## Correspondence

## The fukutin protein family – predicted enzymes modifying cell-surface molecules L. Aravind<sup>1,2</sup> and Eugene V. Koonin<sup>1\*</sup>

Fukuyama type congenital muscular dystrophy (FCMD) is an autosomal recessive disorder that is observed predominantly in Japanese populations [1]. Recently, the cause of this syndrome was discovered to be lesions in the gene encoding the protein fukutin; these lesions involve retroposon insertion and point mutations, which result in a truncated protein [2]. Consistent with the phenotypic patterns of the disease, the fukutin mRNA has been found in skeletal muscles, heart, brain and pancreas [2]. Fukutin contains a signal peptide and is localized to the Golgi and secretory granules [2]. Here, we report on a detailed computer analysis of the fukutin protein sequence, resulting

in the prediction that it is an enzyme that modifies cell-surface glycoproteins or glycolipids.

A gapped BLASTP [3] search of the non-redundant (NR) database at the National Center for Biotechnology Information using the fukutin sequence as the query revealed significant hits (with evalues  $< 10^{-9}$ ) not only to Caenorhabditis elegans proteins, such as T07D3.4 and T07A5.1, but also to the uncharacterized protein RP688 from the intracellular parasitic bacterium Rickettsia prowazekii  $(e < 10^{-4})$ . The RP688 sequence was used for further analysis of this protein family by iteratively searching the NR database using the PSI-BLAST program, which was run with the cut-off of e = 0.001 [3]. At convergence, not only fukutin and its C. elegans homologs, but also bacterial proteins involved in polysaccharide/phosphorylcholine modification and a yeast protein involved in mannosyl phosphorylation of oligosaccharides were retrieved from the database. Reverse searches with these sequences retrieved the original members of the fukutin family without any false positives. Fukutin, therefore, belongs to a family of

proteins associated with the modification of the cell surface.

A multiple alignment of the fukutin protein family was constructed using Gibbs sampling, as implemented in PROBE [4], in conjunction with the -m4 option of PSI-BLAST (Figure 1). The alignment shows prominent conservation in an amino-terminal block, followed by a weakly conserved carboxy-terminal region. The most notable feature of the amino-terminal region is the presence of the strictly conserved signature G[TS]hhGhhx4hhxaxxDxD (in single-letter amino acid code and in which 'h' is a hydrophobic amino acid, 'a' is an aromatic amino acid and 'x' denotes any amino acid). A pattern search with this motif recovers from the NR database the fukutin family in its entirety without any false positives. The carboxy-terminal region contains a motif with a conserved aspartate residue flanked by hydrophobic residues (Figure 1). Secondary structure prediction using the PHD program [5] suggests a compact  $\alpha/\beta$ fold for the fukutin domain (Figure 1).

LicD2 from *Streptococcus pneumoniae* is involved in the addition

## Figure 1

Secondary structure		.HHHHHHHHhheEEEeEEEeee
LicD2_Spn_5001693	14	LALLDYIDETCKKHDIPYFLSYGTMLGAIRHKGMIPWDDDIDISLYREDYERLLKIIEEENHPRYKVLSYDTSSWYFHNFASI-LDTSTVIEDHVKYKRHDTSLFIDVFPIDRFTDLSIVDKSYKYV 139
LicD1_Spn_5001692	14	LEILDYIDTLCKKHNINYIINYGTLIGAVRHEGFIPWDDDIDLSMPREDYQRFINIFQKEKS-KYKLLSLETDKNYFNNFIKI-TDSTTKIIDTRNTKTYESGIFIDIFPIDRFDD-PKVIDTCYKL-ESFKLLSFSKH 148
CpsG_Spn_3907605	1	mdaikefokickennidfflrg <mark>g</mark> svlgavkydgfipwddmdiavpregydklpgifkdriiagkyqvlayqycdtlhcyfprlfle <mark>de</mark> rkrlglprntnlglhlidiipldgapnhsflrklyfgkvyw <mark>y</mark> r 133
LicD_Hi_97170	14	LNILDYFHALCERHQIHYSLGGGTLIGAIRHKGFIPWDDDIDVYMHRDEYQRFVDVWFQETHEYYNIETVEDILAQKTGEMAKIFDCRTQITEIAGKKSAMFIDIFIYDGVPNDPKIICSFMKKYRRT 141
RP688_Rp_3861225	108	YQLMKDTHELLTKNNIKYWIESGTLLGAVRHQGIIPFDDDLDIGIMHEDEIHLQQILPQFEQLGYRVKHNKIYVICGERCLDIFLFHKEKDKFIHVIYDKYPNDFFYDHELYPLKKYKFG 227
ORF_Pi_2967843	32	kenlslikricnkynldfilff <mark>gtligavrehdfishdedidivmpisdle</mark> rfkdilfilrengfevar <mark>f</mark> ergfmsiirngeyidiyfftpyaedrrlstcicelcevkyinn 145
YJR061w_Sc_1176338	459	nSlirnfokfykangliswlshgtlygylydglkfpwdvdhdlqmpikhlhylsqyfnqsliledpregngrflldvgsaitvgvhgngennidarfididsgiyiditglsvssdaakqymskfy <mark>e</mark> e 586
MNN4_Sc_3915759	480	nsmirtfokftkangiiswlshgtlygylyngmafpwdndfdlqmpikhlqllsqyfnqsliledprqgngryfldvsdsltvringngknnidarfidvdtglyiditglastsapsrdylnsyize 607
T07D3.4_Ce_2315761	192	VEELAQLRDELIEFDMYPFLNGGTFLGWYRECSIIPHTTDMDLSVFAKDYNPIYVELLHSYWNPSSFEVWRMLGMVEDSFEIIQTKKWFEYPIDLFLMYEGIENGKLTHHWVGGIA-TDGTKYKFT 317
T07A5.1_Ce_3879520	472	aeklaefrdvlltfnmfaflng <mark>gt</mark> llgwyrecgfiphtadidlamfaedfhpeithfllsrtssfqllrslgmlndsyeltvtpktgyivnmdlflmykdvhkngsvinwvggaksrfc <mark>r</mark> s 592
Fuk_Hs_3370993	278	KELLQLAAKTLNKLGVPFWLSSGTCLGWYRQCNIIPYSKDVDLGIFIQDYKSDIILAFQDAGLPLKHKFGKVEDSLELSFQGKDDVKLDVFFFYEETDHMWNGGTQAKTGKKFKYL 393
Consensus/90%		.p.lhpp.h.p.shhl.GThlGhh+ppshIPas.DhDl.hpchpahhbDhDlhp
1		Current Biology

A multiple alignment of fukutin family proteins. The proteins designations include: gene name\_species abbreviation\_GenBank identification number. The CpsG protein is encoded by the operon involved in capsular polysaccharide biosynthesis in *S. pneumoniae.* The species abbreviations are as follows: *Spn, S. pneumoniae; Rp, Rickettsia prowazekii; Pi, Prevotella intermedia; Hi, H. influenzae; Sc,*  Saccharomyces cerevisiae; Ce, C. elegans; and Hs, Homo sapiens. The alignment is colored according to the 90% consensus with the following consensus terms: p (RKHEDNQST), polar; b (FLIMYWKREQ), big; s (STCVNDPGAS), small; h (LIVMFWYAC), hydrophobic; I (LVIMA), aliphatic; c (RKHED), charged; G,A or S, tiny; a (FYWH), aromatic; + (RKH), positive; S or T, alcohol. The coloring scheme used is: grey, charged; green, small; blue, hydrophobic/ aromatic; pink, tiny; yellow, alcohol; purple, polar; orange, big. Protein secondary structure was predicted using the PHD program [4], with the multiple alignment as the query, and is shown above the alignment (H indicates  $\alpha$  helix and e indicates  $\beta$  strand).

of phosphorylcholine residues to lipoteichoic acid, an important component of the cell wall in Gram-positive bacteria [6]. It has been proposed that LicD2 is an enzyme that catalyzes the transfer of phosphorylcholine to its teichoic acid substrates from a CDP-choline donor [6]. Similarly, the Haemophilus influenzae homolog - LicD is required for the addition of phosphorylcholine to the Gramnegative bacterial lipopolysaccharide [7]. The yeast protein MNN4 participates in a similar reaction in which a mannosyl phosphate residue is added from a GDP-mannose donor to both N-linked and Olinked oligosaccharides on proteins ([8] and references therein). These studies, taken together with the sequence conservation pattern, suggest that the fukutin family proteins are phosphoryl-ligand transferases. The presence of the DxD motif and a distal aspartate residue that is conserved in most of the sequences suggests that these enzymes coordinate a divalent cation, which is similar to a number of nucleotidyltransferases (for example, see [9]).

Drawing from knowledge of the functions of these proteins and the subcellular localization of fukutin, we predict that this protein modifies cell-surface molecules, most probably through the attachment of phosphoryl-sugar moieties. The brains of individuals with FCMD show an abnormal pattern of gangliosides [10], which suggests that the transferase activity of fukutin could participate in one of the many steps of glycolipid modifications. Glycolipid modifications are an important factor in determining the adhesive properties of cells [11]. Thus, fukutin might regulate neural migration and muscle organization by affecting the biogenesis of crucial adhesion molecules. In this regard, it might be of interest that disruption of the basement membrane, linked to an M-laminin

deficiency, has been reported in FCMD patients [12]. All the bacterial homologs of fukutin are from pathogenic bacteria and experimental evidence from H. influenzae [7] and S. pneumoniae [6] shows that the surface modifications catalyzed by these proteins (LicD and LicD2, respectively) are necessary for the adhesion of the bacteria to mammalian lung cells and pathogenesis. The sequence conservation described here, together with the apparent absence of fukutin family members in nonpathogenic bacteria, might suggest that the genes coding for these proteins have been horizontally transferred from eukaryotes to the pathogenic bacteria. A member of the fukutin family was detected in an expressed sequence tag from Trypanosoma brucei (AI077222), which suggests a possible role for this protein in the surface modifications of this eukaryotic pathogen (data not shown).

The proposed function and the phylogenetic distribution of the fukutin family members resemble those of the glycosyl transferases of the Fringe family that are involved in the modification of surface molecules of the Notch pathway in development [13]. Like the fukutin family proteins, Fringe homologs are seen in pathogenic but not in freeliving bacteria. Thus horizontal gene transfer from eukaryotes might be a common theme in the origin of the enzymes involved in surface modification in bacterial pathogens, resulting in similarities in adhesion mechanisms between these bacteria and eukaryotic cells.

## References

- Arahata K, Ishii H, Hayashi YK: Congenital muscular dystrophies. *Curr Opin Neurol* 1995, 8:385-390.
- Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-lida E, Nomura Y, et al.: An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature 1998, 394:388-392.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search

programs. Nucleic Acids Res 1997, 25:3389-2402.

- Neuwald AF, Liu JS, Lipman DJ, Lawrence CE: Extracting protein alignment models from the sequence database. Nucleic Acids Res 1997. 25:1665-1677
- Rost B: PHD: predicting onedimensional protein structure by profilebased neural networks. *Methods Enzymol* 1996. 266:525-539.
- Zhang JR, Idanpaan-Heikkila I, Fischer W, Tuomanen EI: Pneumococcal *licD2* gene is involved in phosphorylcholine metabolism. *Mol Microbiol* 1999, 31:1477-1488.
- Weiser JN, Shchepetov M, Chong ST: Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of *Haemophilus influenzae*. *Infect Immun* 1997, 65:943-950.
- 8. Jigami Y, Odani T: Mannosylphosphate transfer to yeast mannan. *Biochim Biophys Acta* 1999, **1426**:335-345.
- Aravind L, Leipe DD, Koonin EV: Toprim – a conserved catalytic domain in type IA and II topoisomerases, DnaGtype primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res* 1998, 26:4205-4213.
- Izumi T, Hara K, Ogawa T, Osawa M, Saito K, Novo ML, Fukuyama Y, Takashima S: Abnormality of cerebral gangliosides in Fukuyama type congenital muscular dystrophy. *Brain Dev* 1995, 17:33-37.
- Lloyd KO, Furukawa K. Biosynthesis and functions of gangliosides: recent advances. *Glycoconj J* 1998, 15:627-636.
- Xu H, Christmas P, Wu XR, Wewer UM, Engvall E: Defective muscle basement membrane and lack of M-laminin in the dystrophic dy/dy mouse. Proc Natl Acad Sci USA 1994, 91:5572-5576.
- Yuan YP, Schultz J, Mlodzik M, Bork P: Secreted fringe-like signaling molecules may be glycosyltransferases. *Cell* 1997, 88:9-11.

Address: <sup>1</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. <sup>2</sup>Department of Biology, Texas A&M University, College Station, Texas 77843, USA. E-mail: koonin@ncbi.nlm.nih.gov

The editors of *Current Biology* welcome correspondence on any article in the journal, but reserve the right to reduce the length of any letter to be published. All Correspondence containing data or scientific argument will be refereed. Items for publication should either be submitted typed, double-spaced to: The Editor, *Current Biology*, Current Biology Publications, 34–42 Cleveland Street, London W1P 6LE, UK, or sent by electronic mail to cbiol@currentbiology.com