

Structural Characterization of FAD-binding Domain of CgIE, a Putative Dehydrolipoamide Dehydrogenase in Meningitic *E. coli* K1

Yi-Ming Wang^{1,2} and Hong Cao^{1*}

¹Department of Microbiology, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, 510515, China

²The First School of Clinical Medicine, Southern Medical University, Guangzhou, 510515, China

Abstract

To characterize the structural features of FAD-binding domain of *E. coli* K1 CgIE, a dehydrolipoamide dehydrogenase (DLDH) by homology modeling. Sequence similarity of N-terminal residues 1-70 with the α -subunit of FAD-binding domain from CgIE of *E. coli* K1 and other DLDHs provided a basis for the design of the FAD-binding domain of CgIE. As a result of finding no single satisfied template for the homology modeling for CgIE, two templates (PDB code 2q7vA and 1jehA) were obtained by an online homology modeling procedure for multi-templates modeling. To obtain a high quality target protein, a computational bioinformatic software Accelrys Discovery Studio client 2.5 and several automated online servers were utilized. Due to the relatively low identity of the alignments, two templates were taken into consideration attempting to improve the homology model. The quality of the refined model was assessed on the basis of both geometric and energetic aspects including MD simulations, energy minimizations, Ramachandran Plot and other measurements.

Keywords: *E. coli* K1; dehydrolipoamide dehydrogenase; FAD; homology modeling; Rossmann-fold

Background

CgIE is a putative dehydrolipoamide dehydrogenase (DLDH), which is the E3 component of pyruvate dehydrogenase complex, and shares significant protein sequence homology (50% identity and 70% similarity) to the *ibeA* invasin contributing to invasion of the blood-brain barrier (BBB) during neonatal *E. coli* meningitis [1]. The *ibeA* (*ibe10*) and *cglE* genes are encoded by a 20.3 kb genetic island GimA. It is reported that GimA, consisted of four operons, *ptnIPKC*, *cglDTEC*, *gcxKRCI* and *ibeRAT*, encodes 10 enzymes, 3 transporters, 1 regulatory protein and 1 invasin (*ibeA*), implicating that GimA might be involved in novel pathways that result in the regulation of invasion genes in *E. coli* K1 [2]. Previous researches showed that *cglE* and *ibeA* are present in the genetic island GimA as a pair of homologous protein that are encoded by two different operons, *cgl*(GimA2) and *ibe*(GimA4), at different locations. Similar pair of proteins are 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 may be important for energy metabolism in the both microbes [3]. On the other side, bioinformatic analysis indicates a DLDH characteristic FAD-binding domain and homologous flavoprotein regions are present in CgIE, which may act as the E3 component contributing to glycerol metabolism [3]. Glycerol may play a role in regulation of *E. coli* K1 invasion genes (e.g., *ibeA*) as this metabolite has been found to be an important signal in regulation of virulence gene expression in *Staphylococcus aureus* [2]. For example, the production of various exoenzymes and virulence factors, including protein A, alpha-hemolysin, β -lactamase, and toxic shock syndrome toxin 1 (TSST-1) in *S. aureus*, could be blocked by glycerol monolaurate (GML), a mild surfactant. GML also suppresses the induction of vancomycin resistance in *Enterococcus faecalis* [2]. Currently, it is unclear whether and how CgIE contributes to glycerol metabolism in the pathogenesis of *E. coli* K1 meningitis.

Glycerol dehydrogenase (GLDH) (glycerol:NAD(+) 2-oxidoreductase, EC 1.1.1.6) encoded by *cglD* in GimA2 may catalyze the oxidation of glycerol to dihydroxyacetone

(1,3-dihydroxypropanone) with concomitant reduction of NAD(+) to NADH. Dihydroxyacetone phosphate can be converted to pyruvate. As a member of the pyruvate and 2-oxoglutarate dehydrogenase multienzyme complexes, dehydrolipoamide dehydrogenase (DLDH, EC 1.8.1.4) catalyzes NAD(+)-dependent oxidation of dihydrolipoamide in vivo and can also act as a diaphorase catalyzing in vitro nicotinamide adenine dinucleotide (reduced form) (NADH)-dependent reduction of electron-accepting molecules such as ubiquinone and nitroblue tetrazolium (NBT) [4]. These complexes catalyze the oxidative decarboxylation of pyruvate and 2-oxoglutarate with the formation of acetyl-CoA and succinyl-CoA, respectively:



Up to now, it remains unknown how CgID and CgIE cooperatively contribute to glycerol metabolism. To understand the underlying mechanisms and biochemical functions of CgIE, it is necessary to know the three-dimensional structure of this enzyme. Due to the high sequence similarities between CgIE and DLDH of various species from bacteria to human, the homology modeling can be used as an efficient method for the three-dimensional structure construction of protein. Predicting a protein's structure from its amino acid sequence has been one of the most challenging issues in structural biology [6].

The goal of homology modeling is to build a structural model of a protein based on high sequence similarity to a template protein with its known structure which has been explored by X-ray crystallographers and NMR spectroscopists [7]. Once a known structure has been

***Corresponding author:** Hong Cao, Department of Microbiology, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, 510515, China, E-mail: gzhcao@fimmu.com

Received November 24, 2013; **Accepted** December 30, 2013; **Published** December 31, 2013

Citation: Wang YM, Cao H (2013) Structural Characterization of FAD-binding Domain of CgIE, a Putative Dehydrolipoamide Dehydrogenase in Meningitic *E. coli* K1. J Data Mining Genomics Proteomics 4: 146. doi:10.4172/2153-0602.1000146

Copyright: © 2013 Wang YM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

identified as a template of the target protein, a model can be built by copying backbone elements from this template. Typically a model backbone is constructed for the structurally conserved regions, loops are added, and side-chains are placed [8].

As a result of finding no single satisfied template for the homology modeling for CglE, two templates (PDB code 2q7vA and 1jehA) were obtained by an online homology modeling procedure SWISS-MODEL(<http://swissmodel.expasy.org>) [9] for multi-templates modeling. As accurate sequence alignments are vital factors for homology searches and for building models, a list of online tools to help modify automated alignments were taken into consideration.

Methods

Several online tools were utilized in obtaining the templates involving NCBI BLAST(x), SWISS-MODEL, CLUSTALW(x), and (PS)² server. Discovery Studio, SWISS-MODEL, Swiss-PdbViewer 4.01 and VMD1.8.7 [10] helps generate and visualize the homology model as well as manually adjusting alignments and improving the model under the Windows PC environment [11]. Molecular dynamics (MD) simulations and energy minimizations were carried out by NAMD2.81b, latest version of a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems to refine the model [12].

Structure and sequences

The first step was searching a number of related sequences to find proper proteins as templates. The selection of templates was judged by recommendation of online servers and the quality of the model protein. After considering the sequence identity and coverage of sequence alignments, yeast lipoamide dehydrogenase complexed with NAD (PDB code 1jeh chain A) [13], crystal structure of Deinococcus Radiodurans Thioredoxin Reductase (PDB code 2q7v chain A) were chosen as templates. The template 2Q7V_A covers residues 16 to 164 thus identity is 23.494%; The template 1jehB covers only residues 15 to 58, but shows a relatively high identity (52%). The X-ray crystal structures of both the templates 2Q7V_A and 1JEH_A were taken from the Protein Data Bank and were at 2.75Å resolution and 2.2Å resolution respectively.

Homology modeling

Homology modeling of proteins is the most accurate method for 3D-structure construction, yielding models suitable for a wide spectrum of applications such as structure based molecular design and mechanistic investigations. It is usually the method of choice when a clear relationship of homology between the sequences of a target protein and at least one known structure is found. The approach would give reasonable results based on the assumption that the tertiary structures of two proteins will be similar if their sequences are related [14]. Homology modeling consists of building a protein model using a structural template or multiple templates, with a protein of known structure. The detailed steps in homology modeling are shown in:

- a) Search templates
- b) Sequence alignment
- c) Determine structurally conserved regions(SCRs)
- d) Build backbone and side-chains
- e) Build loops

f) Molecular dynamics and energy minimization(optional)

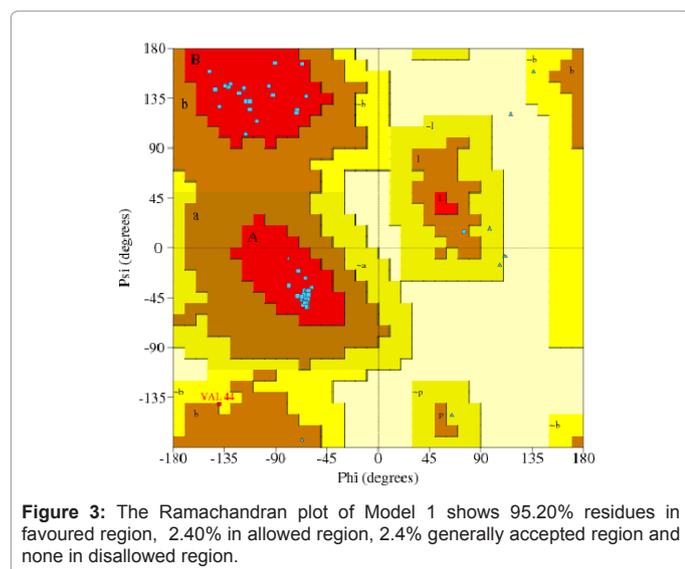
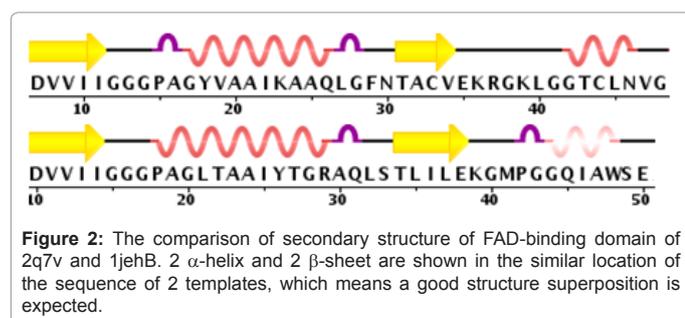
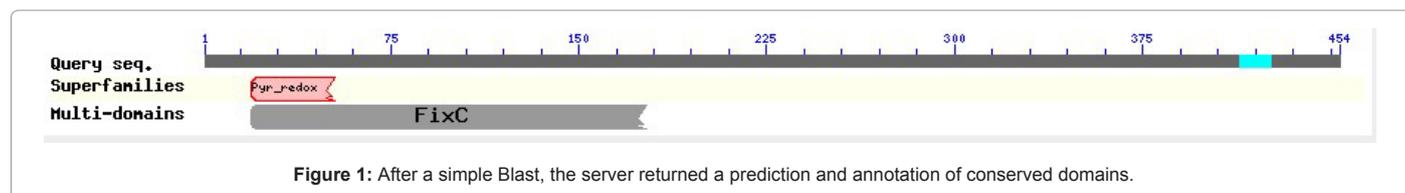
g) Verify the quality of model

Due to the low coverage of alignments between sequence of CglE and all the candidate templates, only a part of the protein could be modeled. The aligned region is a Rossmann-fold NAD(P)(+)- binding region which is a protein structural motif found in proteins that bind nucleotides such as cofactors like NAD, FAD and FMN [15]. To obtain a high quality target protein, a computational bioinformatic software Accelrys Discovery Studio client 2.5 and several automated online servers were utilized. Preliminary models for both templates were returned by online servers or generated by software. As two templates were used, for each of the template, the backbone atom positions of the structure are averaged. The templates are measured by their sequence similarity to the target sequence. When encountering regions of insertions or deletions in the target-template alignment, the server would use constraint space programming (CSP) to generate those parts [9]. The algorithm is highly reliable whereas manual intervention is still necessary. To manually enhance the alignment, an online tool for structural alignment based on the jFATCAT algorithm was used to compare the three-dimensional structures between templates and the result can be used to guide the superposition of the templates [16]. A result of secondary structure prediction made directly from CglE's sequence using another online tool was also exploited.

The template 1jeh is a lipoamide dehydrogenase from Yeast. The model was built from residues 7 to 48 which locate in the FAD binding domain containing 1 α -helix and 2 β -strands. The model based on 2q7v starts from residues 7 to 61 also sits in the FAD binding region. According to the sequence of 1jehB, a disulfide bond is formed between Cys44 and Cys49, which are essential for the dehydrogenase activity [13]. The disulfide bond at the active site of E3 of various species should be broken during the catalysis to form an intermediate with the substrate [17]. Thus there is no such a similar active disulfide bond located in the sequence of 2Q7V, suggesting that this disulfide bond is not necessary for DLDH.

Structure validation

The quality of the refined model was assessed on the basis of both geometric and energetic aspects. MD simulations and energy minimizations were carried out within the CHARMM force field using the steepest descent method followed by the conjugate gradient method [18]. In consecutive MD rounds the temperature was varied between 50 and 300K with a time step of 1fs. This tool provided the graphical representation of energy minimization of obtained protein models. The stereo chemical property was checked by Ramachandran. Ramachandran Plot is the display of the (ϕ, ψ) angle pairs of C₁-C and N-C₁ atoms of residues in an easily comprehensible way [19]. The PROSA test was employed to evaluate the quality of consistency between the native fold and the sequence to examine the energy of residue-residue interactions using a distance based pair potential [20]. The energy was transformed to a score called Z-score. Residues with negative Z-score indicated reasonable side chain interactions. The final structure with the lowest energy was checked by PROCHECK [21] to verify 3D module of Discovery Studio [22] and Qmean server [23]. PROCHECK checks the stereochemical quality of a protein structure, producing a number of PostScript plots analysing its overall and residue-by-residue geometry which is finally judged by two statistics: Ramachandran Plot and G-factors. Verifying 3D was used to access the compatibility of an atomic model with its own amino acid sequence. A highly verified 3D profile score indicates the high quality of a protein



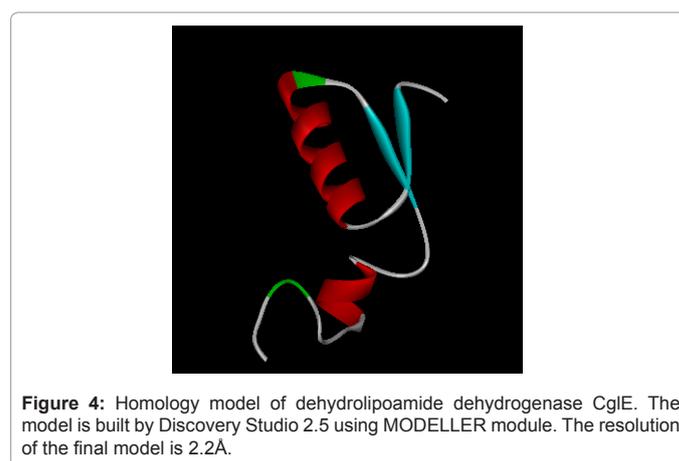
model. The Qmean server is a composite scoring function describing the major geometrical aspects of protein structures which contains five different structural descriptors.

Results and Discussion

Homology modeling of CglE

Studies show that when the sequence identity to the target structure is higher than 40%, the homology models are generally satisfactory. Because of the relatively low identity of the alignments, two templates were taken in consideration attempting to improve the homology model. The result of dual-template modeling showed slight improvements from all the chosen protein structure verification programs compared to single template modeling. Our findings suggest that CglE may act as the E3 component contributing to glycerol metabolism.

Absence of residues from disallowed region of the model constructed by Discovery Studio (Model 1) supports its high geometric quality comparing with the templates returned by SWISS-MODEL and (PS)². Figure 1 represents Ramchandran plot for Model 1.



Using two templates and manually adjusting the alignment could be possible explanation for this geometric arrangement. Models automatically generated by online servers utilized only one template could be possible explanation for residues in disallowed region of web automated models (Model 2, Model 3 and Model 4). Total quality of G-factors was also obtained in acceptable range as shown in Table 1. Acceptable values of G-factor in PROCHECK are between 0 and -0.5 with best model displaying values close to 0, indicating the design models with good quality and acceptable. Qmeans score is a composite score consisting of a linear combination of five terms, which helps to estimate the quality of protein structure model. The model Qmean scores obtained for the models shown in Table 2 are within the reliability (0-1). A final test is the packing quality of each residue as assessed by the Verify 3D program represents the profile obtained with respect to the residues. Verifying 3D uses a score function to analyze the compatibility of the residues with their environment in models. The vertical axis represents the average 3D-1D protein score for each residue in a 21 residue sliding window helps to further validating the models. Residues with a score over 0.2 should be considered reliable. Scores for all refined structures maximally lies above 0.2 which corresponds to acceptable side chain environment as represented in Table 2. Figure 2 represents graphical representation of the verified 3D-1D score for the model 10. To sum up, the geometric quality of the backbone confirmation, the residues interactions and the energy profile of the structures are all well within the limits established for reliable structures for the 10 models. To investigate how well the modeled structure matches the X-ray data of template, the prepared models and their respective templates were superimposed on their backbone atoms. RMSD values of backbone atoms for all models tabulated in Table 2 supported that generated models are reasonably good and quite similar to template. However, from a visual inspection a good overall agreement of secondary structural elements of the homology model 1 and the X-ray structures is observed. Figure 3 represents superimposition of CglE (Model 1) with template 1JEH_A.

Accession	Description	Max score	Total score	Query coverage	E-value
1JEH_A	Yeast E3, Lipoamide Dehydrogenase	48.9	48.9	76%	7.00E-09
1EBD_A	Binding Domain Of The DihydrolipoamideAcetylase	42.7	42.7	80%	1.00E-06
2Q7V_A	DeinococcusRadioduransThioredoxinReductas	42	42	94%	2.00E-06
1ZY8_A	Subcomplex Of Human Pyruvate Dehydrogenase Complex	41.2	67	76%	4.00E-06
1ZMC_A	Human Dihydrolipoamide Dehydrogenase Complexed To Nad+	41.2	67	76%	4.00E-06
3RNM_A	The Subunit Binding Of Human DihydrolipoamideTransacylase	41.2	67	76%	4.00E-06
1GRT_A	Human Glutathione Reductase A34eR37W MUTANT	40.8	40.8	81%	6.00E-06
2GRT_A	Human Glutathione Reductase , Oxidized Glutathione Complex	40.8	40.8	81%	6.00E-06
2EQ7_A	Lipoamide Dehydrogenase From ThermusThermophilus Hb8	40.4	63.9	76%	9.00E-06
3LAD_A	Lipoamide Dehydrogenase From AzotobacterVinelandii	39.3	39.3	76%	2.00E-05
3R9U_A	Thioredoxin-Disulfide Reductase From Campylobacter Jejuni	37.7	37.7	76%	7.00E-05
1LPF_A	Lipoamide Dehydrogenase From Pseudomonas Fluorescens	37.7	37.7	76%	8.00E-05

Table 1: The Blast result of CgIEthrough PDB online server. Yeast E3, Lipoamide Dehydrogenase and DeinococcusRadioduransThioredoxinReductas were chosen templates of following homology modeling.

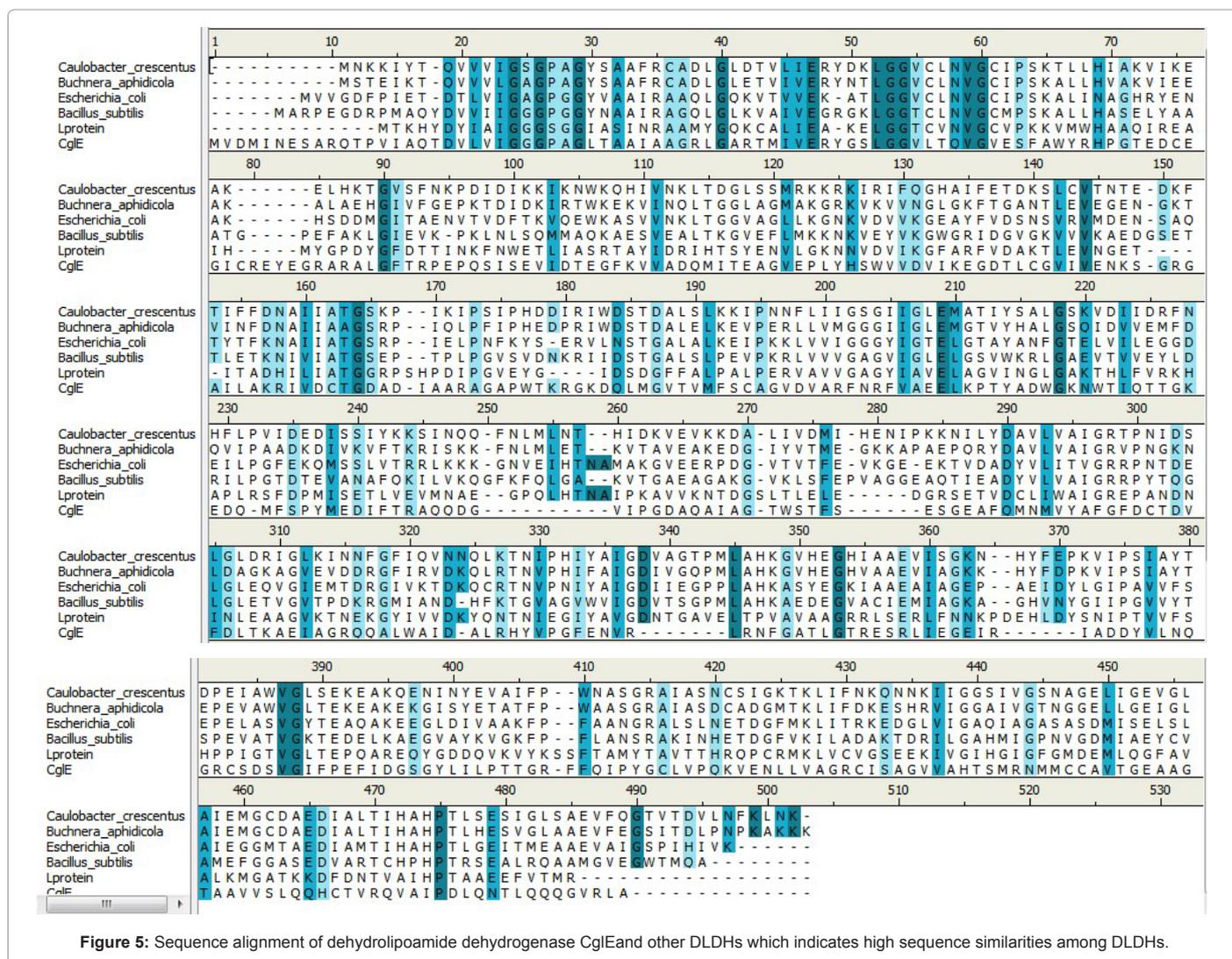


Figure 5: Sequence alignment of dehydrolipoamide dehydrogenase CgIE and other DLHDs which indicates high sequence similarities among DLHDs.

Validation of the refined structure

The minimized CgIE model has identical backbone coordinates with the experimental structure of 1ojtA. After the model was refined by MD simulations and energy minimization, the final structure of CgIE (residue 1 to 70) was displayed in Figure 4. From Figure 4, 7

α -helices and 8 β -strands can be recognized. Among the 149 residues, 3 residues were found in the disallowed regions of Ramachandran plot. The statistical score of the Ramachandran plot shows that 95.2% are in the most favored regions, 2.4% in the additional allowed regions, and 2.4% in generally allowed regions. The above results indicate that the

Models	Favored	Allow	General	Disallow	G-factor	Qmean	zscore	3D-1D
Model 1	95.20%	2.40%	2.40%	0%	-0.07	0.703	-0.09	45.16%
Model 2	92.90%	2.40%	4.80%	0%	-0.26	0.514	-1.02	26.79%
Model 3	89.20%	8.10%	2.70%	0%	-0.08	0.53	-0.75	35.29%
Model 4	91.70%	5.60%	0%	2.80%	0	0.57	-0.51	4.00%
Model 5	90.50%	7.10%	2.40%	0%	-0.1	0.531	-0.93	10.71%
Model 6	97.60%	2.40%	0%	0%	-0.08	0.614	-0.53	12.50%

Table 2: The Blast results of several online evaluation tools of 6 homology models. In most of the evaluations, Model 1 shows the better result than other models.

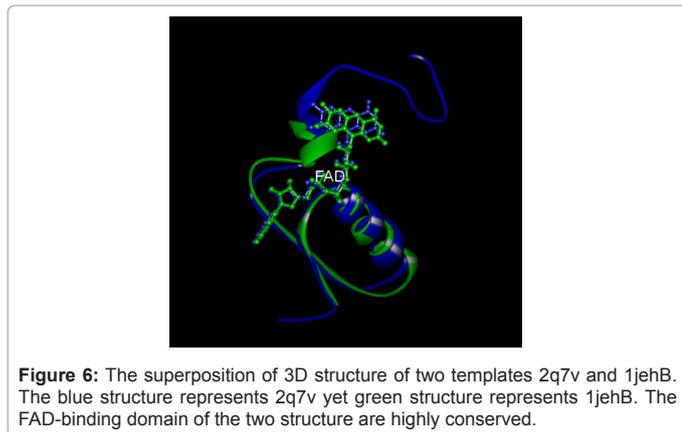


Figure 6: The superposition of 3D structure of two templates 2q7v and 1jehB. The blue structure represents 2q7v yet green structure represents 1jehB. The FAD-binding domain of the two structure are highly conserved.

homology model is reliable. Tools available at SWISS-MODEL were also taken into consideration. The ANOLEA results represent the Y-axis of the plot, the energy for each amino acid of protein chain. Negative energy values (in green) represents the favorable energy environment where as the values (in red) unfavorable energy environment forgiven amino acid (Figures 5 and 6).

References

- Cao H, Chen L, Yang J, Gong S, Zhou H, et al. (2007) Comparative Infectomic Analysis and Molecular Characterization of CgIE, the Invasin IbeA Homologous Protein, Which Contributes to the Pathogenesis of Meningitic *E. coli* K1 Infection. *Systemics, Cybernetics and Informatics* 6: 65-70.
- Huang SH, Germon P, Jong A (2004) The Invasion-Associated Genetic Island GimA in Meningitic *E. coli* K1. Focus on Meningitis Research (Editor: Strong, Phyllis V); Nova Biomedical Books, 99-116.
- Huang SH, Chen YH, Kong G, Chen SH, Besemer J, et al. (2001) A novel genetic island of meningitic *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* 1: 312-322.
- Yan LJ, Yang SH, Shu H, Prokai L, Forster MJ (2007) Histochemical staining and quantification of dihydrolipoamide dehydrogenase diaphorase activity using blue native PAGE. *Electrophoresis* 28: 1036-1045.
- Stephens PE, Lewis HM, Darlison MG, Guest JR (1983) Nucleotide sequence of the lipoamide dehydrogenase gene of *Escherichia coli* K12. *Eur J Biochem* 135: 519-527.
- Koehl P, Levitt M (1999) A brighter future for protein structure prediction. *Nat Struct Biol* 6: 108-111.
- Moult J (1996) The current state of the art in protein structure prediction. *Curr Opin Biotechnol* 7: 422-427.
- Bower MJ, Cohen FE, Dunbrack RL Jr (1997) Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol* 267: 1268-1282.
- Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31: 3381-3385.
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33-38, 27-8.
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781-1802.
- Toyoda T, Suzuki K, Sekiguchi T, Reed LJ, Takenaka A (1998) Crystal structure of eucaryotic E3, lipoamide dehydrogenase from yeast. *J Biochem* 123: 668-674.
- Bodade RG, Beedkar SD, Manwar AV, Khobragade CN (2010) Homology modeling and docking study of xanthine oxidase of *Arthrobacter* sp. XL26. *Int J Biol Macromol* 47: 298-303.
- Rao ST, Rossmann MG (1973) Comparison of super-secondary structures in proteins. *J Mol Biol* 76: 241-256.
- Prlc A, Bliven S, Rose PW, Bluhm WF, Bizon C, et al. (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26: 2983-2985.
- Kim H (2002) Activity of human dihydrolipoamide dehydrogenase is reduced by mutation at threonine-44 of FAD-binding region to valine. *J Biochem Mol Biol* 35: 437-441.
- MacKerell AD Jr, Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56: 257-265.
- Sheik SS, Sundararajan P, Hussain AS, Sekar K (2002) Ramachandran plot on the web. *Bioinformatics* 18: 1548-1549.
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35: W407-410.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283-291.
- Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277: 396-404.
- Benkert P, Tosatto SC, Schomburg D (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* 71: 261-277.