Cell, Vol. 88, 9–11, January 10, 1997, Copyright ©1997 by Cell Press

Letter to the Editor

Secreted Fringe-like Signaling Molecules May Be Glycosyltransferases

Pattern formation during development requires the regulated expression of numerous signaling molecules. One of these, Drosophila Fringe (FNG) is a novel secreted protein with a key role in dorsal-ventral aspects of wing formation (Irvine and Wieschaus, 1994). Recently, multiple functions have been assumed for a Xenopus homologue (lunatic Fringe, lFNG) including the induction of mesoderm; a complex expression pattern supports this notion (Wu et al., 1996).

Complex and general functions of Fringe-like proteins are also indicated by the presence of two divergent C. elegans and at least six human homologues identifiable by sequence database searches (Figure caption). Thus, conservation patterns within the FNG family emerge that allow the use of sensitive motif and profile searches (for details see Bork and Gibson, 1996). Indeed, we found significant similarities of FNG-like proteins to Drosophila Brainiac (BRN; Goode et al., 1996) and, surprisingly, also to the Lex1 family of biosynthetic galactosyltransferases (Figure 1).

BRN has been proposed to be required for proper contact or adhesion between germline and follicle cells (Goode et al., 1992,1996). BRN and FNG share several features: i) they are developmentally regulated, secreted signaling molecules without known receptors, ii) they are required during (dorso-ventral) epithelial patterning (Irvine and Wieschaus, 1994; Wu et al., 1996; Goode et al., 1996), iii) they interact genetically with the Notch and/or EGF receptor pathways (Kim et al., 1995; Goode et al., 1996), suggesting that they might modify the signaling mediated by these receptors, and iv) FNG and BRN both have at least two C. elegans and several human homologues suggesting the presence of multigene families (Figure 1).

Lex1 of Haemophilus influenzae is essential for the biosynthesis of its extracellular lipooligosaccharides (LOS) (Cope et al., 1991) as is its homologue in another parasitic bacterium, Pasteurella haemolytica (Potter and Lo, 1995). In two other parasites with a similar LOS architecture, Neisseria meningitidis and Neisseria ghonorrhoeae, two highly related proteins of the Lex1 family have been independently characterized in each organism as galactosyltransferases (Gotschlich, 1994; Jennings et al., 1995) that add galactose to glucose or N-acetylglucosamine residues of the LOS (Jennings et al., 1995). The LOS of all these parasitic bacteria contain epitopes that are antigenically and structurally very similar to carbohydrates present in human glycosphingolipids; the parasites are thus able to mimic the latter (Mandrell et al., 1992). Furthermore, the bacterial galactosyltransferases are significantly similar (blast (Altschul et al., 1994) value for the probability of matching by chance $p = 1.9 \times 10^{-7}$) to a putative secreted protein from C. elegans (Figure 1) as well as to human and mouse ESTs (blast p-values $<10^{-11}$).

The two most conserved regions of all three subfamilies FNG, BRN, and Lex1 (motif 3 and 4 in Figure 1) are also the major hallmarks of the putative glycosyltransferase superfamily (a more precise prediction of the substrate specificity is not possible given the limited sequence similarity between the subfamilies; Paulson and Colley, 1989). In the context of an enzymatic function it has to be noted that only negatively charged amino acids are invariant, thus pointing to catalytic residues similar to those in characterized glycosyltransferase families (Saxena et al., 1995; Strokopytov et al., 1995; Qian et al., 1994). This is supported by secondary structure predictions (Rost et al., 1994) that are consistent around all conserved regions (Figure 1) and that predict the conserved negatively charged residues to be located in exposed loops indicating a catalytic role. The alternating arrangement of α-helices and β strands suggest an α/β folding type for the central portion of each subfamily, similar to that of other glycosyltransferases (Strokopytov et al., 1995; Qian et al., 1994; Saxena et al., 1995).

The conclusion that signaling molecules involved in pattern formation such as Fringe and Brainiac, may be secreted glycosyltransferases might come as a surprise, but is not completely unexpected: i) Glycosyltransferases have been implicated in developmental processes for a long time (Shur, 1977a,b); ii) many extracellular, highly expressed glycosyltransferases have been shown to exist in humans (Lammers and Jamieson, 1988; Fujita-Yamaguchi and Yoshida, 1981); iii) the expression of secreted glycosyltransferases increases during embryonic development (Cho et al., 1996); iv) the extracellular carbohydrate moieties change during development (Masteller et al., 1995) as a function of the expressed glycosyltransferases (Kukowska-Latallo et al., 1990); and v) transmembrane galactosyltransferases have been shown to transmit intracellular signals after binding substrates via their extracellular part (Gong et al., 1995).

Secreted glycosyltransferases may use their ability to recognize specific carbohydrate moieties on cell surface molecules to trigger particular receptors (Shur, 1993), but they might also play a crucial role in epithelial pattern formation by modifying these carbohydrate moieties at particular locations recognizable by various carbohydrate-binding domains of extracellular proteins. Numerous distinct ESTs from multicellular organisms including Arabidopsis (no match was found in yeast or other completely sequenced unicellular genomes) suggest a vast superfamily of glycosyltransferases that might belong to a system of posttranslational modification independent from the Golgi apparatus. The carbohydrate status of the cell during development might even be a function of neighboring cells and not only of its own expression set of glycosyltransferases.

Figure 1. Alignment of Five Regions Conserved in the FNG, BRN and Lex1 Subfamilies

First four columns: gene names, species, accession numbers, position of the motifs in the respective sequence. The numbers in angle brackets between the 5 aligned motifs indicate the number of residues that are not displayed. Amino acids conserved in at least 60% of the sequences are highlighted; hydrophobic residues conserved in at least 85% of the sequences are given in bold. The consensus line summarises residue properties conserved in at least 75% of all sequences (t - turn-like or polar, h - hydrophobic, s - small, a - aromatic). Secondary structures predicted (Rost et al., 1994) independently for each of the three subfamilies (E/e - strand, H/h - helix with 81%/72% expected accuracy) are given beneath the alignment. A region of 44 residues has been removed from the C. elegans U64858 sequence as it is very likely an artificially translated intron because of the presence of splice sites and similarity to an intronic region in Z35641. U64858 has been elongated due to exons detected upstream. The N-terminal regions of the FNG and BRN subfamilies that preceed the alignment with the Lex1 family are divergent and have been proposed to be a propeptide for the fringe subfamily (Wu et al., 1996). All eukaryotic, but also some prokaryotic sequences contain leader peptides.

Starting point for the iterative database searches (for details of the strategy see Bork and Gibson, 1996) was FNG. Blastp searches (Altschul et al., 1994) revealed two C. elegans homologues with blastp values $p < 10^{-4}$; a multiple alignment generated by MACAW (Schuler et al., 1991) identifies several significantly similar regions with $p < 10^{-50}$. The two most conserved regions (motif 3 and 4) of the extended FNG family were then subjected to iterative MoST searches (Tatusov et al., 1994). Using the restrictive ratio of observed versus expected hits $r = 0.005$, motif 3 matched only BRN with $p = 0.022$ (as derived from the expected number of false positives $E = 0.023$ calculated using the MoST program). The majority of the Lex1 subfamily has been identifed using motif 4 (p values down to 0.002 using the threshold $r = 0.05$) without false positives. Members of all three subfamilies were then aligned using MACAW (Schuler et al., 1991) which reidentified the conserved motifs 3 and 4 with significant p values ($<10^{-50}$). All other motifs have significant MACAW p values for more than one subfamily. This was confirmed by Blastp and MoST searches with all motifs excluding motif 3. Nevertheless, support for correct alignment of motif 1 of the Lex1 family and motif 2 of the BRN comes mostly from the secondary structure context.

Numerous ESTs (e.g. hL1, hB1, hB2, hF3) were detected by tblastn (Altschul et al., 1994) searches suggesting a high expression rate. Each individual EST was aligned with the protein families using a dynamic programming algorythm as implemented in PAIRWISE (Birney et al., 1996). Overlapping ESTs were merged. Four Arabidopsis thaliana ESTs were identified as belonging to the superfamily (H36894, H37262, Z37212, Z34632). In addition to the hF1 and hF2 ESTs found by (Wu et al., 1996), N47633, R65917, Z45166, and the merged R65818/H87611/N51686/H55274 appear to be independent fragments similar to distinct members of the Fringe subfamily (probability for chance matches $p < 10^{-12}$). The ESTs provided additional support for the homology between the subfamilies. For example, the translated human EST N47633 matches members of the Fringe and the Lex1 subfamily with blastp p values of 1.7×10^{-31} and 2.5×10^{-10}, respectively (using the PAM150 amino acid similarity matrix).

Yan P. Yuan, Jörg Schultz, Marek Mlodzik, and Peer Bork
EMBL
Meyerhofstr. 1
69012 Heidelberg
and Max-Delbrück-Center
for Molecular Medicine
Berlin-Buch
Germany

References

Altschul, S.F., Boguski, M.S., Gish, W., and Wootton, J.C. (1994). Nat. Genet. *6*, 119–129.

Birney, E., Thompson, J.D., and Gibson, T.J. (1996). Nucl. Acid. Res. *14*, 2730–2739.

Bork, P., and Gibson, T. (1996). Meth. Enzym. *266*, 162–183.

Cho, S.K., Yeh, J., Cho, M., and Cummings, R.D. (1996). J. Biol. Chem. *271*, 3238–3246.

Cope, L.D., Yogev, R., Mertsola, J., Latimer, J.L., Hanson, M.S., McCracken, G.H., and Hansen, E.J. (1991). Mol. Microbiol. *5*, 1113–1124.

Fujita-Yamaguchi, Y., and Yoshida, A. (1981). J. Biol. Chem. *256*, 2701–2706.

Gong, X., Dubois, D.H., Miller, D.J., and Shur, B.D. (1995). Science *269*, 1718–1721.

Goode, S., Morgan, M., Liang, Y.-P., and Mahowald, A.P. (1996). Dev. Biol. *178*, 35–50.

Goode, S., Wright, D., and Mahowald, A.P. (1992). Development *116*, 177–192.

Gotschlich, E.C. (1994). J. Exp. Med. *180*, 2181–2190.

Irvine, K.D., and Wieschaus, E. (1994). Cell *79*, 595–606.

Jennings, M.P., Hood, D.W., Peak, I.R.A., Virji, M., and Moxon, E.R. (1995). Mol. Microbiol. *18*, 729–740.

Kim, J., Irvine, K.D., and Carroll, S.B. (1995). Cell *82*, 795–802.

Kukowska-Latallo, J.F., Larsen, R.D., Nair, R.P., and Lowe, J.B. (1990). Genes Dev. *4*, 1288–1303.

Lammers, G., and Jamieson, J.C. (1988). Biochem. J. *256*, 623–631.

Mandrell, R.E., McLaughlin, R., Kwaik, Y.A., Lesse, A., Yamasaki, R., Gibson, B., Spinola, S.M., and Apicella, M.A. (1992). Infect. Immun. *60*, 1322–1328.

Masteller, E.L., Larsen, R.D., Carlson, L.M., Pickel, J.M., Nickoloff, B., Lowe, J., Thompson, C.B., and Lee, K.P. (1995). Development *121*, 1657–1667.

Paulson, J.C., and Colley, K.J. (1989). J. Biol. Chem. *264*, 17615–17618.

Potter, M.D., and Lo, R.Y. (1995). FEMS Microbiol. Lett. *129*, 75–81.

Qian, M., Haser, R., Buisson, G., Duee, E., and Payan, F. (1994). Biochemistry *33*, 6284–6294.

Rost, B., Sander, C., and Schneider, R. (1994). Comput. Apll. Biosci. *10*, 53–60.

Saxena, I.M., Brown, R.M., Fevre, M., Geremia, R.A., and Henrissat, B. (1995). J. Bacteriol. *177*, 1419–1424.

Schuler, G.D., Altschul, S.F., and Lipman, D.J. (1991). Proteins Struct. Funct. Genet. *9*, 180–190.

Shur, B.D. (1977a). Dev. Biol. *58*, 23–39.

Shur, B.D. (1977b). Dev. Biol. *58*, 40–55.

Shur, B.D. (1993). Curr. Opin. Cell Biol. *5*, 854–863.

Strokopytov, B., Penninga, D., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L., and Dijkstra, B.W. (1995). Biochemistry *34*, 2234–2240.

Tatusov, R.L., Altschul, S.F., and Koonin, E.V. (1994). Proc. Natl. Acad. Sci. USA *91*, 12091–12095.

Wu, J.Y., Wen, L., Zhang, W.-J., and Rao, Y. (1996). Science *273*, 355–358.