Comparative Infectomic Analysis and Molecular Characterization of CglE, the Invasin IbeA Homologous Protein, Which Contributes to the Pathogenesis of Meningitic *E. coli* K1 Infection

Hong Cao, Lidan Chen, Jun Yang, Shuji Gong, Hao Zhou, Sheng-He Huang
Department of Microbiology, Southern Medical University,
Guangzhou, Guangdong 510515, China

and

Yong Jiang
School of Basic Medicine, Southern Medical University,
Guangzhou, Guangdong 510515, China

and

Sheng-He Huang
Childrens Hospital Los Angeles, the University of Southern California,
Los Angeles, CA 90027, USA

ABSTRACT

CglE is a putative dihydrolipoamide dehydrogenase (DLDH) that shares significant protein sequence homology (50% identity and 70% similarity) to the IbeA invasin contributing to the pathogenesis of *E. coli* meningitis. However, the biochemical, biological and pathogenic functions of CglE are unknown. In order to characterize the role of CglE in the pathogenesis of meningitic infections, infectomic, bioinformatics and molecular approaches were used to analyze and predict its structure and function. **First**, our comparative infectomic studies showed that CglE and IbeA are present in the genetic island GimA as a pair of homologous proteins that are encoded by two different operons, *cgl* (GimA2) and *ibe* (GimA4), at different locations. A similar pair of proteins is also present in *Silicibacter sp* which belongs to the most abundant and ecologically relevant marine bacterial groups. Meningitic *E. coli* K1 and *Silicibacter sp* have to survive under harsh environments (cerebrospinal fluid and ocean) with poor nutrition, suggesting that this pair of proteins is important for energy metabolism in the both microbes. **Secondly**, bioinformatic analysis indicated that CglE is a putative DLDH, which is the E3 component of pyruvate dehydrogenase complex. A FAD-binding domain and homologous flavoprotein regions are present in CglE. DLDH has been identified as virulence factors contributing to the pathogenesis of hepatitis C virus and pneumococcal infections. The sequence of CglE is homology to that of IbeA, an invasion protein of meningitic *E. coli,* and the surface probability and hydrophilicity are very similar to each other. **Thirdly**, the expression, purification and biochemical analysis of CglE protein was further carried out to determine whether CglE shares similar functions as IbeA and whether it is a new class of DLDH. The *cglE* gene was amplified by PCR and subcloned into pET22b(+) then expressed in *E. coli* BL21(DE3) as a fusion protein with His6 tag at its C-terminus induced by IPTG. CglE fusion protein was purified. Like IbeA, CglE is able to bind to an IbeA-binding protein, vimentin. In further studies, we will examine whether CglE is a new class of DLDH and how it contributes to the pathogenesis of meningitic infections.

Keywords: *E. coli* K1, meningitis, IbeA, CglE, pathogenesis, infectomics, bioinformatics

1. INTRODUCTION

Bacterial meningitis is one of the most serious infections

of the central nervous system (CNS) with high morbidity and mortality. *Escherichia coli* is the most common gram-negative bacillus that causes meningitis during the neonatal period, but it is not clear how circulating *E coli* could invade the brain microvascular endothelial cells (BMECs) and cross the blood-brain barrier (BBB) [1-4].

Meningitic *E.coli* K1 possess traits that distinguish them from other pathogenic and commensal strains of *E. coli*. Characteristically, meningitic strains of *E.coli* are composed of a restricted number of O serogroups (O1, O2, O7, O18 and O83), produce S fimbriae, express *ibeA* and are predominately carrying K1 capsule (over 84%). These features imply that meningitic strains possess a defined set of virulence determinants that allow the bacterium to penetrate the BBB and enter the CNS. Such clusters of virulence genes, termed genetic islands, are present in meningitic pathogens *Nersseria meningitidis* and *Haemophilus influenzae*. A 20.3-kb **g**enetic **i**sland of **m**eningitic *E. coli* containing *ibeA* (GimA) [5] has been found to be present in *E. coli* K1 strains but absent in *E. coli* K12 strains. The GimA island consists of 15 genes that form 4 operons: *ptnIPKC*, *cglDTEC*, *gcxKRCI* and *ibeRAT*. The G+C content (46.2%) of GimA is substantially different from that (50.8%) of the rest of the *E. coli* chromosome. GimA encodes 15 proteins in which IbeA and IbeT are invasion protein mediated penetration of the BBB [6] and most of other proteins contribute to substrate transportation and carbon source metabolism. CglE is a putative dihydrolipoamide dehydrogenase (DLDH) as it shares sequence homology to IbeA and DLDH of various species from bacteria to human [5]. However, the biochemical, biological and pathogenic functions of CglE are unknown. In order to characterize the role of CglE in the pathogenesis of meningitic infections, infectomic, bioinformatics and molecular approaches were used to analyze and predict its structure and function.

2. MATERIALS AND METHODS

Bacterial strains, culture conditions, plasmids and media

pUC10-8 was a plasmid containing *cglE* [5]. BL21(DE3) and expression vector pET22b(+) were used for protein expression. DH5a was used as the host strain in subcloning and in preparing the plasmids for DNA sequence determination. BL21(DE3) was the host strain for protein expression of *cglE*. Strains containing plasmids were grown at 37℃ in Luria broth (LB; 10 g of tryptone, 10g of NaCl, and 5g of yeast extract/L) with ampicillin (100 mg/mL) for the positive selection.

Chemicals, enzymes and antibodies

Taq DNA polymerase, DNA restriction enzymes (BamH I & Hind III), and plasmid isolation kit were the products of TaKaRa. pGEM-T vector and T4 DNA ligase were from Promega. DNA Marker (λ DNA/Hind III), Protein Marker, Bovine serum albumin (BSA), EDCI were from Sigma. All other chemicals were local products of analytic grade. Mouse anti-His tag antibody, mouse anti-BSA antibody, goat anti-mouse IgG conjugated to horseradish peroxidase (HRP) and goat anti-rabbit IgG conjugated HRP were purchased from Sigma.

Infectomic and bioinformatics approaches

In order to identify *E. coli* structures that contribute to the invasion of the BBB, Tn*phoA* transposon mutagenesis was performed with a rifampin-resistant strain derived from RS218 (O18:K1:H7), an invasive *E. coli* K1 strain E44 for generating a collection of mutants. A genetic island named GimA including the *ibeA* gene was identified. Functional annotation of GimA has been performed by combination of bioinformatics and comparative infectomic approaches. A Hidden Markov Model of the GeneMark finding program was used to predict unknown genes in GimA. Gene starts and translation initiation sites in a DNA sequence packed with prokaryotic genes were specifically targeted by using this approach. The models of protein coding regions, the second order Markov models and the RBS model were determined by the GeneMarkS training program. Other types of DNA sequence analysis were made with program developed by Durham N.C software. DNA and deduced protein sequences were used to search the DNA and protein databases at NCBI/NIH using BLAST and PSI-BLAST. Sequences are compared to existing sequences with known functions at the protein level to identify homologous regions and then to predict the functionality of the novel genes. Multiple alignment of protein

sequences was done with Clustal W and Boxshade programs. Phylogenetic analysis was made by using BioNavigator (http://www.bionavigator. com). Therefore, detection and functional annotation of novel genes from the genetic island GimA made use of infectomic and several conceptually different approaches, such as Hidden Markov model based gene finding, sequence comparison, motif finding and phylogenetic analysis.

Construction of expression plasmid pET4C

The *cglE* gene was amplified with PCR from pUC10-8 using the primer set: P1(*Bam*H I) 5'gccggatccaatggtgg-acatgattaatg 3' and P2(*Hin*d III) 5' ctaaagctttaatgccagac-gaacgccc 3'. The condition for PCR was as follows: incubate at 94℃ for 2 min., followed by 30 cycles of 1 min at 94℃, 30s at 60℃, 1 min at 72℃; and 10 min at 72 ℃ in the last cycle. The PCR-amplified fragment of about 1.36kb was subcloned into a pGME-T vector to form the pTCE construct. The 1.36kb fragment was excised from pTCE construct with BamH I and Hind III and ligated into the equivalent sites of pET22b(+) vector to form the new plasmid pET4C which was transformed into *E.coli* BL21(DE3). The resultant construct was confirmed by enzyme digestion and sequence analysis.

Expression and purification of CglE protein

Positive colony was chosen and cultured overnight in LB medium with 100mg/L ampicilin, and then inoculated into fresh LB medium at 37ºC until its A_{600} reached 0.6, protein expression was induced by adding isopropyl β-*D*-thiogalactopyranoside (IPTG; 1m*M*). Bacteria were lysed in 2×SDS loading buffer for SDS-PAGE analysis. The expressed proteins was determined by Western blot, using 2000 fold His-tagged antibody and HRP conjugated goat-anti-mouse IgG.

For the preparation of the fusion proteins, the bacteria were harvested by centrifugation (4000 rpm for 15 min at 4ºC) and resuspended in 1.25ml BugBuster Protein Extraction Reagent (Novagen). The soluble proteins were obtained by centrifugation at 16,000*g*, 4 ºC for 20 min., and the pellet was washed again with BugBuster buffer. Then the protein preparations isolated from the supernatants and pellets were separately run on 10% polyacrylamide gel with SDS. The results indicated that CglE predominately resided in the insoluble fraction. The insoluble protein was dissolved in 8 *M* urea (20m*M* Tris-HCl pH 8.0, 0.5 *M* NaCl), and diluted 30 times by refolding buffer (20m*M* Tris-HCl pH 8.0, 0.5 *M* NaCl, 15%glycerol). CglE protein with a histidine tag was purified by binding to Ni-NTA resin, according to the manufacturer's instructions (Novagen).

DNA sequencing and structure analysis of *cglE*

The DNA and deduced protein sequences of *cglE* were searched in the DNA and protein databases at NCBI/NIH by using BLAST and PSI-BLAST. Similar sequences were found and compared with the existing sequences with known functions at the protein level by using the software of DNAMAN. Putative structure was compared to IbeA by using the software of DNASTAR.

3. RESULTS

Comparative infectomic studies of CglE

The genetic island GimA consists of four operons, *ptnIPKC*(GimA1), *cglDTEC* (GimA2), *gcxKRCI*(GimA3) and *ibeRAT*(GimA4). The operon *cgl*(GimA2) presumably contributes to glycerol metabolism, whereas the operon *ibe*(GimA4), which codes for an invasin (IbeA) and a putative regulatory protein (IbeR), may contribute to invasion and regulation. Our comparative infectomic studies showed that CglE and IbeA are present in GimA as a pair of homologous proteins that are encoded by two different operons, *cgl* (GimA2) and *ibe* (GimA4), at different locations. The functionality of both genes was initially defined by comparing the deduced their open reading frames (ORFs) to existing sequences with known functions at the protein level, and the results showed that a similar pair of proteins is also present in *Silicibacter sp* which belongs to the most abundant and ecologically relevant marine bacterial groups [15].

Bioinformatic analysis of CglE

As suggested by bioinformatics analysis of the sequences, CglE is a putative DLDH since it shares sequence homology to DLDH of various species from bacteria to human (Fig.1). By using BLAST in protein databases at NCBI/NIH, we know that this protein contains pyridine nucleotide- disulphide oxidoreductase site, which

includes both class I and class II oxidoreductases and also NADH oxidases and peroxidases, and a Lpd site which is a Pyruvate/2-oxoglutarate dehydrogenase complex including a dihydrolipoamide dehydrogenase (E3) component. A FAD-binding domain and homologous flavoprotein regions are present in the very N-terminal region of CglE.

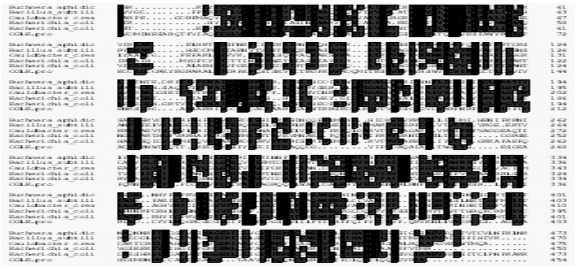

Fig.1 Sequences comparison between CglE and known DLDH

We found that the sequence of *cglE* is homologous to *ibeA*, and both sequences homology is over 50%, and both proteins are outer membrane protein and highly analogical in surface probability (Fig.2) and hydrophilicity (Fig.3).

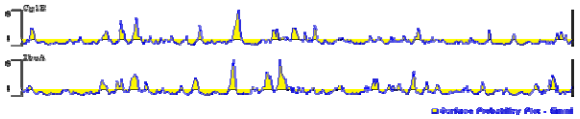

Fig.2 Comparison of surface probability between and CglE and IbeA

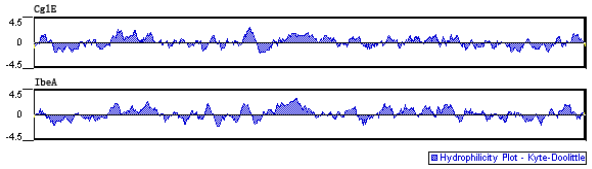

Fig.3 Comparison of hydrophilicity between CglE and IbeA

Identification of recombinant plasmid pET4C

The full-length *cglE* gene from pUC10-8 amplified by PCR was subcloned into pET22b(+) to make construct pET4C. The BamH Ⅰ - HindⅢ fragment from pET4C which encode CglE, was eluted from 1 % agarose gel after disgestion (Fig.4).

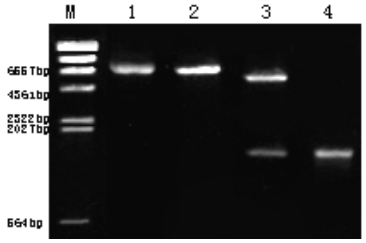

Fig.4 Restriction endonucleases and PCR analysis of recombinant plasmid pET4C

M::λ Hind III DNA marker; Lane 1:pET4C; Lane 2: pET4C/BamH Ⅰ; Lane 3: pET4C/BamH I + Hind III; Lane 4: pET4C/PCR

Expression and characterization of CglE

The plasmid pET4C encodes a fusion protein with 6 histidine residues at its C-terminus. Expression from pET4C was induced with IPTG. In SDS-PAGE, a prominent band of approximately 50-kD was observed (Fig.5, lane 1), matching the predicted molecular weight. Western blot analysis of expressed protein CglE from recombinant BL21 (DE3) is also shown in Fig.5.

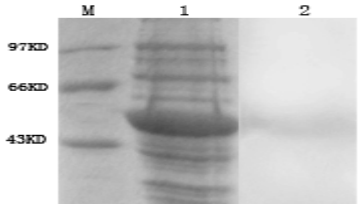

Fig.5 Detection of CglE by SDS-PAGE and Western-blotting

M: protein low molecular markers; Lane 1:CglE by SDS-PAGE; Lane 2:purified CglE binds to IbeA polyclonal antibody by WB

4. DISCUSSION

In the present studies on meningitic *E. coli* infection, it is known that several genes may contribute to *E.coli* invasion of BMECs and cross the BBB [7]. In the GimA islands of meningitic *E.coli*, fourteen novel genes have been identified in addition to the *ibeA*, the functions of the gene-coding proteins were assigned to the functional categories of proteins relating to carbon source metabolism and substrate transportation [5]. Functional annotation of CglE has been performed by combination of imfectomic, bioinformatics and molecular approaches in this study. First, the functionality of the *cglE* gene was initially defined by comparing the deduced open reading frames (ORFs) of CglE to existing sequences with known

functions at the protein level. The results showed that CglE and IbeA are present at different locations of the genetic island GimA as a pair of homologous proteins that are encoded by two different operons, *cgl* (GimA2) relating presumably to glycerol metabolism and *ibe* (GimA4) contributing to invasion and regulation [5]. CglE is a putative dihydrolipoamide dehydrogenase (DLDH) that shares significant protein sequence homology (50% identity and 70% similarity) to the IbeA invasin contributing to the pathogenesis of *E. coli* meningitis. Glycerol may play a role in regulation of invasion genes since glycerol has been shown to be an important signal in regulation of virulence gene expression in *Staphylococcus aureus*[8,9]. It may interpret that the carbon source metabolism such as glycerol during bacterial penetration process was required by the bacterial pathogenesis. A similar pair of proteins is also present in *Silicibacter sp* which belongs to the most abundant and ecologically relevant marine bacterial groups [15]. Meningitic *E. coli* K1 and *Silicibacter sp* have to survive under harsh environments (cerebrospinal fluid and ocean) with poor nutrition, suggesting that this pair of proteins is important for energy metabolism in the both microbes.

As suggested by our bioinformatics analysis, CglE contains several domains such as Pyr、Lpd、NADH domains which are the same as the primary structures of all other DLDH that have been published [10], and also contains the FAD-binding domain. It is known that the sequence of DLDH currently included an FAD-binding domain, a conserved pyridine redox active site, and an NAD-binding domain [11-13]. DLDH is the E3 component of pyruvate dehydrogenase complex that plays an important role in energy metabolism. Since DLDH has been identified as virulence factors contributing to the pathogenesis of hepatitis C virus and pneumococcal infections, whether CglE is a new class of DLDH and how it contributes to the pathogenesis of meningitic infections are further investigated.

Furthermore, the sequences homology is over 50% (56% identity) between CglE and IbeA, both proteins are outer membrane protein and highly analogical in surface probability, hydrophilicity, transmembrane probability and antigenic index, which suggest both proteins share the comparability in structure and function [6,14]. The relevant functionalities of CglE were then confirmed by experimental approaches. In our experiment, the purified CglE protein can binding with the polyclonal antibody and binding protein of IbeA, their interactions indicate that CglE have the same antigenicity and binding affinity as IbeA. As showed by the experimental data, the structures and functions for both CglE and IbeA greatly matched the analysis and prediction by infectomics and bioinformatics. Whether the function of CglE is similar to that of IbeA will be further addressed by detecting the invasion ability of the *cglE* mutant and the ability of the purified CglE protein to block the invasion.

5. ACKNOWLEDGEMENTS

This work was supported by Guangdong Natural Science Foundation C020094.

6. REFERENCES

[1] S.H. Huang, M. Stins, K.S. Kim, “Bacterial penetration across the blood-brain barrier during the development of neonatal meningitis”, **Microbes Infect**., Vol.2, No.5, 2000, pp.1237-1244.

[2] S.H. Huang, A. Jong, “Cellular mechanisms of microbial proteins contributing to invasion of the blood-brain barrier”, **Cell Microbiol**., Vol.3, No.5, 2001, pp.277-287.

[3] K.S. Kim, “*E. coli* invasion of brain microbascular endothelial cells as a pathogenetic basis of meningitis”, **Subcell Biochem**., Vol.33,2000,pp.47-59.

[4] X. Saez-Llorens, G.H. McCracken Jr, “Bacterial meningitis in children”, **Lancet**, Vol.361,2003, pp.2139-2148.

[5] S.H. Huang, Y.H. Chen, G.Y. Kong, “A novel genetic island of meningitis *Escherichia coli* K1 containing the ibeA invasion gene (GimA): functional annotation and carbon-source-regulated invasion of human brain microvascular endothelial cells”, **Funct Integr Genomics**. Vol.1, 2001, pp.312-322.

[6] S.H. Huang, Z.S. Wan, Y.H. Chen, “Further characterization of *Escherichia coli* brain endothelial

cell invasion gene *ibeA* by deletion, complementation and protein expression", **J Infect Dis**., Vol.183, 2001, pp.1071-1078.

[7] Chen, L., Cao, H. Progress in Meningitic *E. coli* crossing the blood-brain barrier. **Chin J. Zoon**. Vol.21, No.5, 2005, pp.426-428.

[8] J.L. Smith, M.M. Bencivengo, C.A. Kunsch, "Enterotoxin A synthesis in *Staphylococcus aureus*: inhibition by glycerol and maltose", **J. Gen. Microbiol**., Vol.132, 1986, pp.3375-3380.

[9] A. Ruzin, R.P. Novick, "Equivalence of lauric acid and glycerol monolaurate as inhibitors of signal transduction in *Staphylococcus aureus"*, **J Bacteriol.,** Vol.182, 2000, pp.2668-2671.

[10] A. W. Smith, H. Roche, M. C. Trombe, "Characterization of the dihydrolipoamide dehydrogenase from *Streptococcus pneumoniae* and its role in pneumococcal infection", **Molecular Microbiology**. Vol.44, 2002, pp.431-448.

[11] M. S. Patel, L. G. Korotchkina, "The biochemistry of the pyruvate dehydrogenase complex", **Biochemistry and Molecular Biology Education,** Vol.31, 2003, pp.5-15.

[12] R.N Perham, H.W. Duckworth, "Mobility of polypeptide chain in the pyruvate dehydrogenase complex revealed by proton NMR", **Nature**, Vol.292, 1981, pp.474-477.

[13] N. Kruger, F.B.Oppermann, H. Lorenzl, "Biochemical and Molecular Characterization of the Clostridium magnum Acetoin Dehydrogenase Enzyme System", **J Bacteriol.**, Vol.176, 1994, pp.3614-3630.

[14] M. Zou, L. He, C.H. Wu, H. Cao, Z.H. Xie, Y. Ouyang, Y. Wang, A. Jong, S.H. Huang. "PSF is an IbeA-binding protein contributing to meningitic *Escherichia coli* K1 invasion of human brain microvascular endothelial cells", **Med Microbiol Immunol.**, 2007. In press.

[15] M.A. Moran, A. Buchan, J.M. Gonzalez, *et. al*. "Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment", **Nature**, Vol.432, 2004, pp.910-913.