Computational Approach to Prioritize Drug Targets in Metabolic Pathway of *Laribacter hongkongensis*

Iftikhar Aslam Tayubi
Faculty of Computing and Information Technology,
King Abdulaziz University, Rabigh, Saudi Arabia.
iftikhar.tayubi@gmail.com

ABSTRACT

The availability of complete genome sequences of pathogenic bacteria accelerates the process of drug target identification. Further, the emergence of rapidly mutating strains evokes the necessity for the discovery of new drug targets. Hence, the objective of this study is to determine the highly expressed genes responsible in cellular pathways of *Laribacter hongkongensis*, where pyrimidine biosynthesis was our major pathway of interest. *Laribacter hongkongensis* is a gram-negative, seagull- or spiral rod-shaped potentially bacteria with global pathogenic occurrence causing diseases such as the community-acquired Gastroenteritis and Traveller's Diarrhoea. The highly expressed genes in related biological pathways were determined, from which the pseudogenes, hypothetical proteins and enzymes common to human species were systematically filtered out and eliminated. From the remaining enzymes, 2 key enzymes were selected based on their position of occurrence in the metabolic pathways. Subsequently, the 3-dimensional model of key enzyme Dihydroorotase from pyrimidine biosynthesis was predicted using MODELLER and validated using SAVS. The interaction of Dihydroorotase enzyme with three inhibitors was assessed by GOLD software. The study resulted in shortlisting of the highly expressed genes specific to the pathway of interest and in addition, resulted in determining the conserved residues at the binding site of DHO which can be targeted for designing inhibitor molecules, comprising a comprehensive approach which can be further, applied to other pathogens of clinical importance.

INTRODUCTION

One of the promising trends in the post-genomic era is the identification of novel drug targets using *in silico* approaches. The availability of genomic data for various pathogenic micro-organisms in combination with the characterization of their metabolic pathways has greatly enhanced the scope for the recognition of candidate drug targets. The viability of these drug targets can be evaluated based on two attributes namely, "essentiality" and "selectivity"[1]. Unveiling the essential genes and their associated protein products may form the basis for the discovery of new, antibacterial lead compounds[2].

Laribacter hongkongensisis is human pathogen prominently associated with community-acquired gastroenteritis and traveller's diarrhea [3].The isolation of *L. hongkongensis* from patients who resided in or have traveled to Asia, Europe, America, and Africa implied that the bacterium is likely to be of global importance. It is a Gram-negative, facultative anaerobic, motile, seagull or S-shaped, bacillus belonging to the *Neisseriaceae* family of β-proteobacteria[4]. Previous studies suggest *L. hongkongensis* isolates recovered from freshwater fish and affected patients fell into separate clusters,which suggest *hongkongensis* clones are probably more virulent or adapted to human than others [5].

There is evidence which suggests that all genomes evolve a unique codon usage pattern to optimize gene expression levels [6, 7]. Codon Usage Table (CUSP) available with the software package European Molecular Biology Open Software Suite (EMBOSS) is used to create a table depicting the codon usage frequency of all the genes expressed in the metabolic pathways of *Laribacter hongkongensis HLHK9*. A numerical indicator of gene expression known as Codon adaptation index (CAI) is then used to predict the highly expressed genes based on the Codon Usage Table as a reference. Genes assuming a CAI value greater than 0.70 are considered to be highly expressed [8]. This approach has yielded over 1000 proteins which could further be scrutinized for the recognition of novel drug targets. From this extensive list of proteins, we have chosen one particular key enzyme Dihydroorotase (DHO), for further investigation. DHO is a zinc metalloenzyme that functions in the pathway for the biosynthesis of pyrimidine nucleotides by catalyzing the reversible interconversion of carbamoyl aspartate and dihydroorotate. Furthermore,

this enzyme lacks an appropriate human homolog. These attributes make DHO a promising drug target [9]. The structure of DHO has been predicted and it's interaction with three inhibitors namely 5-fluoorotic acid (FOA), 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid (HDDP) and 4,6-dioxo-piperidine-2-(*S*)-carboxylic acid were analyzed.

Despite the availability of vaccines and antibiotics to treat gastroenterital diseases, rapidly mutating strains evoke the need for new antibacterial compounds. In this regard, we hereby propose a promising *in silico* approach for the identification of potential drug targets in *Laribacter hongkongensis HLHK9* based on codon usage preferences to determine gene expressivity.

MATERIALS AND METHODS

Identification of Potential Drug Targets

The complete set of genes expressed in the metabolic pathways of *Laribacter hongkongensis HLHK9* was downloaded from the KEGG database (ftp.genome.ad.jp/pub/kegg/). CUSP, a program from EMBOSS GUI (http:/bioinfo.hku.hk/EMBOSS/) was used to generate a codon frequency table for the set of genes obtained from *Laribacter hongkongensis HLHK9*. Based on the codon usage table generated via Codon Adaptation Index (CAI), a measure of gene expressivity was estimated. CAI utilizes a reference set of highly expressed genes from the species to analyse the relative merits of each codon, and based on the frequency of occurrence of codons in a gene sequence, CAI values are generated which determine the expression levels. The genes having CAI value greater than 0.70 were predicted to be highly expressed genes in *Laribacter hongkongensis HLHK9*. Further,the list of highly expressed genes was subjected to BLAST-P analysis against human genome as reference, to identify the set of human non-homologs. The list of human non-homologous sequences derived was further classified based on the pathways common to *Laribacter hongkongensis HLHK9* and pathways common to *Laribacter hongkongensis HLHK9* and *Homo sapiens*.The genes located in the upstream of the individual pathways were

considered as key enzyme targets that are essential for the survival of *Laribacter hongkongensis HLHK9*.

Homology Modeling

A Homology model is built for Dihydroorotase (DHO) of *Laribacter hongkongensis HLHK9* using MODELLER9v10 (http://salilab.org/modeller/modeller.html). The 3-D structure of DHO protein and it's coordinates from *E.coli* (PDB ID: 1J79) was used as the structural homologue to build the model. Since, the template selected exhibited high identity with our sequence of interest, basic modeling was employed to compute the theoretical model for Dihydroorotase (DHO) of *Laribacter hongkongensis HLHK9*. The quality of the obtained model was verified using the PROCHEK server (http://nihserver.mbi.ucla.edu/SAVES/).

Docking Analysis

The possible binding sites of DHO is determined via Q-site finder (http//bmbpcu36.leeds.ac.uk/qsitefinder/).The protein residues participating in the binding pocket formation was predicted by comparing with active sites of the template used for the modeling.

Molecular docking analysis was performed using GOLD (Genetic Optimization of Ligand Docking) software with the three inhibitors targeted at the binding site of the receptor DHO. The GOLD program employs genetic algorithm (GA) to explore the full range of ligand conformational flexibility and the rotational flexibility of selected receptor hydrogens. Grid was prepared for the protein with the center and the size of the bounding box set on 10 Å. The coordinates of the enclosing box (x = 121 Å; y =87 Å; z = 45 Å) were defined starting from the set of active site residues.

Figure 1: Structure of 5-fluoorotic acid (FOA), 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid (HDDP) and 4,6-dioxo-piperidine-2-(S)-carboxylic acid

RESULTS

Potential Drug Targets of *L.hongkongensis*

The results of CAI yielded 1012 protein targets having CAI value greater than 0.7 and were considered as highly expressed enzymes. The elimination of pseudo genes and hypothetical proteins revealed 426 protein targets from 53 pathways as highly expressed genes. Discarding 172 enzymes from *L.hongkongensis* that shared a significant similarity with the host (*Homo sapiens*), aided in eliminating undesired protein-drug interactions. Thirty one enzymes were obtained from the pathways that are common to the host and the pathogen, whereas two enzymes were obtained from pathways that are unique to the pathogen. The list of key enzymes is shown in the **Table 2**.

Pathways unique to *L.hongkongensis*

The two pathways identified as unique in this study are Pyramidine Metabolism and Peptidoglycan synthesis. Enzymes from these pathways which do not exhibit similarity with *Homo sapiens* are shown in **Table 3.**

Pyramidine Metabolism

De novo pyrimidine nucleotide biosynthesis is also referred to as the "orotate pathway" which refers to the formation of uridine monophosphate (UMP) from carbamoyl phosphate (CP) [10]. In view of the importance of carbamoyl phosphate from both arginine and pyrimidine synthesis, coordinating the channeling of this intermediate into both the pathways is crucial. Carbamoyl phosphate synthetase A provides a pool of carbamoyl phosphate specific to the arginine pathway, whereas, Carbamoyl phosphate synthetase P provides carbamoyl phosphate pool specific to the pyrimidine pathway. Aspartate and carbamoyl phosphate are coupled to form carbamoyl aspartate (ureidosuccinate), which is further, cyclized by a separate enzyme to form dihydroorotate. After oxidation to orotate, the nucleotide is formed by phosphoribosyl transferase [11].

Peptidoglycan biosynthesis

Peptidoglycan or murein is the polymeric mesh of the bacterial cell wall which plays important role in protecting the bacteria against osmotic lysis. Biosynthesis of cell wall can be divided into two phases - an intracellular cytoplasmic phase involving six

enzymatic steps and three steps that occur exterior to the plasma membrane. The cytoplasmic phase comprises of four ADP forming ligases namely MurC, MurD, MurE and MurF [12]. The key enzyme that we have identified from this pathway in *L.hongkongensis* is MurC, also known as UDP-N-acetylmuramate-L-alanine ligase. MurC along with the other cytoplasmic ligases catalyzes the assembly of the peptide moiety of the peptidoglycan by the successive additions of L-alanine, D-glutamate, L-lysine and D-alanine to UDP *N*-acetylmuramic acid [13,14].

Homology Modelling

The three dimensional structure of DHO, considered as the most potential target for the study in *L.hongkongensis* was predicted and subjected to validation using PROCHEK server. The Ramachandran plot shows 92.6% of residues in most favored region, 6.4% in additionally allowed region, 0.7% residues in generously allowed region and 0.3% residues in the disallowed region. The residue namely LEU101, present in the disallowed region was subjected to loop refining and the final model obtained had 92.2% of residues in the core region. The residues in the disallowed region were shifted to the allowed region thereby stabilizing the overall conformation of the protein structure.

Figure 2: The 3D structure of DHO obtained after energy minimization. The α-helix is represented by red, β-sheet by cyan and loops by grey lines.

Figure 3: Ramachandran plot of DHO built using MODELLER software.

Docking Analysis

Ligand binding site of DHO

The active site of the receptor DHO was determined using the Q-Site Finder. (The active site of the protein includes HIS13, HIS15, ARG17, MET39, ASN41, LEU 101, TYR102, LEU135, VAL136, HIS137, HIS175, CYS219, LEU220, PRO221, ASP248, ALA250, HIS252, ALA264, GLY265 and ILE266.

Figure 4: Active site residues taken for docking analysis. Purple color indicates the protein and red color indicates the active site.

Docking of DHO was performed with three potential inhibitors namely 5-fluoorotic acid (FOA), 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid (HDDP) and 4,6-dioxo-piperidine-2-(*S*)-carboxylic acid. The three final docked conformation obtained for different inhibitors were evaluated based on the number of hydrogen bonds formed and bond distance between atomic co-ordinates of the active site and inhibitor.

The residues HIS13, HIS15 and HIS 175 are involved in hydrogen bonding interactions with the three inhibitors. The hydrogen bond interactions between the three inhibitors and DHO along with their bond distances and their corresponding scores are shown in table 4**.**

Figure 5a: DHO in complex with 5-fluoorotic acid (FOA).

Figure 5b: DHO in complex with 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid (HDDP) .

Figure 5c: DHO in complex with 4, 6-dioxo-piperidine-2-(*S*)-carboxylic acid.

DISCUSSION

With the advent of microbial genomics, it is now possible to delve into the function of pathogenic proteins and evaluate them as drug targets, thereby expediting the process of drug discovery. The identification of drug targets from microbial genomes is of significant importance for designing antimicrobials against emerging pathogenic strains. Previous studies suggest *L.hongkongensis* is one of the leading causes of gastroenteritis disorders.We have attempted to discover potential drug targets *in silico* based on highly expressed genes. The approach undertaken in determining gene expression levels is based on codon usage bias among the genes expressed in the metabolic pathways of *L.hongkongensis*.

The pattern of codon usage in the genome of every organism is unique and largely governed by the process of translational selection [15, 16, 17]**.** Owing to the degeneracy of the genetic code, many codons are synonymous for the same amino acid. Among the synonymous codons, some are more preferentially used than the others. This codon bias is nonrandom and characteristic of a particular species [18]. Hence, codon preferences differ considerably within and between organisms [19, 20, 21]. Experimental evidence in

E.coli and *S.cerevisiae* suggest that the codon bias is strongly linked to levels of gene expression [18,22].The effect of translational bias on gene expression was ascertained by a numerical indicator known as Codon Adaptation Index (CAI) [18]. The algorithm uses a reference set of genes which represents the synonymous codon bias [19, 23, 24]. In this study, the codon usage table generated from all the genes present in the metabolic pathways of *L.hongkongensis* is the reference set. This set yields the frequency of occurrence of every codon in all the genes expressed implying codon usage bias, which can then be correlated to levels of gene expression. Based on this reference set, we have determined the CAI value for all the genes in the metabolic pathways of *L.hongkongensis*. Genes having a CAI value of 0.7 and above were considered to be highly expressed [25]. About 1012 genes had a CAI value of 0.7 and above as shown in **figure 6**. From this exhaustive list of highly expressed genes and their associated proteins, we have employed a systematic approach to narrow down the search for potential drug targets.

Figure 6: Graph showing the No. genes with CAI value greater than 0.7

In this regard, the pseudogenes, hypothetical proteins and enzymes common to metabolic pathways of human and *L.hongkongensis* have been eliminated. As a result of this subtractive approach, 52 unique enzymes were shortlisted having unique occurrence in the metabolic pathways of *L.hongkongensis* which qualifies them to be candidate drug targets. We further filtered out and listed 2 of them as key enzymes shown in table 3. Each of these key enzymes has been chosen by virtue of its upstream occurrence in the respective metabolic pathways. Hence, unique enzymes from pathways occurring relatively downstream were not chosen for further analyses. Our contention is enzyme/protein targets impeded upstream in the pathway will be an effective approach to nullify the effects of the entire cascade. We have considered one key enzyme DHO from the Pyrimidine biosynthesis pathway for structure prediction and docking analysis with potential inhibitors.

One of the candidate antimicrobial enzyme targets of Gram-negative bacteria present in the *de novo* pyrimidine biosynthesis, a zinc metalloenzyme that catalyzes the third step of

the pyrimidine biosynthesis pathway [26]. Dihydroorotase (DHOase) is the enzyme that catalyzes a key step involving the cyclical dehydration of N-carbamoyl-l-aspartate (CAasp) to l-dihydroorotate (DHO). The dihydroorotase gene has been identified as an essential gene in several organisms, including *Bacillus subtilis* [27], *Staphylococcus aureus* [28] and *Mycobacterium tuberculosis* [29].

In this work the 3D model of DHO in *L.hongkongensis* has been constructed and it's interaction with three potential inhibitors has been analyzed (**Table 3** & **Table 4**)**.** The three final docked conformations obtained with the various inhibitors were evaluated based on the number of hydrogen bonds formed and bond distance between atomic co-ordinates of the active site and inhibitor.

The 22 amino acid active site pocket provides a cavity for the three inhibitors to interact with the receptor (DHO). The inhibitor 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid (HDDP) binds with the receptor with the highest GOLD score of 30.50, comparatively the inhibitor 5-fluoorotic acid (FOA) binds with a score of 28.63 and 4,6-dioxo-piperidine-2-(*S*)-carboxylic acid with a score of 27.59.

From the analysis of the H-bond formations between the three inhibitors and the DHO receptor,it is evident that 4,6-dioxo-piperidine-2-(*S*)-carboxylic acid form eight h-bonds, inhibitor 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid (HDDP) form five h-bonds and 5-fluoorotic acid (FOA) forms seven h-bonds.The aminoacid residues HIS13, HIS15 and HIS175 are involved in the interaction with the three inhibitors, thereby, suggesting them as the critical amino acids for the inhibition of the sigma 1 receptor activity.In addition, the residues VAL136 and HIS137 proved to be important conserved aminoacids.

CONCLUSION

Based on current data, it is clear that *L.hongkongensis* DHO has been endowed with conserved residues which plays instrumental role for binding to a variety of inhibitors. This study is the first report on an *insilico* approach towards the identification of novel drug targets in *L.hongkongensis* based on gene expression levels. It also assesses the

various interactions of the drug target DHO with its inhibitors thereby providing preliminary insight into the prospects of structure based drug design.

References:

1. Sakharkar, K. R., Sakharkar, M. K. and Chow, V. T. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. I*n Silico Biol.* 4, 0028.
2. Mushegian, A. R. and Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci.* USA 93, 10268-10273.
3. Woo, P. C. Y., S. K. P. Lau, J. L. L. Teng, T. L. Que, R. W. H. Yung, W. K. Luk, R. W. M. Lai, W. T. Hui, S. S. Y. Wong, H. H. Yau, and K. Y. Yuen. 2004. Association of Laribacter hongkongensis in community-acquired gastroenteritis with travel and eating fish: a multicentre case-control study. *Lancet* 363:1941–1947.
4. Woo PC, Lau SK, Teng JL, Yuen KY (2005) Current status and future directions of Laribacter hongkongensis, a novel bacterium associated with gastroenteritis and traveller's diarrhoea. *Curr Opin Infect Dis* 18: 413–419.
5. Teng JL, Woo PC, Ma SS, Sit TH, Ng LT, Hui WT, Lau SK, Yuen KY: Ecoepidemiology of Laribacter hongkongensis, a novel bacterium associated with gastroenteritis. *J Clin Microbiol* 2005, 43:919-922.
6. Karlin S, Mrázek J (2000). Predicted highly expressed genes of diverse prokaryotic genomes. *J Bacteriol*. 182(18), 5238-50.
7. Karlin S, Barnett MJ, Campbell AM, Fisher RF, Mrazek J. (2003). Predicting gene expression levels from codon biases in alpha-proteobacterial genomes. *Proc Natl Acad Sci* U S A. 100(12), 7313-8.
8. Wu G, Nie L, Zhang W. (2006). Predicted highly expressed genes in Nocardia farcinica and the implication for its primary metabolism and nocardial virulence. Antonie Van *Leeuwe Teng nhoek*. 89(1), 135-46.
9. Whittington DA, Rusche KM, Shin H, Fierke CA, Christianson DW. (2003). Crystal structure of LpxC, a zinc-dependent deacetylase essential for endotoxin biosynthesis. *Proc Natl Acad Sci* U S A. 100(14), 8146-50.
10. Tatibana, M. (1978). Interrelationship of purine and pyrimidine metabolism. In Uric Acid, W.N. Kelley and I.M. Weiner, eds (Berlin, Germany: Springer-Verlag), pp. 125-154.
11. Wagner, K.G., and Backer, A.I. (1992). Dynamics of nucleotides in plants studied on a cellular basis. *Intl. Rev. Cytol*. 134, 1-84.
12. van Heijenoort J.(2001). Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology*. 11(3), 25R-36R. Review.
13. Doublet P, van Heijenoort J, Mengin-Lecreulx D. (1994). The glutamate racemase activity from Escherichia coli is regulated by peptidoglycan precursor UDP-N-acetylmuramoyl-L-alanine. *Biochemistry*. 3;33(17), 5285-90.

14. Rogers, H.J., Perkins, H.R., Ward, J.B. (1980). Biosynthesis of peptidoglycan. In Microbial Cell Walls and Membranes (Rogers, H.J., Ed.), pp. 239–297. Chapman and Hall, London.
15. Horn D. (2008). Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*.9,2.
16. Andersson SG, Kurland CG. (1990). Codon preferences in free-living microorganisms. *Microbiol Rev*. 54(2), 198-210
17. Kurland, C. G. (1993). Major codon preference: theme and variation. *Biochem. Soc. Trans.* 21, 841-846.
18. Paul M. Sharp and Wen-Hsiung Li. (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucl. Acids Res* 15, 1281-1295.
19. Karlin S, Campbell AM, Mrázek J. (1998). Comparative DNA analysis across diverse genomes. *Annu Rev Genet.* 32,185-225.
20. Irwin B, Heck JD, Hatfield GW. (1995). Codon pair utilization biases influence translational elongation step times. *J Biol Chem.* 270(39), 22801-6.
21. Sharp P. M., G. Matassi, (1994). Codon usage and genome evolution *Curr. Opin. Genet. Dev* 4,851-860.
22. Ikemura T. [1981]. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol*. 146(1), 1-21.
23. Carbone A, Zinovyev A, Képès F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics.* 19(16), 2005-15.
24. Shahila Mehboob, Debbie C. Mulhearn, Kent Truong, Michael E. Johnson, Bernard D. Santarsiero. Structure of dihydroorotase from Bacillus anthracis at 2.6 Å resolution. Acta Crystallogr Sect F *Struct Biol Cryst Commun*. 2010 November 1; 66(Pt 11): 1432–1435.
25. Samant S, Lee H, Ghassemi M, Chen J, Cook JL, Mankin AS, Neyfakh AA. Nucleotide biosynthesis is critical for growth of bacteria in human blood. *PLoS Pathog*. 2008 Feb 8; 4(2):e37.
26. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol*.48, 77–84.
27. Sakharkar, K. R., Sakharkar, M. K. and Chow, V. T. (2006). Gene fusion in Helicobacter pylori: making the ends meet. *Antonie Van Leeuwenhoek* 89, 169-180.
28. Forsyth, R. A. et al. (2002). *Mol. Microbiol*.43, 1387–1400.
29. Sharp PM, Li WH. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15(3), 1281-95.