Molecular cloning, sequence analysis and homology modeling of galE encoding UDP-galactose 4-epimerase of Aeromonas hydrophila

Shivani Agarwal 1, Keshav Gopal 1, Gagan Chhabra 2, Aparna Dixit 1,∗

1 Gene Regulation Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi – 110067, India; 2 Gene Regulation Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi -110067, India; Aparna Dixit - Email: adixi2100@gmail.jnu.ac.in; Phone: +91-11-26742164; Fax: +91-11-26704085; ∗Corresponding author

Received October 24, 2009; Accepted November 15, 2009; Published November 17, 2009

Abstract:
Aeromonas hydrophila, a ubiquitous gram-negative bacterium present in aquatic environments, has been implicated in illness in humans, fish and amphibians. Lipopolysaccharides (LPS), a surface component of the outer membrane, are one of the main virulent factors of gram-negative bacteria. UDP-galactose 4-epimerase (GalE) catalyses the last step in the Leloir pathway of galactose metabolism and provides precursor for the biosynthesis of extracellular LPS and capsule. Due to its key role in LPS biosynthesis, it is a potential drug target. The present study describes cloning, sequence analysis and predicted 3D-structure revealed that the GalE of A. hydrophila consists of the signature sequences of the epimerase super family. The present study reports the molecular modeling / 3D-structure prediction of GalE of A. hydrophila. Further, the potential regions of the enzyme that can be targeted for drug design are identified.

Keywords: Aeromonas hydrophila, lipopolysaccharide, virulence, UDP-galactose 4-epimerase, molecular phylogenetic.

Abbreviations: UDP-GlcNAc, UDP-N-acetylglucosamine; GalE, UDP-galactose 4-epimerase; LPS, Lipopolysaccharides.

Background:
Aeromonas hydrophila is a member of the family Aeromonadaceae, associated with disease conditions mainly in fish, amphibians and humans [1]. Identification of strains of A. hydrophila capable of causing illness in apparently healthy individuals, by infecting open wounds and possibly by ingestion of the microorganism in food or water, has generated immense interest in this organism [2]. Some of the known virulence factors responsible for pathogenesis of A. hydrophila are O-antigen lipopolysaccharide, capsules, exotoxins, enterotoxins, and certain exoenzymes [3 & 4]. LPS has been reported to be involved in adherence and may play a role in antigenic variation [5-7]. Importance of the enzymes involved in galactose metabolism in bacterial virulence has been demonstrated [8-10]. GalE is one of the enzymes involved in galactose metabolism that mediates the incorporation of galactose in extra cellular polysaccharide material such as the O-side chain of lipopolysaccharide. The essential role of UDP-galactose 4-epimerase in virulence of many other gram-negative bacteria is well documented [11-15]. The fact that the epimerase mutants have altered LPS core biosynthesis with significant reduction in their ability to adhere and invade the host cells makes epimerase a potential drug target. GalE from different species exhibits a significant degree of interspecies variation at their gene and quaternary structure. In the present study, we report cloning, characterization of the galE, including its putative promoter, and structure modeling of the deduced amino acid sequence of the GalE of A. hydrophila

Methodology:
Bacterial strains and vector:
A. hydrophila (AH17) isolated from pond water was obtained from Dr. I. Karunasagar, College of Fisheries, Mangalore, India. Escherichia coli DH5α and BL21 (DE3) strains were from GIBCO BRL, USA and Novagen, USA, respectively. Plasmid pBCKS+ was procured from Stratagen (USA).

Cloning and sequencing of galE of A. hydrophila:
Genomic DNA from A. hydrophila (AH17) was isolated essentially as described earlier [16]. The galE of A. hydrophila was PCR amplified using the genomic DNA as a template and the forward and reverse primers (5′-AGTCTGAGAAAAACGCGCTGTG-3′, 5′-TTAATCGGGATATCCCTGTGGATGG-3′, respectively), designed on the basis of available sequence information of galE of E. coli (Acc. No. NC_009093) obtained from Mirosynth, Switzerland. The PCR amplified product was purified and the ends of purified PCR product were phosphorylated using T4 polynucleotide kinase (NEB, USA), followed by ligation to Smal I digested dephosphorylated pBCKS (+) vector. Competent E. coli DH5α cells (Novagen, USA) were then transformed with the ligation mix and the transformants were analyzed by colony PCR and were further confirmed by restriction enzyme digestion for the release of the insert. The construct thus made was designated as pAHGauE. The integrity of the galE insert was verified by automated DNA sequencing (Applied Biosystem Model 393A).

Phylogenetic analysis:
Sequence analysis tools of the ExPAsy Molecular Biology Server of Swiss Institute of Bioinformatics were used to process nucleic acid sequence for the deduced amino acid sequence. The deduced amino acid sequence of the GalE of A. hydrophila was aligned with the GalE of other species and A. hydrophila AH17 strain was carried out using ClustalW (Version 1.83) [17]. The phylogenetic tree was inferred using Phylip’s inference package, Version 3.5c.
Structure Modeling and Visualization of Model:

The most appropriate template for Homology modeling of *A. hydrophila* GalE (Accession No. AJ785765) was identified using BlastP analysis. The available structure of GalE from *Escherichia coli* in the Protein Database (PDB) (PDB entry 1udcA, resolution =1.65, R value=0.177) was referred [18]. The target and the template sequences were aligned using ClustalW. Homology modeling program Swiss-Model was employed to generate a comparative 3D- structure model of *A. hydrophila* GalE [19]. Swiss-Model [20] is a server for automated comparative modeling of three-dimensional (3D) protein structures. No other refinements were applied. Swiss PDB viewer software [21] was employed as a tool to envisage the generated structural model.

Validation of the generated model:
The generated 3D-model was assessed/reviewed evaluated at various structure verification servers viz. PROCHECK [22] that relies on Ramachandran plot [23], WHAT_CHECK, a subset of WHATIF programme [24 & 25], and VERIFY3D [26 & 27].

---

**Figure 1:** Nucleotide and deduced amino acid sequence of cloned *galE* of *A. hydrophila*. The open reading frame encodes for a protein of 338 amino acid residues. Initiation and termination codons are shown in bold. Putative -35 and -10 regions are shown as bold and underlined nucleotides. The shaded and the boxed bases represent putative ribosome binding site and the GalR binding site, respectively. Nucleotides pertaining to the primers used for amplification of the gene are underlined.
Discussion:

Sequence analysis of the \textit{galE} of \textit{A. hydrophila}:

Sequencing of the PCR amplified cloned \textit{galE} fragment revealed the insert to be of 1140 bp, representing full length \textit{galE} and its promoter-operator sequences (Figure 1). Sequence analysis revealed the presence of putative RNA polymerase binding site (-35 region) at 57-65 bp and a ribonuclease box (-10 region) at 76-84 bp. Putative binding site for catabolite repressor protein or cyclic AMP receptor protein (CRP) is also present at 49-69 bp, overlapping -35 region. The binding site for the GalR overlaps the translation start site and is present at 131-139 bp. Presence of all the regulatory sequences and components of a promoter upstream of the \textit{GalE} encoding region of the cloned fragment indicate that the organization of \textit{galETK} operon of \textit{A. hydrophila} is similar to that of other gram-negative bacteria, which is
organized in the order of galE, galT and galK. Open reading frame of the cloned galE contains a single protein translation start site ATG at 124-126 bp and a termination codon, TAA, is present at the 1138-1140 bp (Figure 1). Putative ribosome binding site is located 7 bp upstream of the ATG at 112-117 bp position. The encoded protein is of 338 amino acid residues with a theoretical pi of 5.64 and molecular weight of 36501.36. Genome database search (Blalt N) showed varying degrees of similarity to nucleotide sequence of the galEs of other species. Blast P of the deduced amino acid sequence showed that the GalE of A. hydrophila shares the percentage identities ranging from ~60-95% (95% with Shigella boydii and 63% with Photobacterium profundum) with different species of bacteria. It is of interest to note that while the GalE of the A. hydrophila AH17 shows significant identity with other bacteria, it showed only 59% identity and 85% similarity with the GalE of another strain of A. hydrophila, AH3 [15], though the active site and the catalytic sites/residues have remained conserved between the two. Bacteria of Aeromonas spp. are highly heterogeneous group of bacteria, and the differences in the GalEs of the two Aeromonas hydrophila strains may only be an indication of heterogeneity.

Phylogenetic analysis:
Analysis of amino acid sequence alignment of the GalE (Figure 1 in supplementary material) revealed that the A. hydrophila GalE consists of two characteristic Tyr-X-Lys-Lys couple (positions 140, 225 and 227) showing an identity of 95%, whereas with D. rerio (94%) and H. sapiens (93%), it exhibited only 51% and 53% identities with that of A. hydrophila, respectively, in which it is an important disease causing bacteria, wherein structural motifs have been identified to remain conserved. Since A. hydrophila has also been reported to infect humans, it is important to compare the depicted motif with that of some Homo sapiens (4-Epimerase), with which it shows only 51% identity. A superimposition of the A. hydrophila GalE onto the human epimerase monomer along with UDP-GlcNAc and NADH is shown in Figure 5.

Superimposition of modeled structure of A. hydrophila GalE onto a Homo sapiens GalE monomer (subunit A of PDB entry 1HZJ) matches residues Ala305, Ala306 and Cys307 with a beta strand (B12) in C-terminal domain, whereas corresponding residues Pro297, Ala298 and Tyr299 of A. hydrophila GalE form coiled structure. When comparing A. hydrophila GalE to Homo sapiens GalE, and ignoring single amino acid differences, a stretch of six amino acids GGSLPE make a loop between ß12 and ß2 in Homo sapiens GalE, which is absent in A. hydrophila GalE. Moreover, ß2 of A. hydrophila GalE adopts a slightly different orientation compared with Homo sapiens GalE. A. hydrophila GalE also differs from that of fish species (D. rerio) by about 47%, therefore, the regions of differences between the two, can be targeted for drug design against the pathogen enzyme.

Supersimposition of the catalytic site of the Homo sapiens and A. hydrophila GalE along with the UDP-GalNAc and NADH is shown in Figure 5B. It is well known, that in addition to catalyze the interconversion of UDP-galactose and UDP-glucose, the human epimerase is also capable of interconverting UDP-GalNAc and UDP-GlcNAc [30]. Markedly, E. coli epimerase has not been reported for this activity. It is clear from the superimposition of A. hydrophila GalE onto Homo sapiens GalE (Figure 5B) that substitution of Cys307 in Homo sapiens GalE. It can be suggested that the substitution of more bulky Tyr299 in the A. epimerase with a Cys307 in the human epimerase most likely inhibit UDPGalNAc from binding in the A. hydrophila GalE active site as has been reported for the E. coli GalE. These positions can be taken into consideration for designing suitable inhibitors against A. hydrophila GalE, hydrophila exhibited greater degree of differences between the hosts, fish and human. Structure modeling of the A. hydrophila GalE resulted in identification of the structural differences between the GalE of the host and the pathogen. These differences can be targeted for drug design against the pathogen.
Figure 4: Ramachandran plot of the predicted model of *A. hydrophila* GalE: This figure is generated by PROCHECK. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. Glycine is represented by triangles and other residues are represented by squares.
Figure 5: (A) Superimposition of homology modeled structure of *A. hydrophila* GalE onto a *Homo sapiens* GalE monomer. *A. hydrophila* GalE is shown in red and *Homo sapiens* GalE in blue. α2 and β12 correspond to human GalE. Structural differences between the two structures are encircled. (B). Superimposition of the catalytic site of human epimerase/UDPGlcNAc/NADH structure onto the catalytic site of the predicted model of the *A. hydrophila* GalE. The human protein is shown in blue with UDPGlcNAc (in CPK color scheme) and NADH (in green) while the *A. hydrophila* protein is depicted in red. Figure shows the substitution of Tyr 299 in the *A. hydrophila* epimerase with a Cys307 in the human enzyme.

Acknowledgements:
This work is supported by a research grant from the Indian Council of Agricultural Research, New Delhi, India to AD. The Council of Scientific and Industrial Research, New Delhi and the University Grants Commission, New Delhi are acknowledged for providing research fellowships to SA and KG.

References:
[22] GN Ramachandran, V Sasisekharan, Advances in Protein Chemistry 23: 283 (1968) [PMID: 4882249]

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.
### Supplementary Figure 1

**Figure 1**: Multiple sequence alignment of the deduced amino acid sequence of epimerase.

The deduced amino acid sequence of *A. hydrophila* GalE was aligned with amino acid sequences of GalE from other species. Genbank Accession Numbers are given in bracket in the top section. The signature sequences of the epimerase super family are shown as shaded residues. The conserved N-terminal NAD binding domain (GXXGXXG) and the catalytic couple (TyrXXXLys) are shown as boxed sequences. (*), (:) and (.) denote single fully conserved residues; residues with conserved strong groups and residues with conserved weak groups, respectively.