of Fn3 modules from three different proteins have been determined and found to be similar to immunoglobulin domains (Kinemage 1; DeVos et al., 1992; Leahy et al., 1992; Main et al., 1992).

Whereas the presence of the Fn3 modules in known mosaic proteins may not be altogether unexpected, their existence in tapeworm antigens is a surprise. All three proteins identified (*onca* and *oncb* from *Taenia taeniaformis*, a parasite that infects rodents, and a 45-kDa protein from *Taenia ovis*, a sheep tapeworm) have been used for vaccination and shown to protect hosts against infection (Johnson et al., 1989; Cougle et al., 1991). The Fn3 module is known to act as an adhesive domain in numerous settings and may very well be involved in anchoring larval cestodes to the host intestinal wall, a process that would be expected to be suppressed by immunization with the respective antigens.

Receptor protein tyrosine phosphatases are involved in diverse cell cycle activities (for a recent review see Charbonneau & Tonks [1992]). In many cases, their extracellular parts contain several Fn3 modules, often accompanied by amino-terminal immunoglobulin-like domains. With the detection of Fn3 domains in the extracellular parts of DPTP, PTPζ and CD45, further regions of these phosphatases can be classified. Moreover, all sequenced receptor protein tyrosine phosphatases with large extracellular parts are now known to contain Fn3 modules, suggesting functional analogies. This is supported by the fact that

Reprint requests to: Russell F. Doolittle, Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093-0634.

```
Fn3 cons.          P  h h        h h Wttt     h  h h h                th h  L P   Y h h A    t
Ali cons.          P  h h t  tt th h Wt t   t h th h h tt      t h     th httLtt   Y h h Ah   t

45kd_Taeov-1    18 RDIFAWEP.PTSNSIGLTWQRH-6-REVLTLKAVPTSEPN-04-AYAKLGSGKVTLDGLKPNATYLVTATANISGD..
45kd_Taeov-2   115 NNIFHWGP.VTNQSIQVRWDQI-6-ALIVTLTAEMASDPG-04-ESALFGKGKVTVDGLESDTLYIATVMVFRNGR..
  Oncb_Taeta    12 PQHFRWSQ.VGSQSIQLSWDDH-7-DQVHLFVIPRSSSLQ-03-KQVAFSAKTVILEGLHGNTLYYVILTVSAGEE..
Onca_Taeta-1     7 PQNVQIEA.IDPHTARMTWDPP-4-GSIIGYTIQWSTDDL-03-QVKVASDNSYDFKDLQSQQTIVASILAHHRPD..
Onca_Taeta-2   111 SFRATYCE..DLEELDMLIHDP-4-GMFGGFEVLMRAGDA-09-VNLTAEERRYKLKGLVPSLPYAVTVRGVALPS..
  Cd45_Human   393 PQIIFCRS.EAAHQGVITWNPP-1-RSFHNFTLCYIKETE-03-LNLDKNLIKYDLQNLKPYTKYVLSLHAYIIAK..
    Cd45_Rat   347 P.HVQCKN.STNSTTLVSWAEP-1-SKHHGYILCYKKTPS-03-ENLANDVNSFEVKNLRPYTEYTVSLFAYVIGR..
  Cd45_Mouse   239 P.KPSCGD.PAARKTLVSWPEP-7-SKPHGYVLCYKNNSE-02-KSLPNNVTSFEVESLKPYKYYEVSLLAYVNGK..
Dptp_Drome-3   444 SIQCHHDNVTRGNSIAINWDVP-4-GKVVSYLIHLLGNPM-12-RRIDEPHHKTLYESVSPNTNYTVTVSAITRHK..
  Rtpz/Human   315 P.ENVQADPENYTSLLVTWERP-5-TMIEKFAVLYQQLDG-08-LTDGYQDLGAILNNLLPNMSYVLQIVAICTNG..

  Tena_Chick 1409 PSGLVVMN.ITDSEALATWQPA-1-AAVDNYIVSYSSEDE-03-QMVSGNTVEYDLNGLRPATEYTLRVHAVKDAQ..
  Rtp97Drome   174 PQNLTILD.VSANSITMSWHPP-4-GAIAGYHVFHIHDNQ-10-NSVETL.IHFELQNLRPYTDYRVIVKAFTTKN..
  Keck_Human 1114 PPKVRLEG.RSTTSLSVSWSIP-4-SRVWKYEVTYRKKGS-04-NVRRTEGFSVTLDDLAPDTTYLVQVQALTQEG..
    Finc_Rat   612 VQVIITETPSQPNSHPIQWNAP-2-SHITKYILRWRPKTS-06-ATIPGHLNSYTIKGLTPGVIYEGQLISIQQYG..

beta strands         aaaaa       bbbbbb        cccccccc          ddddd   eeeeeee    fffffffff
```

Fig. 1. Multiple alignment of the putative Fn3 modules reported here with selected Fn3 domains from other proteins as revealed by our pattern search program (Rohde & Bork, 1993). The numbers after the SWISSPROT code names refer to the residue position in the native proteins. Only conserved regions have been included (the carboxy-terminal segments have been omitted). The strand assignment (bottom line) is implied by the known three-dimensional structures of several Fn3 modules (DeVos et al., 1992; Leahy et al., 1992; Main et al., 1992). The consensus line of the alignment (Ali cons.) highlights common features: bold capitals, conserved amino acids; h, hydrophobic; t, polar or turnlike and probably located at the surface. Only a few exceptions per position are allowed. This consensus is very similar (bold) to a general Fn3 consensus (Fn3 cons) based on an analysis of a multiple alignment of 240 Fn3s (data not shown) in which 5% exceptions per position were tolerated.

Table 1. *Database search statistics*[a]

Query	Two best hits	FASTA %id	Opt	Max	PROFILE score	PROPAT mismatch
CD45_Rat	Tenascin	27.1	113	488	5.22	4
	rptp 99A	28.9	110			
Dptp_Drome-3	rptp 99A	28.8	91	483	5.61	5
	LAR	31.0	89			
Rtpz/Human	rptp delta	26.1	106	495	5.31	4
	N-CAM 1	23.7	102			
Onca_Taeta-1	Onca_Taeta-3	40.1	158	535	5.28	6
	Undulin	26.8	94			
Onca_Taeta-2	vir.polym.(*)	23.1	82	505	5.28	7
	GCSF	26.4	82			
Onca_Taeta-3	Onca_Taeta-1	40.1	158	481	5.28	7
	col.canc.supr.	30.6	72			
Oncb_Taeta	45kd_Taeov-2	31.0	128	513	4.61	7
	LAR	22.7	106			
45kd_Taeov-1	Oncb_Taeta	27.2	120	519	4.25	5
	Fibronectin	29.3	82			
45kd_Taeov-2	Oncb_Taeta	31.0	128	521	4.25	6
	Tenascin	19.0	80			

[a] The significance of Fn3 assignments was assessed by three methods. First, each putative Fn3 region, including the portions shown in Figure 1 and an additional 5 and 10 residues on the amino- and carboxy-terminal ends, respectively, was subjected to a FASTA database search (Pearson & Lipman, 1988). The maximal score possible (max) is defined by self comparison. The output was sorted for optimized scores (opt), which include weights for similar amino acids and gap penalties (Pearson & Lipman, 1988) and was filtered according to a threshold for structural homology (Sander & Schneider, 1991). Known Fn3 sequences from different subgroups (Bork & Doolittle, 1992) ranked among the best hits in all runs. The only (probably) false positive in the FASTA searches was a viral polymerase (*), which does not match any of the otherwise fully conserved positions (Fig. 1). Profile or pattern search methods are usually more sensitive in the detection of distant relatives because they weight structurally and functionally important positions. We combined two different methods: (1) PROFILESEARCH (Gribskov et al., 1987) based on an alignment of 240 known Fn3s that have a pairwise identity less than 70% (data not shown); the false positives scored up to 4.95, but several known Fn3s had scores below this cutoff; (2) PROPAT, a property pattern search method (Rohde & Bork, 1993); in this case the first false positive had a mismatch score of 6. All candidates that ranked ahead of the first false positive by either method were considered. Only the *oncb* sequence and the second domain of the 45-kDa tapeworm protein fell below the level of clear significance, and the certain relationship between *oncb* and both domains of the 45-kDa protein, as well as the presence of several wholly conserved features (Fig. 1), implies genuine homology.

Fn3 modules of receptor protein phosphatases tend to cluster together in a comparative sequence analysis of all known Fn3 domains (data not shown).

Reasonable three-dimensional models of newly identified Fn3 modules can be built based on the known three-dimensional structures of other Fn3 domains. In this regard, previous studies have shown that ligand-binding sites are mostly located in loop regions (Fig. 1; Kinemage 1). In the case of the Fn3 modules in tapeworm antigens, this observation might prove helpful in the analysis of the host infiltration process.

References

Bork, P. (1992). Mobile modules and motifs. *Curr. Opin. Struct. Biol. 2*, 413-421.

Bork, P. & Doolittle, R.F. (1992). Proposed acquisition of an animal domain by bacteria. *Proc. Natl. Acad. Sci. USA 89*, 8990-8994.

Charbonneau, H. & Tonks, N.K. (1992). 1002 protein phosphatases? *Annu. Rev. Cell Biol. 8*, 463-493.

Cougle, W.G., Lightowlers, M.W., Bogh, H.O., Rickard, M.D., & Johnson, K.S. (1991). Molecular cloning of *Taenia taeniaformis* antigen genes. *Mol. Biochem. Parasitol. 45*, 137-146.

DeVos, A.M., Ultsch, M., & Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science 255*, 306-312.

Doolittle, R.F. (1992). Reconstructing history with amino acid sequences. *Protein Sci. 1*, 191-200.

Gribskov, M., McLachlan, A.D., & Eisenberg, D. (1987). Profile analysis: Detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA 84*, 4355-4358.

Johnson, K.S., Harrison, G.B.L., Lightowlers, M.W., O'Hoy, K.L., Cougle, W.G., Dempster, R.P., Lawrence, S.B., Vinton, J.G., Heath, D.D., & Rickard, M.D. (1989). Vaccination against ovine cysticercosis using a defined recombinant antigen. *Nature 338*, 585-587.

Krueger, N.X. & Saito, H. (1992). A human transmembrane protein tyrosine phosphatase, PTPz, is expressed in brain and has an N-terminal receptor domain homologous to carbonic anhydrases. *Proc. Natl. Acad. Sci. USA 89*, 7417-7421.

Leahy, D.L., Hendrickson, W.A., Aukhil, I., & Erickson, H.P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science 258*, 987-991.

Main, A.L., Harvey, T.S., Baron, M., Boyd, J., & Campbell, I.D. (1992). Solution structure of the fibronectin type III module. *Cell 71*, 671-678.

Pearson, W.R. & Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA 85*, 2444-2448.

Rohde, K. & Bork, P. (1993). A fast sensitive pattern matching approach for protein sequences. *CABIOS 9*(2), 183-189.

Saito, H., Streuli, N., Krueger, N.X., & Tsai, A.Y.M. (1992). CD45 and a family of receptor-linked protein tyrosine phosphatases. *Biochem. Soc. Trans. 20*, 165-169.

Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins Struct. Funct. Genet. 9*, 56-68.

Streuli, M., Krueger, N.X., Tsai, A.Y.M., & Saito, H. (1989). A family of receptor-linked protein tyrosine phosphatases in humans and *Drosophila*. *Proc. Natl. Acad. Sci. USA 86*, 8698-8702.

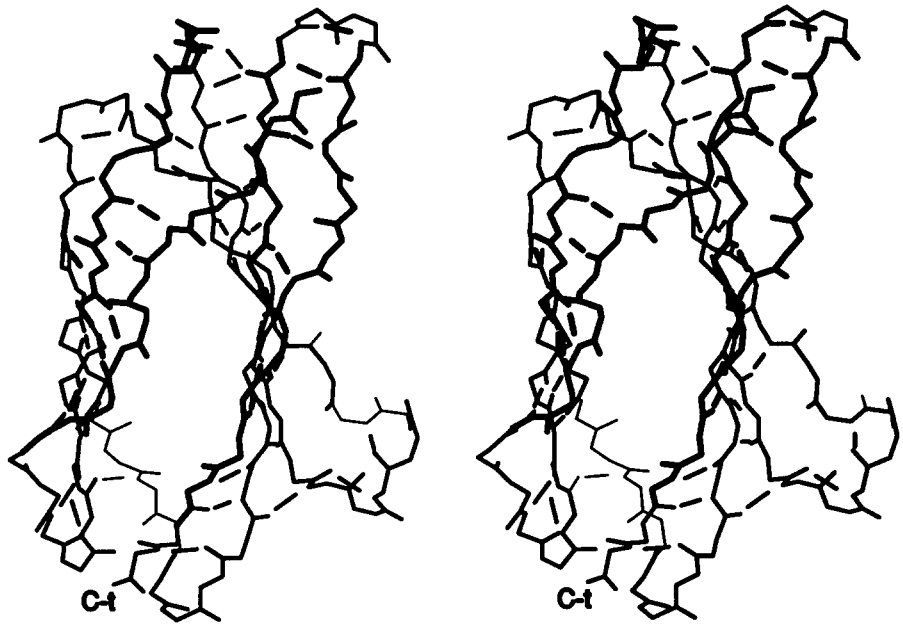

Figure added in proof. Stereo diagram of the Fn3 domain from tenascin (Brookhaven file 1TEN), showing the main chain and the hydrogen bonds of the β structure. (Modified from kinemage file 2Bork.kin, Diskette Appendix.)