ISSN: 0975-766X
CODEN: IJPTFI
Available Online through
www.ijptonline.com
Research Article

IDENTIFICATION OF DRUG TARGET IN FOOD PATHOGEN - *STREPTOCOCCUS PYOGENES*

Polani B Ramesh Babu*, Krishnamoorthy, P and Abenaya M
Department of Bioinformatics, School of Bioengineering
Bharath University, Chennai, Tamil Nadu, India.
Email: bioinfohod@gmail.com

Received on 10-11-2012 *Accepted on 24-11-2012*

Abstract

Streptococcus pyogenes (S.pyogenes) is the cause of many important human diseases, ranging from mild superficial skin infections to life-threatening systemic diseases. This pathogen of interest for the therapeutic drug target is the one causing contamination in food samples. The samples (contaminated vegetables) subjected to serial dilution was used as inoculums for the Nutrient agar (NA) plates that were further sub-cultured using streaking techniques. Pure colonies were isolated from the selective medium to isolate a particular strain in NA slants and subjected to Gas Chromatography (GC) for identification. The organism identified was *S pyogenes*. The entire set of proteins from this pathogen and the ones encoded in human genome were determined. Pyruvate kinase involved in carbohydrate metabolic pathway of S.pyogenes was chosen based on gene expression as a potential drug target candidate. This enzyme was further analyzed for protein structure prediction. Three protein inhibitor UDP D-Xululose was selected and screened for the inhibitor study using protein ligand docking program Genetic Optimisation of ligand Docking (GOLD). The docking studies proved that the inhibitors had better interactions with the drug targets and modeled proteins can be used as target sites for *S pyogenes* contaminations and this study is a modest approach towards target based drug designing.

Introduction

Food-borne diseases (FBD) are defined by the World Health Organization as “diseases of infectious or toxic nature caused by, or thought to be caused by the consumption of food or water”. Among FBDs, food-borne infections

are caused by many different disease-causing pathogens that can contaminate foods, while food-borne poisoning is caused by poisonous chemicals, or other harmful substances that are present in food. In many countries, national health care organizations record FBD outbreaks, defined as the occurrence of two or more cases of a similar illness resulting from the ingestion of a common food.

S. pyogenes is a spherical, Gram positive bacterium that is the cause of group A Streptococcal infections [1-3]. S.*Pyogenes* has an incubation period of approximately 1–3 days. It is an infrequent, but usually pathogenic. The *S. pyogenes* genome sequence was determined by using the whole-genome shotgun approach. A biological target is frequently used in pharmaceutical research to describe the native protein in the body whose activity is modified by a drug resulting in a desirable effect. Currently, our Target methodology builds on three main pillars: protein structure prediction, expression data analysis and metabolic/regulatory pathway modeling

S. pyogenes can also cause disease in the form of postinfectious "nonpyogenic" (not associated with local bacterial multiplication and pus formation) syndromes [3-5]. These autoimmune-mediated complications follow a small percentage of infections and include rheumatic fever and acute postinfectious glomerulonephritis.

Pathways are biologically meaningful subgraphs, eg signaling cascades or metabolic pathways that account for supply and consumption of any intermediate metabolites. Each of the methods described above can provide valuable clues pointing to target proteins. But the crux lies in their clever combination, interconnecting data from different sources. In recent work, we have shown that in real life situations clustering alone may not be able to reconstruct pathways from gene expression data. Set of pathways is extracted from a protein/gene network. Then, these pathways are scored with respect to gene expression data. They can be constructed from interactions that are observed in different tissues or species. The expression data provide an orthogonal view on these interactions and can thus be used to validate the drug target.

The molecular structure, the genetic aspects and the interaction of these fascinating molecules with the immune system remains their most interesting aspects. The key role of Streptococcus Superantigens (SAgs) [1,2,6] , a family of highly mitogenic proteins secreted individually or in certain combinations by many strains in the pathogenesis of acute, chronic, and some autoimmune diseases has offered new insights into elucidating patho-physiological effects of these

molecules [6-9]. Many questions remain unanswered, but new achievements will certainly emerge in the coming years, such as the design and use of novel therapeutical strategies (drugs and vaccines) in the management of SAgs-induced diseases.

Materials and Methods

Sampling

Contaminated vegetable samples were obtained from a restaurant in Chennai. 10gm of the sample was retrieved aseptically into sterile vials, using sterilized spatulas. The samples were transported to the laboratory within 1hour in a clean thermal container filled with freezing mixture. The media required for the enumeration of organism i.e. Nutrient agar – 2.8g/100ml and Mac conkey – 5.15g/ml were prepared in conical flasks and autoclaved (121^{o}C/15min). Sterile water (9ml and 10ml) was prepared in test tubes. Glass rods, spreaders, pipette tips were also sterilized. The whole testing procedure was performed inside a laminar airflow unit that is properly sterilized by UV irradiation method for 20minutes. After irradiation, the air blower was switched on. 1 gm of the food sample was retrieved from the collection vial, using sterile (with 70% ethanol) spatulas/pipettes. It was then added to 10ml of sterile saline/distilled water in test tubes and macerated thoroughly using a glass rod. The sample was allowed to settle down for a few minutes so that the microbial flora will come to the aqueous phase. This forms the 'Master stock'.

Culture and Isolation of pure colonies of *S pyogenes*

The sample was then serially diluted by transferring 1ml of the master stock into 9 ml of sterile water. The procedure was repeated till the 10^{th} tube such that the dilution will gradually increase from 10^{-1} to 10^{-10}. Using sterile tips, 0.1ml of the sample from the 3^{rd}, 5^{th} and 7^{th} dilutions were plated in Nutrient agar and Mac Conkey agar media by spread plate technique, with the help of sterile glass rods. The plates were incubated at 37^{o}C for 24 hours.

Individual colonies were identified from the Spread plates, based on their morphology [5]. These colonies were then grown on Nutrient agar plates by quadrant streaking technique to obtain pure individual colonies to carry out further works. Streak plates allow bacteria and fungi to grow on a semi-solid surface to produce discrete colonies. These colonies can be used to help identify the organism, purify the strain free of contaminants, and produce a pure genetic clone.In order to obtain well-isolated discrete colonies, the quadrant streak technique should be used. This

allows sequential dilution of the original microbial material (broth culture or colonies on a plate or slant) over the entire surface of a fresh plate. As the original sample is diluted by streaking it over successive quadrants, the number of organisms decreases. Usually by the third or fourth quadrant only a few organisms are transferred on the inoculating loop and these produce a few isolated colonies.

Bioinformatics tools used to identify genes in *S pyogenes*

CD-HIT stands for Cluster Database at High Identity with Tolerance (10,11,12). The program (CD-HIT) takes a fasta format sequence database as input and produces a set of 'non-redundant' (nr) representative sequences as output. In addition cd-hit outputs a cluster file, documenting the sequence 'groupies' for each nr sequence representative. The idea is to reduce the overall size of the database without removing any sequence information by only removing 'redundant' (or highly similar) sequences. CD-HIT uses a 'longest sequence first' list removal algorithm to remove sequences above a certain identity threshold. Additionally the al gorithm implements a very fast heuristic to find high identity segments between sequences, and so can avoid many costly full alignments. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (13-15). The metabolic pathway database records networks of molecular interactions in the cells and we used it to identify enzymes in carbohydrate metabolic pathway.

GOLD is a program for calculating the docking modes of small molecules in protein binding sites and is provided as part of the GOLD Suite, a package of programs for structure visualisation and manipulation (Hermes), for protein-ligand docking (GOLD) and for post-processing and visualisation of docking results (16,17). We have used GOLD in this study because, it is very highly regarded within the molecular modeling community for its accuracy and reliability.

Pfam is a database of protein families that includes their annotations and multiple sequence alignments generated using hidden Markov models. 74% of protein sequences have at least one match to Pfam. This number is called the sequence coverage. The Pfam database contains information about protein domain and families. Pfam-A is the manually curated portion of the database that contains over 10,000 entries. For each entry a protein sequence alignment and a hidden Markov model is stored. These hidden Markov models can be used to search sequence databases with the HMMER package written by Sean Eddy. Because the entries in Pfam-A do not cover all known

proteins, an automatically generated supplement is provided called Pfam-B. Pfam-B contains a large number of small families derived from clusters produce by an algorithm called ADDA. Although of lower quality, Pfam-B families can be useful when no Pfam-A families are found.

Results and Discussion

Our results showed that the food pathogen was isolated from contaminated food was Streptococcus pyogenes which was further identified using Gas chromatography (GC). Carbohydrate metabolic pathway of S.pyogenes was chosen for identification of the drug target and enzymes involved in common microbial metabolic pathway, namely Pyruvate Kinase was selected based on gene expression as a potential drug target candidate. This enzyme was further analyzed for protein structure prediction.

Amino acid sequences were extracted from KEGG database (Fig 1). CD-HIT was performed to generate the non-paralogous sequences. The outputs of CD-HIT were the base results for drug target identification. The percentage of similarity between the target and template was selected in the range of 40% to 90%, for which the target and the template should have the same function. The domain analysis was done for the target and the template sequence to find whether they share the same function. The pfam database was used to find the domains present in the sequence. The pfam result clearly indicates that the target and the template sequence have the same Pyruvate Kinase (Fig 2). We short listed the sequences to solve the structure and refined a stable structure in order to find the active site of the protein and the inhibitor (Fig 3). Docking was further initiated using protein-modeling program that contained tools to visualize, analyze, modify and simulate protein structures (Fig 4). By docking studies the mechanism of action can be concluded based on the results for the prediction of therapeutic drug targets.

The protein inhibitor UDP–D Xylulose was selected and screened for the inhibitor study using GOLD (Table 1). The molecular docking results show a fitness score of 56.05 between UDP-D Xylulose and Pyruvate kinase. The analog binds with the protein with a positive fitness, which proves that the series of analogs have a potential activity in inhibiting *S pyogenes* by down regulating the protein responsible for the survival of the bacteria. The docking studies proved that the inhibitors had better interactions with the drug targets and modeled proteins can be used as target sites for *S. pyogenes* contaminations.

Figure-1: Aminoacid sequence of Pyruvate kinase

```
>pdb|2E28|A Chain A, Crystal Structure Analysis Of Pyruvate Kinase From Bacillus
Stearothermophilus
Length=587

 Score =  402 bits (1032),  Expect = 4e-133, Method: Compositional matrix adjust.
 Identities = 221/501 (44%), Positives = 310/501 (62%), Gaps = 29/501 (6%)

Query  1    MNKRVKIVATLGPAVEIRGGKKYGEDGYWAGQLDVEESAKKIAELIEAGANVFRFNFSHG  60
            M ++ KIV+T+GPA                       ES  K+ +L+EAG NV R NFSHG
Sbjct  1    MKRKTKIVSTIGPA---------------------SESVDKLVQLMEAGMNVARLNFSHG  39

Query  61   DHKEQGDRMATVRLAEEIARQKVGFLLDTKGPEMRTELFADDAKEFSYVTGEKIRVATTQ  120
            DH+E G R+A +R A +    + V  LLDTKGPE+RT    + A E     G K+ ++ ++
Sbjct  40   DHEEHGRRIANIREAAKRTGRTVAILLDTKGPEIRTHNMENGAIELKE--GSKLVISMSE  97

Query  121  GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDDGKLGLKVIDKDIATRQFIVEVENDGII  180
             +  T + I++       + D+V VG  IL+DDG + L+V   D    + +  V N G++
Sbjct  98   -VLGTPEKISVTYP---SLIDDVSVGAKILLDDGLISLEVNAVDKQAGEIVTTVLNGGVL  153

Query  181  AKQKGVNIPNTKIPFPALAERDNADIRFGLEQGLNFIAISFVRTAKDVEEVREICRETGN  240
              +KGVN+P  K+  P + E+D ADI FG+ QG++FIA SFVR A DV E+RE+
Sbjct  154  KNKKGVNVPGVKVNLPGITEKDRADILFGIRQGIDFIAASFVRRASDVLEIRELLEAHDA  213

Query  241  DHVQLFAKIENQQGIDNLDEIIEAADGIMIARGDMGIEVPFEMVPVFQKMIITKVNAAGK  300
             H+Q+ AKIEN++G+ N+DEI+EAADG+M+ARGD+G+E+P E VP+ QK++I K N  GK
Sbjct  214  LHIQIIAKIENEEGVANIDEILEAADGLMVARGDLGVEIPAEEVPLIQKLLIKKSNMLGK  273

Query  301  AVITATNMLETMTEKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID  360
             VITAT ML++M   PR TR+E SDV NA+ DGTDA MLSGE+A G+YPVE+V+TM  I
Sbjct  274  PVITATQMLDSMQRNPRPTRAEASDVANAIFDGTDAVMLSGETAAGQYPVEAVKTMHQIA  333

Query  361  -RNAQTLLNEYGRLDSSAFPRTNKTDVIASAVKDATHSMDIKLVVTITETGNTARAISKF  419
             R  Q L +       +   +T  TD I  +V     ++D+  +VT T +G T + ++K+
Sbjct  334  LRTEQALEHRDILSQRTKESQTTITDAIGQSVAHTALNLDVAAIVTPTVSGKTPQMVAKY  393

Query  420  RPDADILAVTFDEKVQRALMINWGVIPVLAEKPASTDDMFEVAERVAVEAGLVQSGDNIV  479
            RP A I+AVT +E V R L + WGV    A    +TD+M +VA   AV +GLV+ GD +V
Sbjct  394  RPKAPIIAVTSNEAVSRRLALVWGVYTKEAPHVNTTDEMLDVAVDAAVRSGLVKHGDLVV  453

Query  480  IVAGVPVG-TGGTNTMRVRTV  499
            I AGVPVG TG TN M+V  +
Sbjct  454  ITAGVPVGETGSTNLMKVHVI  474
```

Figure 2: Target Sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To				
PK	Pyruvate kinase, barrel domain	Family	CL0151	2	373	3	370	**2**	**345**	478.2	9.1e-144	n/a	Show
PK_C	Pyruvate kinase, alpha/beta domain	Domain	n/a	384	498	384	497	1	**116**	115.3	9.6e-34	n/a	Show

Template Sequence

Significant Pfam-A Matches

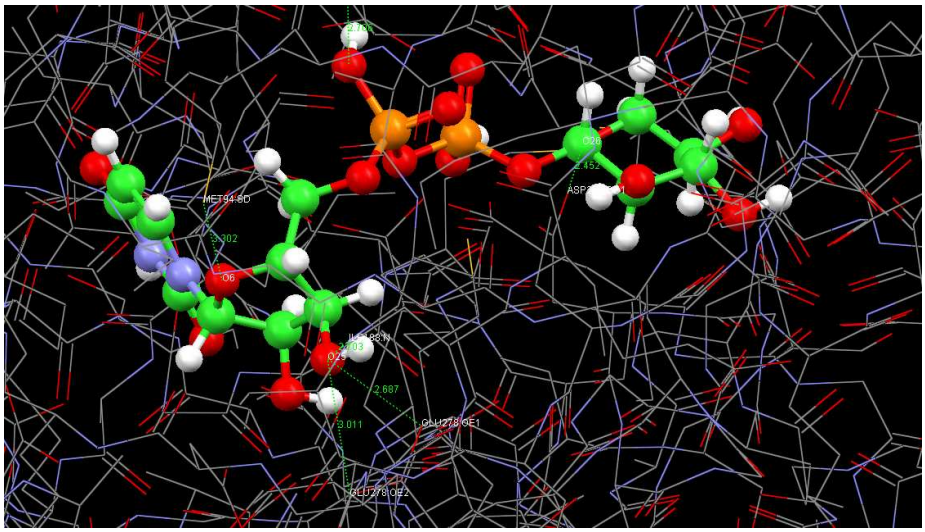
Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To				
PK	Pyruvate kinase, barrel domain	Family	CL0151	2	346	3	345	**2**	**347**	533.4	1.4e-160	n/a	Show
PK_C	Pyruvate kinase, alpha/beta domain	Domain	n/a	358	473	358	472	1	**116**	122.2	6.8e-36	n/a	Show
PEP-utilizers	PEP-utilising enzyme, mobile domain	Family	n/a	498	577	501	575	**4**	**79**	81.6	1.9e-23	541	Show

Figure 3: Shortlisted sequence of Pyruvate kinase to match the target site.

```
>P1;target
sequence:target:1: :499: ::::
MNKRVKIVATLGPAVEIRGGKKYGEDGYWAGQLDVEESAKKIAELIEAGANVFRFNFSHG
DHKEQGDRMATVRLAEEIARQKVGFLLDTKGPEMRTELFADDAKEFSYVTGEKIRVATTQ
GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDDGKLGLKVIDKDIATRQFIVEVENDGII
AKQKGVNIPNTKIPFPALAERDNADIRFGLEQGLNFIAISFVRTAKDVEEVREICRETGN
DHVQLFAKIENQQGIDNLDEIIEAADGIMIARGDMGIEVPFEMVPVFQKMIITKVNAAGK
AVITATNMLETMTEKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID
RNAQTLLNEYGRLDSSAFPRTNKTDVIASAVKDATHSMDIKLVVTITETGNTARAISKFR
PDADILAVTFDEKVQRALMINWGVIPVLAEKPASTDDMFEVAERVAVEAGLVQSGDNIVI
VAGVPVGTGGTNTMRVRTVK*
```

Figure 4



3-Dimentional image of Inhibitor (UDP-D-Xylulose) complex with Pyruvate Kinase.

Table-1:

Atom in protein	Atom in ligand	Hydrogen bond distance	Score
MET94.8D	06	3.302	56.05
GLU278:0E2	025	3.011	56.05
GLU278:0E1	025	2.687	56.05
ILE188:N	025	2.303	56.05
GLU93:0E1	022	2.706	56.05
ASP274:0D1	028	2.452	56.05

GOLD score representing the inhibitor (UDP-D-Xylulose) and Pyruvate kinase interaction. The molecular docking results show a fitness score of 56.05 between UDP-D-Xylulose with Pyruvate Kinase.

Acknowledgements:

The authors thank Bharath University, Chennai, India for their encouragement and support in performing this work.

References:

1. Alouf, J. E. (1980). "Streptococcal toxins (streptolysin O, streptolysin S, erythrogenic toxin)." Pharmacol Ther **11**(3): 661-717.
2. Alouf, J. E. and H. Muller-Alouf (2003). "Staphylococcal and streptococcal superantigens: molecular, biological and clinical aspects." Int J Med Microbiol **292**(7-8): 429-40.
3. Alouf, J. E., Muller-Alouf, H. and Kohler, W. (1999). The Comprehensive Sourcebook of Bacterial Protein Toxins, Academic Press, London.
4. Arcus, V. L., T. Proft, et al. (2000). "Conservation and variation in superantigen structure and activity highlighted by the three-dimensional structures of two new superantigens from Streptococcus pyogenes." J Mol Biol **299**(1): 157-68.
5. Beres, S. B., G. L. Sylva, et al. (2002). "Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence." Proc Natl Acad Sci U S A **99**(15): 10078-83.
6. Bisno, A. L. (1991). "Group A streptococcal infections and acute rheumatic fever." N Engl J Med **325**(11): 783-93.
7. Bisno, A. L. and D. L. Stevens (1996). "Streptococcal infections of skin and soft tissues." N Engl J Med **334**(4): 240-5.
8. Caparon, M. G., D. S. Stephens, et al. (1991). "Role of M protein in adherence of group\ A streptococci." Infect Immun **59**(5): 1811-7.
9. Ferretti, J. J., W. M. McShan, et al. (2001). "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Proc Natl Acad Sci U S A **98**(8): 4658-63.

10. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics. 2012 Oct 11.
11. Wei D, Jiang Q, Wei Y, Wang S. A novel hierarchical clustering algorithm for gene sequences. BMC Bioinformatics. 2012 Jul 23;13:174.
12. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010 Mar 1;26(5):680-2
13. Feng X, Xu Y, Chen Y, Tang YJ. MicrobesFlux: a web platform for drafting metabolic models from the KEGG database. BMC Syst Biol. 2012 Aug 2;6:94.
14. Moutselos K, Kanaris I, Chatziioannou A, Maglogiannis I, Kolisis FN. KEGGconverter: a tool for the in-silico modelling of metabolic networks of the KEGG Pathways database. BMC Bioinformatics. 2009 Oct 8;10:324.
15. Lu IL, Wang H. Protein-specific Scoring Method for Ligand Discovery. J Comput Biol. 2012 Nov;19(11):1215-26.
16. Dietzen M, Zotenko E, Hildebrandt A, Lengauer T. On the applicability of elastic network normal modes in small-molecule docking. J Chem Inf Model. 2012 Mar 26;52(3):844-56.
17. Plewczynski D, Łaźniewski M, Augustyniak R, Ginalski K Can we trust docking results? Evaluation of seven commonly used programs on PDB bind database. J Comput Chem. 2011 Mar. 32(4):742-55.

Corresponding Author:
Polani B Ramesh Babu*,
Email: *bioinfohod@gmail.com*