FEBS Letters 580 (2006) 581–584

GG: A domain involved in phage LTF apparatus and implicated in human MEB and non-syndromic hearing loss diseases

Jinhu Guo[a,b,*], Haipeng Cheng[a], Shouyuan Zhao[b], Long Yu[b]

[a] *Department of Neurobiology, Pharmacology and Physiology, The University of Chicago, IL 60637, USA*
[b] *State Key Laboratory of Genetic Engineering, Institute of Genetics, Fudan University, Shanghai 200433, PR China*

Received 15 October 2005; revised 16 December 2005; accepted 22 December 2005

Available online 3 January 2006

Edited by Takashi Gojobori

Abstract Here, we report the identification of a novel domain – GG (domain in KIAA1199, FAM3, POMGnT1 and Tmem2 proteins, with two well-conserved glycine residues), present in eukaryotic FAM3 superfamily (FAM3A, FAM3B, FAM3C and FAM3D), POMGnT1 (protein O-linked mannose β-1,2-*N*-acetylglucosaminyltransferase), TEM2 proteins as well as phage gp35 proteins. GG domain has been revealed to be implicated in muscle–eye–brain disease and non-syndromic hearing loss. The presence of GG domain in Bacteriophage gp35 hinge connector of long tail fiber might reflect the horizontal gene transfer from organisms. And we proposed that GG domain might function as important structural element in phage LTF.

Keywords: GG domain; LTF apparatus; Muscle–eye–brain disease; Non-syndromic hearing loss; Gp35; POMGnT1; KIAA1199; FAM3 superfamily

1. Introduction

"Viruses straddle the definition of life. They lie somewhere between supra molecular complexes and very simple biological entities" [1]. Bacteriophage T4 apparatus containing the long-tail fibers (LTF) is responsible for host cell recognition and infection and initial attachment to susceptible bacteria [2,3]. Among the components of LTF, the distal half-fiber is composed of triple copies of gp34, gp36, gp37, respectively, as well as one copy of gp35 (30 kDa) [4]. In analyzing the protein sequence of gp35, we found a novel protein domain present in not only gp35 and but also a variety of eukaryotic proteins, including FAM3 superfamily, TMEM2 and POMGnT1.

2. Materials and methods

In analyzing the protein sequence of gp35, we found some homologues in similar phage strains. To find more homologues, we conducted PSI-BLAST searching at http://www.ncbi.nlm.nih.gov/blast/ [5] using the sequence of gp35 (gi|32753733, 52-157aa), with threshold value as 0.01. In the first two iterations, several phage proteins were retrieved, e.g., *Enterobacteria* phage JS98T4 gp35-like, tail fiber hinge (gi|52139849, *E*-value: 5e − 13), *Enterobacteria* phage RB69gp35 hinge connector of long tail fiber, proximal connector (gi|32753733, *E*-value: 3e − 50); *Enterobacteria* phage T4 gp35 hinge connector of long tail fiber, proximal connector (gi|5354251, *E*-value: 3e − 23), *Enterobacteria* phage T4 tail fiber protein gp35 - phage T4 (gi|2145006, *E*-value: 4e − 11); Enterobacteria phage gp35 tail fiber hinge (gi|66391730, *E*-value: 3e − 07), *Aeromonas* phage 31 gp35 (gi|62114858, *E*-value: 3e − 05); *Bacteriophage* 44RR2.8t hinge connector of long tail fiber proximal connector; gp35 (gi|34733001, *E*-value: 3e − 05), *Enterobacteria phage* RB49hinge long tail fiber protein proximal connector (gi|33348149, *E*-value: 9e − 05). From the second to ninth iterations, many eukaryotic proteins were retrieved, such as FAM3A (Family with sequence similarity 3, member A), FAM3B, FAM3C, FAM3D, TEM2, KIAA1199, POMGnT1 and some other uncharacterized proteins, including *Dictyostelium discoideum* hypothetical protein DDB0204607; gi|66812802, *D. discoideum* hypothetical protein DDB0204608. Additionally, exhaustive searches against all available genome and protein database at GenBank demonstrated that no homologue with significant *E*-value in plants, fungi, insects, Archaea and bacteria.

After 9 iterations, the results converged and retrieved 75 non-redundant protein sequences totally, which were subjected to multiple sequence alignment with ClustalX software [6] and manual editing (Fig. 1A and B, and supplementary materials), colored with Chroma [7]. The phylogenetic tree of these sequences was constructed with ClustalX software. This conserved region was named GG domain (Domain in KIAA1199, FAM3, POMGnT1 and Tmem2 proteins with two well-conserved glycine residues). The secondary structure of GG domain was predicted by Jpred using the alignment profile [8].

3. Results and discussion

GG domain is composed of seven β-strands and two α helices, about 100 amino acid residues in size. It is present in a wide range of proteins, including FAM3 superfamily (FAM3A, FAM3B, FAM3C and FAM3D), POMGnT1, TEM2, phage gp35 proteins and some uncharacterized proteins. Functional roles of FAM3 superfamily almost remain unknown, although some pilot work suggested FAM3B could represent a novel class of effectors involved in the destruction of the β-cells and involved in the pathogenesis of type 1 diabetes [9]. Human *TMEM2* is expressed in cochlea and many other tissues, it is located in the ARNSHL (autosomal recessive non-syndromic hearing loss) linked region on (chromosome 9q13–q21) but no disease-causing mutations were found in the coding region [10].

POMGnT1 is a glycosyltransferase, catalyzing the transfer of GlcNAc to O-mannose of glycoproteins. *POMGnT1* gene is implicated in muscle–eye–brain disease (MEB), caused by mutations in the *POMGnT1* gene, The frameshift mutations

*Corresponding author. Present address: Department of Physiology, ND13.200, University of Texas, Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, USA. Fax: +86 21 65642428 (PR China).
E-mail address: jhguo@fudan.edu.cn (J.Guo).

0014-5793/$32.00
doi:10.1016/j.febslet.2005.12.076

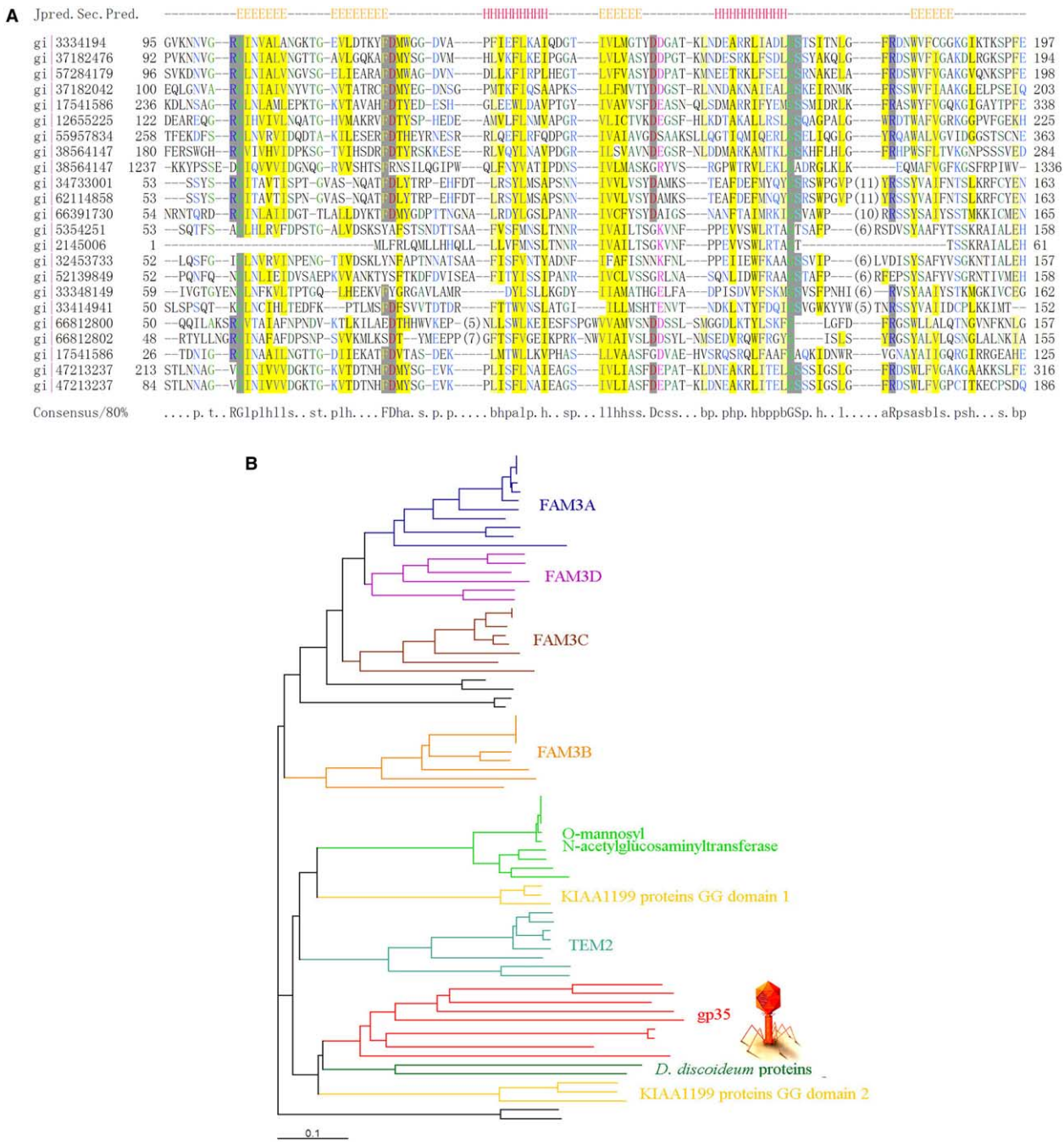

Fig. 1. Alignment (A) and phylogenic tree (B) of representative sequences with GG domain. In A, the sequences are: gi|3334194, *Homo sapiens* Protein FAM3C precursor; gi|37182476, *H. sapiens* FAM3D; gi|57284179, *H. sapiens* family with sequence similarity 3, member A; gi|37182042, *H. sapiens* FAM3B; gi|17541586, *Caenorhabditis elegans* putative protein family member, with a coiled coil-4 domain, of bilaterial origin (4D18); gi|12655225, *H. sapiens* O-linked mannose β-1,2-*N*-acetylglucosaminyltransferase; gi|55957834, *H. sapiens* transmembrane protein 2; gi|38564147, *H. sapiens* KIAA1199 protein; gi|34733001, *Bacteriophage* 44RR2.8t hinge connector of long tail fiber proximal connector; gp35; gi|62114858, *Aeromonas phage* 31 gp35; gi|66391730, *Enterobacteria phage* RB43 gp35 tail fiber hinge; gi|5354251, *Enterobacteria phage* T4 gp35 hinge connector of long tail fiber, proximal connector; gi|2145006, *Enterobacteria phage* T4 tail fiber protein gp35; gi|32453733, *Enterobacteria phage* RB69 gp35 hinge connector of long tail fiber, proximal connector; gi|52139849, *Enterobacteria phage* JS98 T4 gp35-like, tail fiber hinge; gi|33348149, *Enterobacteria phage* RB49 hinge long tail fiber protein proximal connector; gi|33414941, *Bacteriophage Aeh1* gp35 hinge long tail fiber proximal connector; gi|66812800, *Dictyostelium discoideum* hypothetical protein DDB0204607; gi|66812802, *D. discoideum* hypothetical protein DDB0204608; gi|17541586, *Caenorhabditis elegans* putative protein family member, with a coiled coil-4 domain, of bilaterial origin (4D18); gi|47213237, *T. nigroviridis* unnamed protein product. Fig. 1B is the Phylogenic tree of GG domain containing sequences retrieved by PSI-BLAST searches, KIAA1199 protein GG domain 1 denotes the N′ terminal GG domain in KIAA1199 and KIAA1199 protein GG domain 2 denotes the C′ terminal GG domain. Black branches represent unknown proteins difficult to be classified.

(281C > T and 541 del T) in POMGnT1 result in truncated proteins missing both the GG domain and the GNT-I family region (Pfam entry: PF03071) (Fig. 2), and some other mutations result in truncated proteins missing only the GNT-I family region [11]. GNT-I region occurs only in α-1,3-mannosyl-glycoprotein β-1,2-*N*-acetylglucosaminyltransferase family

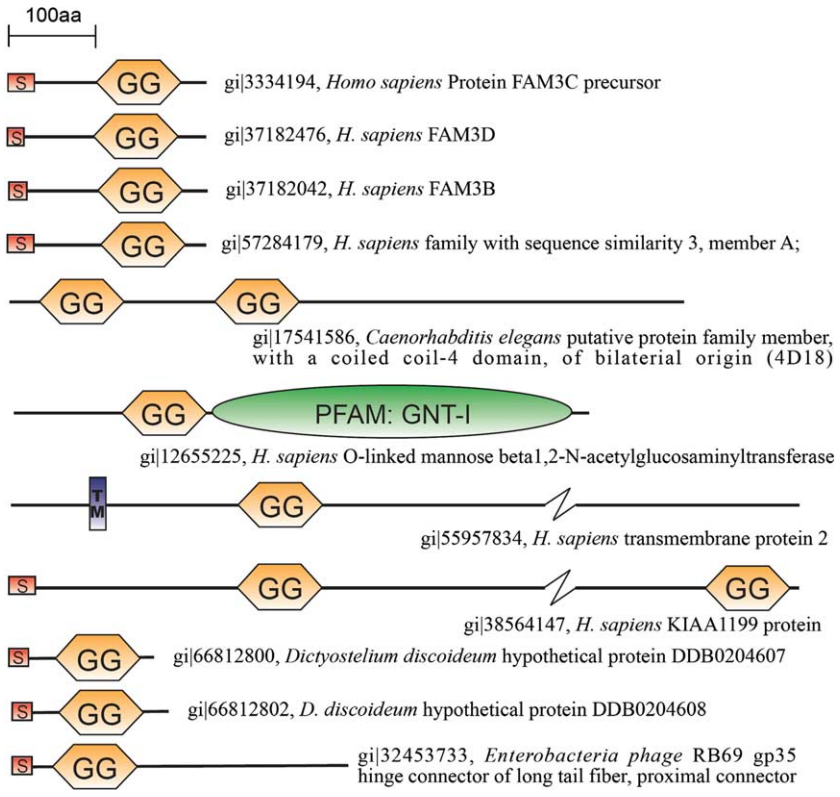

Fig. 2. Domain architecture of representative proteins with GG domain. S indicates signal peptide; TM indicates transmembrane region; PFAM:GNT-I: GNT-I family (PF03071). GG domain occurs in related proteins as singlet or two copies.

(GNT-I, GLCNAC-T I) which transfers *N*-acetyl-D-glucosamine from UDP to high-mannose glycoprotein *N*-oligosaccharide. The catalytic domain in proteins of this family is located at the C-terminus [12,13].

KIAA1199 protein contains two GG domains, and the phylogenetic tree indicated that these two GG domains were originated from separate combination events, instead of intragenic duplication. In the two GG domains, the N′ terminal one is more homologous to the phage gp35 proteins and *Dictyostelium* proteins (gi|66812800 and gi|66812802) (Fig. 1B). Murine *KIAA1199* is specifically expressed in Deiters' cells in the organ of Corti at postnatal day zero (Pn) P0 before the onset of hearing, but expression disappears by day P7 in those cells [14]. Abe et al. reported the R187C mutation of KIAA1199 protein in one family of non-syndromic hearing loss (2003), which is located at the N′ terminal GG domain (Fig. 2). Arginine is negative charged and hydrophilic while cysteine is neutral and hydrophobic. Mutation of this residue might have impact on the structure and function of GG domain.

Noticeably, GG domain also occurs in T4 type phage gp35 proteins, component of the long-tail fibers (LTF) hinge. LTF is an apparatus for Bacteriophage T4 to recognize and infect host cells, and it is also responsible for its initial attachment to susceptible bacteria [4,15]. Among the components of LTF, the distal half-fiber is composed of triple copies of gp34, gp36, gp37, respectively, as well as one copy of gp35 (30 kDa). Gp35 forms the local non-equivalent hinge (the "knee-cap") between proximal and distal half-fibers [4]. Since gp35 proteins contain only GG domain (Fig. 2), we proposed that GG might be an important structural element for the role of gp35 proteins as asymmetrical "knee-cap" in LTF apparatus.

Virions including phage were originated from organisms, among which bacteriophages infect Eubacteria and Archaea [1]. In *Enterobacteria* phages, GG is present exclusively in gp35 proteins of T4-like phages. According to the available data, GG domain does not occur in bacteria. Therefore, it is reasonable to propose that the sequence coding GG domain in *Enterobacteria* phages was obtained from an unidentified Eubacteria or Archaea species, instead of vertical inheritance from common ancestor. The phylogenetic tree suggested that the nucleic acid sequences encoding GG domain in phages might be derived indirectly from the ancestor of genomic sequences encoding *Dictyostelium* proteins (e.g., gi|66812800 and gi|66812802), as well as the sequences encoding the C′ terminal GG domain in KIAA1199 proteins (Fig. 1B and supplementary materials), prior to the transfer through the unknown Eubacteria or Archaea species.

Summarily, GG domain is a widely distributed protein motif, present in both eukaryotic proteins and T4 phage gp35 proteins. Mutations in KIAA1199 and POMGnT1 are associated with mammalian diseases including MEB and non-syndromic hearing loss, which suggest the physiological roles of GG domain. The GG domain in phage GG gp35 proteins might be crucially important structural element in LTF. And GG domain in the biological entities – T4 phages, might be acquired from organisms through horizontal gene transfer.

4. Supplementary data

Supplementary data for this paper are available online.

References

[1] DeFilippis, V.R. and Villarreal, L.P. (2001) Virus Evolution, Lippincott-Raven Publishers, Philadelphia, USA.

[2] Dickson, R.C. (1973) Assembly of bacteriophage T4 tail fibers. IV. Subunit composition of tail fibers and fiber precursors. J. Mol. Biol. 79, 633–647.

[3] Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V. and Rossmann, M.G. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell 118, 419–429.

[4] Cerritelli, M.E., Wall, J.S., Simon, M.N., Conway, J.F. and Steven, A.C. (1996) Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. J. Mol. Biol. 260, 767–780.

[5] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

[6] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

[7] Goodstadt, L. and Ponting, C.P. (2001) CHROMA: consensus-based colouring of multiple alignments for publication. Bioinformatics 17, 845–846.

[8] Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) JPred: a consensus secondary structure prediction server. Bioinformatics 14, 892–893.

[9] Zhu, Y., Xu, G., Patel, A., McLaughlin, M.M., Silverman, C. and Knecht, K., et al. (2002) Cloning, expression, and initial characterization of a novel cytokine-like gene family. Genomics 80, 144–150.

[10] Scott, D.A., Drury, S., Sundstrom, R.A., Bishop, J., Swiderski, R.E. and Carmi, R., et al. (2000) Refining the DFNB7–DFNB11 deafness locus using intragenic polymorphisms in a novel gene, TMEM2. Gene 246, 265–274.

[11] Manya, H., Sakai, K., Kobayashi, K., Taniguchi, K., Kawakita, M., Toda, T. and Endo, T. (2003) Loss-of-function of an *N*-acetylglucosaminylPOMGnT1, POMGnT1, in muscle–eye–brain disease. Biochem. Biophys. Res. Commun. 306, 93–97.

[12] Strasser, R., Mucha, J., Schwihla, H., Altmann, F., Glossl, J. and Steinkellner, H. (1999) Molecular cloning and characterization of cDNA coding for β-1,2-*N*-acetylglucosaminyltransferase I (GlcNAc-TI) from Nicotiana tabacum. Glycobiology 9, 779–785.

[13] Unligil, U.M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H. and Rini, J.M. (2000) X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. EMBO J. 19, 5269–5280.

[14] Abe, S., Usami, S. and Nakamura, Y. (2003) Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of non-syndromic hearing loss. J. Hum. Genet. 48, 564–570.

[15] Desplats, C. and Krisch, H.M. (2003) The diversity and evolution of the T4-type bacteriophages. Res. Microbiol. 154, 259–267.