CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2009**, *8*, 42–62 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

January 2009, Volume 8, Number 4, Pages 42–62

Editor: Ovidiu Ivanciuc

Computational Prediction of Potent Therapeutic Targets of *Pseudomonas aeruginosa* and *In Silico* Virtual Screening for Novel Inhibitors

Pradeep K. Naik,[1] Seneha Santoshi,[1] and Ashima Birmani [1]

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173215, Himachal Pradesh, India.

Received: October 18, 2009; Revised: 24, 2009; Accepted: December 5, 2009; Published: December 17, 2009

Citation of the article:
P. K. Naik, S. Santoshi, and A. Birmani, Computational Prediction of Potent Therapeutic Targets of *Pseudomonas aeruginosa* and *In Silico* Virtual Screening for Novel Inhibitors, *Internet Electron. J. Mol. Des.* **2009**, *8*, 42–62, http://www.biochempress.com.

Inter*net* **Electronic** Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Computational Prediction of Potent Therapeutic Targets of *Pseudomonas aeruginosa* and *In Silico* Virtual Screening for Novel Inhibitors

Pradeep K. Naik,[1,*] Seneha Santoshi,[1] and Ashima Birmani [1]

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173215, Himachal Pradesh, India.

Received: October 18, 2009; Revised: 24, 2009; Accepted: December 5, 2009; Published: December 17, 2009

Internet Electron. J. Mol. Des. **2009**, 8 (4), 42–62

Abstract

Motivation. *Pseudomonas aeruginosa* is an ubiquitous pathogen capable of infecting virtually all tissues. The complete genome sequence of pathogen has provided a plethora of potential drug targets. While these data potentially contain all the determinants of host–pathogen interactions and possible drug targets, computational methods for selecting suitable candidates for further experimental analyses are currently limited.

Method. We have performed comparative analysis of whole genomes and metabolic pathways of the pathogen *P. aeruginosa* and the host *Homo sapiens* including symbiotic organisms. Moreover, the entire approach was built on the assumption that the potential target must play an essential role in the pathogen's survival and constitute a critical component in its metabolic pathway.

Results. We have predicted 6 unique essential genes in the pathogen using comparative genomics and 20 unique metabolic pathways based on comparative metabolomics. After critical evaluations of the targets we have finally considered dapD, gspL and pilA as the potent targets for virtual screening of lead molecules. Virtual screening was carried out using the high throughput virtual screening module of Glide and the hits with better glide score were further optimized by Glide–XP module. The PubChem molecule libraries (ChemDivision database, Diversity dataset, Kinase inhibitor database) were used for screening process.

Conclusions. Along with the high scoring results, the interaction studies provided promising ligands for future experimental screening to inhibit the proliferation of *Pseudomonas aeruginosa*.

Keywords. *Pseudomonas*; essential genes; comparative microbial genomics; comparative metabolomics; drug targets; virtual screening.

Abbreviations and notations

HTVS, high throughput virtual screening	T2SS, Type II protein secretion system
DEG, database of essential genes	

1 INTRODUCTION

The cost of research and development in the pharmaceutical industry has been rising steadily in the last decade, but the amount of time required bringing a new product to market remains around

* Correspondence author; phone: 91–1792–239–227; fax: 91–1792–245v362; E-mail: pknaik73@rediffmail.com.

ten to fifteen years [1]. It necessitates investment in inexpensive technologies that shorten the length of time spent in drug discovery. The target identification stage is the first step in the drug discovery process [2] and as such can provide the foundation for years of dedicated research in the pharmaceutical industry. This stage is complicated by the fact that the identified drug target must satisfy a variety of criteria to permit progression to the next stage. Important factors in this context include homology between target and host to prevent host toxicity such homology must be low or nonexistent [3] activity of the target in the diseased state [4,5] and the essentiality of the target to the pathogen's growth and survival [3,6,7]. The values of some of these selection criteria can be found easily by querying publicly available bioinformatics resources, including metabolic pathway databases such as KEGG (Kyoto encyclopedia of genes and genomes) [8] protein classification sets such as COGs (clusters of orthologous groups) [9] and databases of druggable (potentially useful as drug targets) proteins [5,10,11].

Availability of genome sequences of pathogens has provided a tremendous amount of information that can be useful in drug target identification [12,13].The strategies for drug design and development are progressively shifting from the genetic approach to the genomic and metabolomic approach. Novel drug targets are required in order to design new defense against antibiotic resistance pathogens. Comparative genomics and metabolomics provide new opportunities for finding optimal targets among previously unexplored cellular functions based on an understanding of their related biological processes in bacterial pathogens and their hosts. In general, a target should provide adequate selectivity; yielding a drug which is specific or highly selective against the pathogen with respect to the host [14,15]. Moreover, the target should be essential for growth and viability of the pathogen at least under the condition of infection. A gene is deemed to be essential if the cell cannot tolerate its inactivation by mutation, and its status is confirmed using conditional lethal mutants. The microbial target for treatment should not have any well–conserved homolog in the host, in order to address cytotoxicity issues. This can help to avoid expensive dead–ends when a lead target is identified and investigated in great detail only to discover at a later stage that all its inhibitors are invariably toxic to the host. Genes that are conserved in different genomes often turn out to be essential [9,16–19].

A simple and efficient computational tool designated as "subtractive genomics" and "comparative metabolomics" can determine concordances of putative gene products showing sets of proteins conserved across one set of user–specified genomes, but are not present in another set of user–specified genomes [20–22]. The functions encoded by essential genes are considered to constitute the foundation of life of the organism, and are therefore likely to be common to all cells [23]. Identification and characterization of essential genes for the establishment and/or maintenance of infection may be the basis to elaborate novel and effective antimicrobials against bacteria, especially if these genes are conserved in various bacterial pathogens. With the availability of Database of Essential Genes (DEG) [24], it is now possible to predict the list of essential genes in a

bacterial pathogen. It was suggestedthat searching for drug targets should be done among previously characterized proteins that are specific and essential for a particular pathogen [19]. Computational pathway analysis also facilitates the identification of enzymes that participate in several pathways. This approach relies on the fact that comparison of pathways present in host and in microbial pathogen reveal unique essential pathways present in the pathogen and further essential enzymes that are unique for the pathogen but absent in host thus suitable as the drug targets. Using metabolic pathway information as the starting point for the identification of potential targets has its advantages as each step in the pathway is validated as essential function for the survival of the bacterium.

In this study we have taken *Pseudomonas aeruginosa*, whole genomes and applied subtractive genomics and comparative metabolomics approaches in order to identify a list of potential therapeutic targets. *P. aeruginosa* is a Gram–negative bacterium and an opportunistic human pathogen. It mainly targets the immuno–compromised patients and typically infects the pulmonary tract, urinary tract and even causes blood infections. *P. aeruginosa* is highly resistant to a wide range of antibiotics and disinfectants [25].The pathogen has been reported to have lower outer membrane permeability to small molecules [26]. The presence of several multidrug efflux pumps from the major facilitator superfamily (MFS), multidrug and toxic compound extrusion (MATE) families, ATP–binding cassette (ABC) and small multi–drug resistance (SMR) have increased its intrinsic resistance to many efficient antibiotics. Thus, developing new antibacterial drugs against this pathogen has been a challenge. Concurrently, the recent availability of the human genome sequence [27,28] and the pathogen *P. aeruginosa* [25] represents a major step in drug discovery. These resources provide a basis for addressing the "complexities and conundrums" in drug discovery by computational methods. Here, we demonstrate the unprecedented potential of the available complementary genome datasets, and the application of both subtractive genomics and comparative metabolomics approaches for the identification of essential genes that may be considered as candidates for antibacterial drug discovery, using *P. aeruginosa* as an example. As a proof of concept, many of the genes identified by our approaches are also reported as essential by experimental methods. Furthermore, our approach successfully identified a number of promising protein targets for new antibiotic development.

2 MATERIALS AND METHODS

In this work we discuss the current state of the art for some of the bioinformatics approaches to identifying drug targets. It makes use of database of essential genes (DEG) (http://tubic.tju.edu.cn/deg/) [24] and the comparative genomic as well as comparative metabolomics approaches to compare with the pathogen bacteria versus human as well as its symbiotic bacteria, including identifying new members of successful target classes and their

functions, predicting disease relevant genes.

2.1 Identification of Unique Targets Based on Comparative Genomics

Protein sequences for *Homo sapiens* and *P. aeruginosa* were downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/). The DEG database [24] was downloaded from http://tubic.tju.edu.cn/deg/ and manually compiled to use as a stand–alone database for the BLAST program [29]. The *P. aeruginosa* genes were purged at 60% using CD–HIT to exclude paralogs from further analysis. The standalone BLAST executables including BLASTn, BLASTp and BLASTx were downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/) and installed locally. The set of essential genes in *P. aeruginosa* have been predicted based on homologous sequence search against DEG using BLASTn (E–value 10^{-8}). An interesting approach designated "differential genome display" has been proposed for the prediction of potential drug targets [20,21]. The resultant set of essential genes of pathogenic bacteria was subjected to BLASTn (E–value, 10^{-8}) against complete genome of non pathogenic bacteria (*Pseudomonas putida*; KT2440). The genes that are present in the genome of a pathogenic bacterium, but absent in the genome of a closely related free–living bacterium (non–pathogenic), are therefore likely to be important for pathogenicity and may be considered candidate drug targets. The subtracted essential genes of the pathogen were subjected to BLASTx (E–value, 10^{-5}) with complete human proteome to identify pathogen genes non–homologous in humans. The non–homologous entries were then subjected to BLASTx (E–value, 10^{-10}) against complete proteomes of four strains of symbiotic organisms: *Bacteroides thetaiotaomicron*, *Escherichia coli*, *Lactobacillus acidophilus* and *Lactobacillus johnsoni* for the identification of non–homologous unique essential genes in the pathogen designated as potent therapeutic targets (Table 1).

Table 1. Unique targets indentified in *Pseudomonas aeruginosa* (PA7) after four levels of subtraction based on comparative genomics.

No.	Gene Id	Gene	Protein name	Pathway
1	PSPA7_1473	dapD	tetrahydrodipicolinate succinylase	Lysine biosynthesis (pap00300)
2	PSPA7_1235	narH	nitrate reductase, beta subunit	Nitrogen metabolism (pap00910); Signal Transduction (pap02020)
3	PSPA7_1556	metE	5methyltetrahydropteroyltriglutamate – homocysteine methyltransferase	Methionine metabolism (pap00271)
4	PSPA7_3847	ccmF	cytochrome C–type biogenesis protein CcmF	Nitrogen metabolism (pap00910)
5	PSPA7_5673	hfq	RNA–binding protein Hfq	Not available
6	PSPA7_1508		putative inner membrane protein	Protein export (pap03060)

The identified genes were then classified into different groups based on gene names and subsequently matched against the list of essential *P. aeruginosa* genes identified by mutagenesis [30]. The overall protocol used for identification of therapeutic targets of *Pseudomonas aeruginosa*

based on comparative genomics is represented in Figure 1. Out of the list of unique essential genes mentioned in Table 1 only dapD was considered for virtual screening of lead molecules.

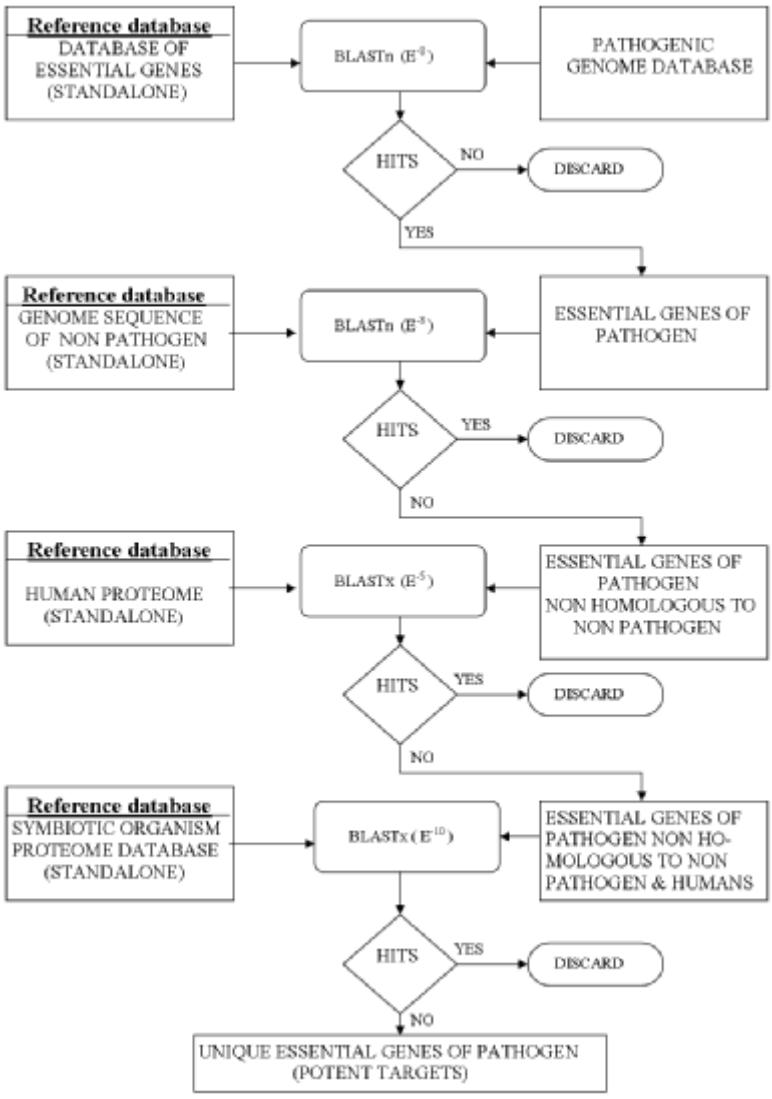


Figure 1. Flowchart for excluding essential genes/proteins of *Pseudomonas aeruginosa* having homology to non–pathogen, human and symbiotic organisms in human host.

2.2 Identification of Unique Enzymes in Unique Pathways

In this approach comparison between the metabolic pathways presnt in the host and the pathogen was done. Metabolic pathway information was obtained from the pathway database Kyoto Encyclopedia of Genes and Genomes [31]. We have compared all the pathways of *Pseudomonas aeruginosa* (KEGG i.d. –pap) with 213 pathways of human being and extracted the unique pathways. An important question to be addressed while choosing potential drug targets is whether the biochemical pathway to be targeted is unique to bacteria. The 20 pathways addressed in this study are all absent in the host *H. sapiens* and therefore unique to the pathogen *Pseudomonas aeruginosa*. The enzyme sequences in FASTA format from each metabolic pathway of

Pseudomonas aeruginosa were taken and did BLASTp with the human proteome with threshold E–value of 10^{-5}. These filtered enzymes were then compared with proteome of non pathogen using BLASTp with an E–value of 10^{-8}. The enzymes which were found to be uniquely present in the pathogen but absent in non pathogen were further compared with proteomes of symbiotic organisms using BLASTp with an E–value of 10^{-10}. All the enzymes which were found to be non homologous with symbiotic organisms were filtered out; they are unique enzymes and can be considered as therapeutic targets as listed in Table 2.

Table 2. Twenty two unique potential targets in *Pseudomonas aeruginosa* (PA7) predicted from comparative metabolomics of *P. aeruginosa* with *H. sapiens*, *P. Putida*, and symbiotic organisms.

c	Gene id	locus	EC id	Gene name	Protein name
AMINO ACID METABOLISM					
1	PSPA7_0967	K01826	EC:5.3.3.10	hpcD	5–carboxymethyl–2–hydroxymuconate isomerase
XENOBIOTICS BIODEGRADATION AND METABOLISM					
	1,2–Dichloroethane degradation				
1	PSPA7_4708	K01560	EC:3.8.1.2	dehII	putative haloacid dehalogenase
	gamma–Hexachlorocyclohexane degradation				
2	PSPA7_4709	K01561	EC:3.8.1.3	dehII	putative haloacid dehalogenase
	Benzoate degradation via hydroxylation				
3	PSPA7_0968	K01827	EC:5.3.3.11	hpcD	5–carboxymethyl–2–hydroxymuconate isomerase
FOLDING,SORTING AND DEGRADATION					
	Type II secretion system				
1	PSPA7_2029	K02452		xcpP	general secretion pathway protein C
2	PSPA7_2033	K02457		gspH2	general secretion pathway protein H
3	PSPA7_1411	K02458		gspI3	general secretion pathway protein I
4	PSPA7_2037	K02461		gspL	general secretion pathway protein L
5	PSPA7_4870	K02282		CpaE	pilus assembly protein
6	PSPA7_4872	K02279		cpaB	Flp pilus assembly protein
7	PSPA7_4873	02651		PilA	pilus assembly protein
8	PSPA7_5780	K02663		PilN	type IV pilus assembly protein
9	PSPA7_5779	K02664		pilO	type IV pilus assembly protein
10	PSPA7_5195	K02675		pilY2	type IV pilus assembly protein
11	PSPA7_5194	K02674		pilY1	type IV pilus assembly protein
12	PSPA7_5193	K02673		pilX	type IV pilus assembly protein
13	PSPA7_5192	K02672		pilW	type IV pilus assembly protein
	Type IV secretion system				
1	PSPA7_3697	K03195		VirB1	conjugation TrbI family protein
2	PSPA7_3698	K03204		VirB9	conjugal transfer protein VirB9
3	PSPA7_3699	03200		VirB5	type IV secretion system protein VirB5
4	PSPA7_3703	K03199		VirB4	ATPase
5	PSPA7_3704	K03198		VirB3	conjugal transfer protein type IV secretion

However, only GspL and pilA were considered for virtual screening of lead molecules. The overall protocol used for identification of therapeutic targets of *Pseudomonas aeruginosa* based on comparative metabolomics is represented in Figure 2.

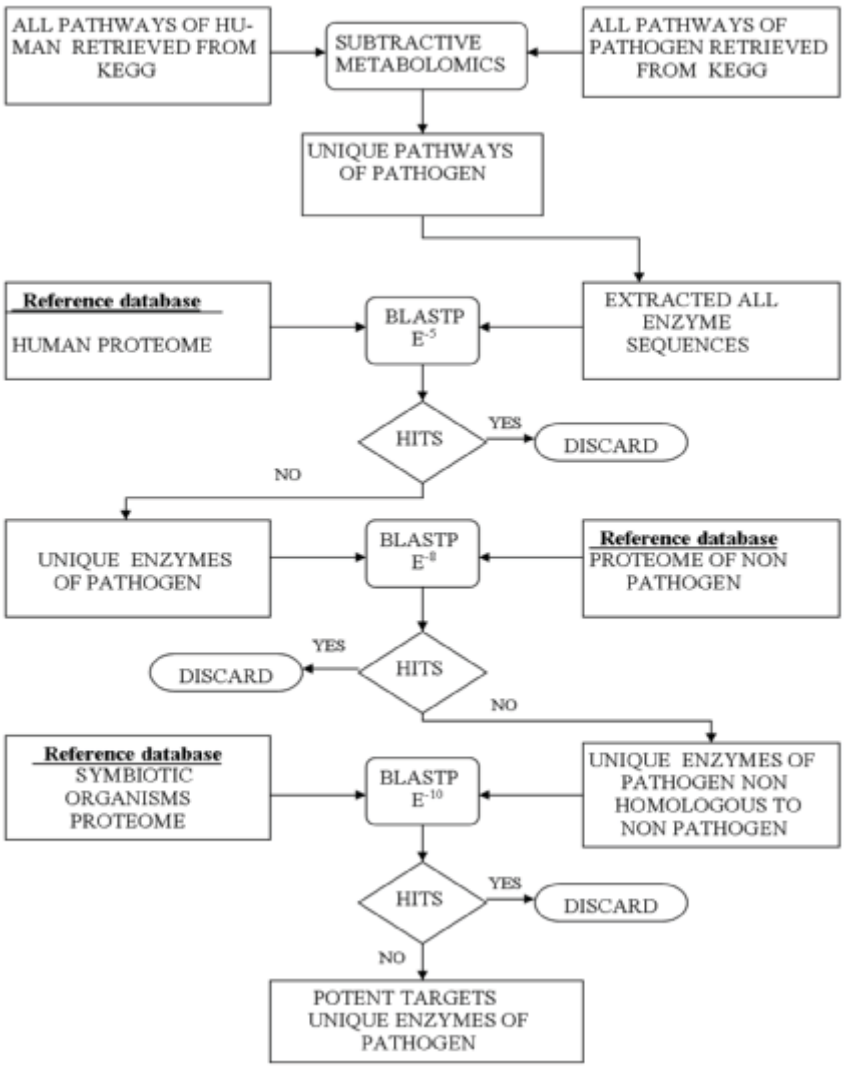


Figure 2. Flowchart for excluding metabolic pathways in *Pseudomonas aeruginosa* having homology to non–pathogen, human and symbiotic organisms in human host.

2.3 Sequence Analysis

Domain analysis (ProDom, Pfam) was carried out to check whether any domain present in the target protein could interfere in the action of the drug. ProDom, a database of protein domain families, is useful for analyzing the protein domain arrangements and helps to analyze homology relationship in modular proteins. The ProDom building procedure MKDOM2 is based on recursive PSI–BLAST searches. The source protein sequences are non–fragmentary, derived from SWISS–PROT and TrEMBL databases. Whereas, Pfam, a database of multiple alignments and HMM profiles of protein domains or conserved protein regions, is actually composed of two sets of families. Pfam–A family is based on curated multiple alignment whereas Pfam–B family is derived from ProDom, a comprehensive set of protein domain family. The screening of the database in this study proved that the domains present in dapD, GspL and pilA of *pseudomonas aeruginosa* are not ubiquitous in human. Finally the proteins that have been selected (viz. dapD, GspL and pilA) were subsequently matched against the list of essential *P. aeruginosa* genes identified by mutagenesis

[30] and were found to be significant.

BLASTp of dapD protein sequence against PDB databank gave three significant hits: 1EFD_N, 1K2V_N and1K7S_N. These proteins were different forms of tetrahydrodipicolinate N–succinyltransferase from *E. coli* and *Mycobacterium bovis* and had given a hit of 37% identity, 56% positive scores and 4% gaps with an expect value of 0.78. Again these sequences were aligned using clustalX [32], obtaining a score of 50. Similarly BLASTp of GspL sequence with PDB databank gave three significant hits: 1W97_L, 1YF5_L and 2BH1_A. These proteins were different forms of general secretion pathway protein L from *Vibrio cholerae* and had given a hit of 34% identity, 60% positive scores and 6% gaps with an expect value of 0.22. Again these sequences were aligned using clustalX [32], obtaining a score of 53. Similarly BLASTp of pilA sequence with PDB databank gave one significant hits: 2JTY. This protein was different forms of pilus assembly protein (Type I fimbrial protein) from *E. coli* and had given a hit of 39% identity, 49% positive scores and 13% gaps with an expect value of 1.8. This provided the broadly studied templates to model the target proteins.

2.4 Homology Model Construction

The homology models of the proteins: dapD, GspL and pilA were built using Prime (Prime version 1.5, Macromodel version 9.1, Schrodinger, LLC, New York, NY, 2005) accessible through the Maestro interface (Schrodinger, Inc.). All water molecules were removed and the bound ligand was kept. During the homology model building, Prime keeps the backbone rigid for the cases in which the backbone does not need to be reconstructed due to gaps in the alignment. The model was screened for unfavorable steric contacts and remodeled using a rotamer library database of PRIME. Explicit hydrogens were added to the protein and the protein model subjected to energy minimization using the Macromodel (Prime version 1.5) force–field MMFFS. Energy minimization and relaxation of the loop regions was performed using 300 iterations in a simple minimization method. Again the steepest descent was carried out until the energy showed stability in the sequential repetition. Model evaluation was performed in PROCHECK v3.4.4 [33] producing plots that were analyzed for the overall and residue–by–residue geometry. Ramachandran Plot [34] provided by the program PROCHECK assured very good confidence for the predicted protein. There were only 0.3% residues in the disallowed region and 0.9% residues in generously allowed regions. Nevertheless, PROCHECK assured the reliability of the structure and the protein was subjected to VERIFY3D [35] available from NIH MBI Laboratory Servers

2.5 Ligand Binding Site Prediction

In silico prediction of binding sites of these proteins was done using SiteMap (Schrodinger Inc.). SiteMap treat entire proteins to locate binding sites whose size, functionality, and extent of solvent exposure meet user specifications. SiteScore, the scoring function used to assess a site's propensity

for ligand binding, accurately ranks possible binding sites to eliminate those not likely to be pharmaceutically relevant. It identifies potential ligand binding sites by linking together "site points" that are suitably close to the protein surface and sufficiently well sheltered from the solvent. Given that similar terms dominate the site scoring function, this approach ensures that the search focuses on regions of the protein most likely to produce tight protein–ligand or protein–protein binding. Subsites are merged into larger sites when they are sufficiently close and could be bridged in solvent–exposed regions by ligand atoms. SiteMap evaluates sites using a series of properties. The binding site with highest site score was taken for docking and virtual screening of the lead molecules.

2.6 Docking Studies

A library of 1,25,000 compounds taken from pubchem drug database (http://zinc.docking.org/vendor0/index_nfs.shtml) were compiled together to form a standalone library. These compounds were available in 3D–MOL2 file. These molecules were imported into maestro and finally prepared using ligPrep. LigPrep is a utility of Schrodinger suit that combines tools for generating 3D structures from 1D and 2D representation, searching for tautomers and steric isomers and perform a geometry minimization of ligands. The Schrodinger Glide program version 4.0 has been used for docking. The best 10 poses and corresponding scores have been evaluated using Glide high throughput virtual screening (HTVS) (Glide HTVS) for each ligand from the virtual libarary. For each screened ligand, the pose with the lowest Glide HTVS score has been taken as the input for the Glide calculation in extra precision mode (Glide XP). To soften the potential for non–polar parts of the receptor, we scaled vander Waals radii of receptor atoms by 1.00 with partial atomic charge 0.25.

3 RESULTS AND DISCUSSION

One major criterion of the modern drug discovery is the specificity and accuracy. The drug should never disturb any normal process of the host system. The objective of the work was to find and locate those essential genes of *P. aeruginosa* that play important roles in the normal functioning of the bacterium within the host and to shortlist them in the view of drug targeting. Identification of non–human homologs in the essential genes of *P. aeruginosa* with subsequent screening of the proteome to find the corresponding protein products are likely to lead to development of drugs that specifically interact with the pathogen. Computational methods have been previously employed for the identification of probable drug targets by predicting only the essential genes in pathogen bacteria [36,15]. However, these works simply compare the bacterial genome sequences with host genome only to discard those essential genes which are homologous to host. However, comparison with the symbiotic bacterial genome sequences if any present within the host cannot be ruled out for

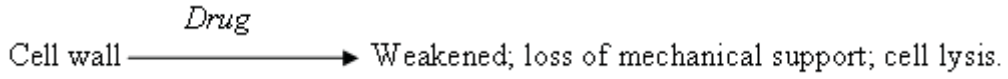
predicting the potential therapeutic targets. Including those genes whose products have sequence similarities with the host and its symbiotic bacterial species if any, may lead to drug reactions with the host and thus led to toxic effects. Therefore, here we have excluded those essential genes having sequence similarities with the host and its symbiotic bacterial genes (in case of the human as host) and considers only unique essential genes which present only in the pathogen bacteria.

3.1 Comparative Genomics of *P. aeruginosa* with the Host

We have used three step processes to predict possible therapeutic targets from the pathogenic bacteria. First the essential genes of pathogenic bacteria were compared with essential genes of non–pathogenic bacteria using BLASTn and those genes which showed similarity between them were rejected. The remaining essential genes from pathogenic bacteria were then subtracted from human genome and symbiotic organisms to obtained unique essential genes (probable therapeutic targets). We identified 183 unique essential genes in bacterial species. It is revealed that out of 183 essential genes only 43 genes are uniquely present in *P. aeruginosa*, which could be related with pathogenicity. After comparison with human proteome (BLASTx), only 33 essential genes were uniquely identified from *P. aeruginosa*. Further comparison with four symbiotic organisms (residing within human being), only 6 unique therapeutic target genes were identified in *P. aeruginosa* (Table 1). These genes were classified using the *P. aeruginosa* annotation table available at http://www.bioinfo.de/isb/2004/04/0028/www.pseudomonas.com and were also confirmed with the list of experimentally determined essential genes of *P. aeruginosa* generated by Jacobs et al., 2003. Nonetheless, since gene disruption data are not available for all the genes in all the pathogens, this approach makes it possible to hazard a "first–order guess" for the probability that any untested gene is essential and may be a probable drug target. The 6 protein components short–listed in our study underlies the important aspects of host–pathogen interaction. It might therefore be legitimate to consider that inactivation of some of these protein through designing potent lead molecules would likely result in inactivation of the pathogen. These proteins would therefore represent promising candidates for further study and characterization with an intention for lead design.

The targeted gene that has been selected further for lead design includes dapD. It encodes for Tetrahydrodipicolinate succinylase. Tetrahydrodipicolinate N–succinyltransferase (dapD) catalyzes the succinyl–CoA–dependent acylation of L–2–amino–6–oxopimelate to 2–N–succinyl–6–oxopimelate as part of the succinylase branch of the meso–diaminopimelate/lysine biosynthetic pathway of bacteria, blue–green algae, and plants. This pathway provides meso–diaminopimelate as a building block for cell wall peptidoglycan in most bacteria, and is regarded as a target pathway for antibacterial agents [37]. Bacterial biosynthesis of lysine has come under increased scrutiny as a target for novel antibacterial agents as it provides both lysine for protein synthesis and meso–diaminopimelate for construction of the bacterial peptidoglycan cell wall [38]. Amino acid lysine is

produced from aspartate through the diaminopimelate (DAP) pathway in most bacteria and higher plants. In bacteria, DAP is not only a direct precursor of lysine, but it is also an important constituent of the cell wall peptidoglycan [39]. The DAP pathway is of special interest for pharmacology, since the absence of DAP in mammalian cells allows for the use of the DAP biosynthetic genes as a bacteria–specific drug target.



3.2 Pathways and Enzymes Unique to *P. aeruginosa* Compared to *H. sapiens*

The targeting of biosynthetic pathways has several advantages. Each step in the pathway is already well validated as an essential function for bacterial growth. Discarding target enzymes from the pathogen which share a similarity with the host proteins ensures that the targets have nothing in common with the host proteins, thereby, eliminating undesired host protein–drug interactions. Metabolic pathways belonging to the pathogen and the host are compared and pathways present in the pathogen but not in the host are considered as unique pathway. Comparative metabolic pathway analysis resulted in 20 unique metabolic pathways in *P. aeruginosa*. Enzymes, from these pathways, which do not show similarity to any of the host proteins and the symbiotic organisms, are listed in Table 2. Amongst these pathways, we have investigated the type–II secretion pathway and type IV pilus assembly protein as potential target for designing the lead molecules in this study.

3.3 Type –II Secretion Pathway

Pseudomonas aeruginosa utilizes a number of distinct pathways to secrete proteins that play various roles during infection. These include the type II secretion system, which is responsible for the secretion of the majority of exoproducts into the surrounding environment, including toxins and degradative enzymes [40]. Type II protein secretion system (T2SS) is widely distributed among most Gram–negative bacteria for secretion of extracellular degradative enzymes and toxins [40]. Based on experimental data with *P. aeruginosa* T2SS component proteins [41,42], a type II secretion apparatus model formed by the multi–protein complex that spans the entire cell envelope has been proposed. T2SS was first discovered in *Klebsiella pneumoniae*, where it was found to be required for secretion of lipoprotein pollulanse [43]. The conservation of T2SS was then identified in many Gram–negative bacteria such as *E. coli*, *Erwinia carotovora*, *Yesiniae entrocolitica*, and so on [44]. The genes encode type II secretion components are usually clustered [45]. So far the precise functions of individual component proteins are not very well characterized. It is clear that mutations in most of the T2SS component genes result in abortion of the secretion process and accumulation of the protein in the periplasm [46]. The T2SS apparatus spans entire cell envelop without extracellular filamentous appendage, the only exception is the type IV pilus. Various experimental approaches have been used for studying the secretion signal of the protein secreted

through T2SS, and it turns out that the multiple sites in the T2SS–secreted proteins are, instead of being treated as secretion signal, needed for recognition of and targeting to the T2SS apparatus [47]. General secretory pathway protein L is involved in the type II secretion system which is responsible for the secretion of the majority of exoproducts into the surrounding environment, including toxins and degradative enzymes [48].

3.4 Type IV Pilus Assembly Protein

Pseudomonas aeruginosa is an ubiquitous pathogen capable of infecting virtually all tissues. A large variety of virulence factors contribute to its importance in burn wounds, lung infection and eye infection. Prominent factors include pili, flagella, lipopolysaccharide, proteases, quorum sensing, exotoxin A and exoenzymes secreted by the type III secretion system [49]. Infection of a host by pathogenic bacteria requires that the bacteria are able to bind to the target tissue to colonize and/or begin the process of invasion. This adhesion event involves specific interactions between receptors on the host tissue and external surface structures produced by the bacterial cell. Gram–positive bacteria utilize a special class of surface–anchored proteins for this purpose, while Gram–negative microorganisms use a more complex structure, called a pilus, to achieve the same end. In both cases, a highly conserved pathway is utilized to export, assemble and anchor these surface structures. As such, these pathways represent targets for antibiotic development. Compounds that prevent the assembly of bacterial surface proteins will cripple the ability of bacteria to interact with and colonize host tissue leading to rapid bacterial clearance from the body [50]. Mutants lacking pili (pilA), showed reduced measured attachment compared with the wild–type strain. Pili mutants also showed reduced pathogenicity in a model insect host. The deletion of pilA in gram negative bacteria decreases adherence to cultured respiratory cell lines. Data suggest that pilA may be an important mediator of the pathogenic process in humans and should be considered as a target in future attempts to generate a protective vaccine [51].

3.5 Virtual Screening of Lead Molecules

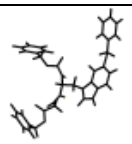
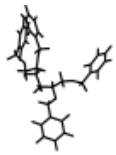
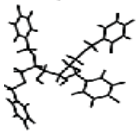
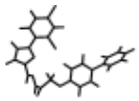
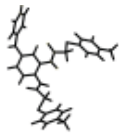
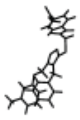
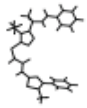
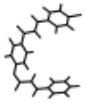
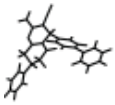
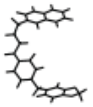
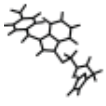
The impact of microbial genomics on drug discovery has led to the identification of novel antibacterial drugs. Enzymes mediate the synthesis of many complex molecules from simpler ones in a series of chemical reactions. Targeting enzymes present in the pathogen but absent in the host ensures the elimination of pseudo drug targets in the pathways [19]. The computational genomics [15] as well as comparative metabolomics approaches stated herein, are likely to speed up drug discovery process by removing hindrances like dead ends or toxicity that are encountered in classical approaches. Our initial work in sequence analysis and domain study suggested to pick up dapD, the key enzyme in lysine biosynthesis, gspL the key enzyme in type II secretory system and pilA , a key type IV pilus assembly protein in *P. aeruginosa*. The strategy is also likely to locate critical pathways and stages in the pathogenecity of *P. aeruginosa*. Sequence as well as structural

analysis was carried out to avoid any interference in the normal function of the human host system. Pairwise sequence alignment study of dapD, gspL and pilA in BLAST was performed against human database as well as the symbiotic organisms. Although, in general, sequence dissimilarity means possible dissimilar binding sites, however, exceptions where targets with significant sequence dissimilarity bind similar molecules do exist. Presumably, screening against such novel targets for functional inhibitors will result in discovery of novel therapeutic compounds active against bacteria, including the increased number of antibiotic resistant clinical strains [52].

Homology modeling protocol was employed to predict the 3D structure of the selected proteins viz. dapD, gspL and pilA. Three dimensional structure predictions by comparative modeling were done by Prime (Schrodinger Inc.). We used both PROCHECK and the VERIFY3D softwares to check the quality of the modeled protein. Ramachandran Plot obtained from the program PROCHECK, which checks the stereochemical quality of a protein structures, producing a number of postscript plots, analyzing its overall and residue–by–residue geometry, assured the reliability of the modeled protein with 91.1%, 92.6%, 91.3% residues in most allowed region and 7.7%, 6.3%, 7.5% in additional allowed region respectively. There were only 0.36%, 0.41%, 0.35% residues in disallowed region and 0.84%, 0.69% and 0.85% respectively in generously allowed region. The assessment with VERIFY3D, which derives a ‘‘3D–1D’’ profile based on the local environment of each residue, described by the statistical preferences for: the area of the residue that is buried, the fraction of side–chain area that is covered by polar atoms (oxygen and nitrogen), and the local secondary structure, also substantiated the reliability of the three dimensional structure. The residues that deviated from the standard conformational angles of Ramachandran plot were found to be more than 10 A° distance to the active site residues, which suggested that those residues would interfere little with the binding of ligands in the active site region. The binding sites were predicted using Sitemap (Schrodinger Inc.). For each targeted protein the binding site with maximum site score was selected for docking and virtual screening of lead molecules using Glide (Schrodinger Inc.). The standard methodology accepted for virtual screening with the program Glide involves, applying the parameters of library screening followed by the standard default setting. Library screening was done to filter the non–docked compounds from database we used for virtual screening. After filtering non–docked compounds, remaining compounds were used for detailed docking. Docking procedure consisted of three interrelated components; (a) identification of binding site, (b) a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface) and (c) a scoring function. For each ligand in the virtual library, the pose with the lowest Glide (HTVS) score was refined using Glide (XP) docking. The best 20 ligands as shown in Table 3–5 respectively, chosen with Glide score proved their reliability in the Glide (XP). Since computational screenings always demand experimental testing in order to confirm the accurate drug molecule(s), the proposed LEAD molecules need to be optimized in further studies. The significance of this work is in providing a relatively inexpensive

approach to screen compounds that are likely to inhibit the action of dapD, GspL and pilA in *Pseudomonas aeruginosa*.

Table 3. Database ID number, chemical structure, IUPAC names and Glide scores of top scored ligands docked with dapD of *Pseudomonas aeruginosa* using docking program Glide

No.	Ligand	ZINC ID	Ligand Name	Glide (HTVS) score	Glide (XP) score
1		ZINC04014946	phenylmethyl 2–[[2–(phenylmethoxycarbonylamino)–3–[5–(phenylmethoxy)–1H–indol–3–yl] propanoyl]amino]acetate	–7.01	–15.65
2		ZINC03994289	(2R,3S,4S,5R)–1,6–bis[(2,6–difluorophenyl)methoxy]–2,5–bis(phenylmethoxy)hexane–3,4–diol	–7.79	–15.19
3		ZINC03994290	(2R,3S,4S,5R)–1,6–bis[(2–fluorophenyl)methoxy]–2,5–bis[(4fluorophenyl) methoxy]hexane–3,4–diol	–8.73	–14.77
4		ZINC03146249	N–[[5–(2,4–dichlorophenyl)furan–2–yl]methylideneamino]–2–(4–phenylphenoxy)acetamide	–7.56	–13.56
5		ZINC02154335	N–[5–(benzoyl)–2–[[2–(4–methylphenyl)sulfanylacetyl]amino]phenyl]–2–(4–methylphenyl)sulfanylacetamide	–7.71	–13.50
6		ZINC00916256	N–(4–acetylphenyl)–2–[[5–[(3–methylphenoxy)methyl]–4–(phenylmethyl)–1,2,4–triazol–3–yl]sulfanyl]acetamide	–7.02	–13.40
7		ZINC02163698	N–[1–[4–ethyl–5–[2–[(5–methyl–4–phenyl–1,3–thiazol–2–yl)amino]–2–oxoethyl]sulfanyl–1,2,4–triazol–3–yl]ethyl]benzamide	–7.66	–13.22
8		ZINC03985979	[4–[3–(4–chlorophenyl)prop–2–enoyl]phenyl] 3–phenylprop–2–enoate	–7.13	–13.11
9		ZINC00728134	2–amino–5–oxo–7–phenyl–4–(4–phenylphenyl)–4,6,7,8–tetrahydrochromene–3–carbonitrile	–7.50	–12.99
10		ZINC02952205	3–[4–[4–(1,3–benzodioxol–5–ylmethyl)piperazin–4–ium–1–yl]–4–oxobutyl]–1H–quinazoline–2,4–dione	–7.86	–12.95
11		ZINC04059724	N–(furan–2–ylmethyl)–2–[1–[(4–methylphenyl)methyl]indol–3–yl]	–7.15	–12.92

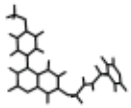
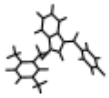
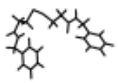
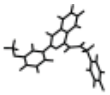
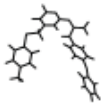
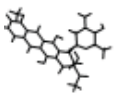
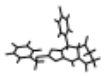
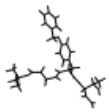
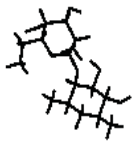
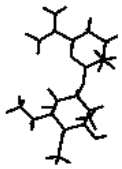
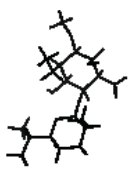
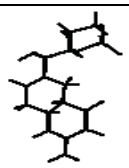
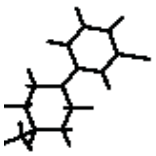
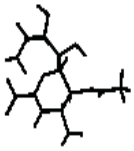
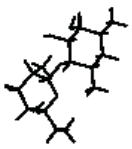
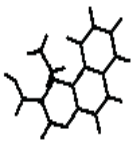
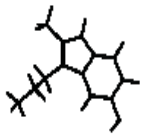
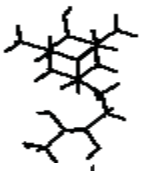
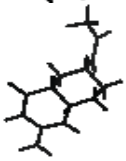
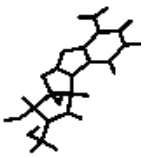
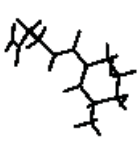
12		ZINC04000222	2–[4–(4–methoxyphenyl)–2–oxochromen–7–yl]oxy–N–(pyridin–4–ylmethyl)acetamide	–7.20	–12.89
13		ZINC03592328	N–(4–fluorophenyl)–2–[(2Z,5R)–2–[2–[(4–methylphenyl)amino]–2–oxoethylidene]–4–oxo–1,3–thiazolidin–5–yl]acetamide	–7.98	–12.82
14		ZINC04059664	3–[(2,5–dimethylphenyl)methylsulfanyl]–1–[(4–fluorophenyl)methyl]indole	–7.33	–12.76
15		ZINC04001913	3–[3–oxo–3–(phenylmethylamino)propyl]disulfanyl–N–(phenylmethyl)propanamide	–7.16	–12.74
16		ZINC02941742	N–[(4–fluorophenyl)methyl]–3–(3–methoxyphenyl)–4–oxophthalazine–1–carboxamide	–7.47	–12.73
17		ZINC03869146	N–(6–{[(4–methylphenyl)thio]methyl}–4–oxo–1,4–dihydro–2–pyrimidinyl)–N''–(4–phenoxyphenyl)guanidine	–7.18	–12.68
18		ZINC03830633	[6–[(3–acetyl–3,5,12–trihydroxy–10–methoxy–6,11–dioxo–2,4–dihydro–1H–tetracen–1–yl)oxy]–3–hydroxy–2–methyloxan–4–yl]azanium	–7.06	–12.66
19		ZINC02212519	(8aR,9S)–9–(4–fluorophenyl)–2–(phenylmethylsulfanyl)–6,7,8a,9–tetrahydro–4H–[1,2,4]triazolo[5,1–b]quinazolin–8–one	–7.11	–12.65
20		ZINC04016474	tert–butyl N–[2–[[1–[(1–amino–3–hydroxy–1–oxobutan–2–yl)amino]–1–oxo–3–[4–(phenylmethoxy) phenyl]propan–2–yl]amino]–2–oxoethyl]carbamate	–7.10	–12.63

Table 4. Database ID number, chemical structure, IUPAC names and Glide scores of top scored ligands docked with GspL of *Pseudomonas aeruginosa* using docking program Glide.

No.	Ligand	ZINC ID	Ligand Name	Glide (HTVS) score	Glide (XP) score
1		ZINC03831196	5–amino–2–(aminomethyl)–6–(4,6–diamino–2,3–dihydroxycyclohexyl)oxyoxane–3,4–diol	–7.45	–10.54
2		ZINC03830242	2–amino–3–[3–amino–6–(1–aminoethyl) oxan–2–yl]oxy–6–methoxy–5–(methylamino) cyclohexane–1,4–diol	–6.86	–7.83
3		ZINC03830241	2–amino–3–[3–amino–6–(1–aminoethyl)oxan–2–yl]oxy–6–methoxy–5–(methylamino)cyclohexane–1,4–diol	–6.91	–7.49

4		ZINC03870170		–6.62	–7.48
5		ZINC02383344	1–(6–chloropyridin–2–yl)piperidin–4–amine	–6.86	–7.24
6		ZINC03870976	3–acetamido–4–amino–2–(3–amino–1,2–dihydroxypropyl)–3,4–dihydro–2H–pyran–6–carboxylic acid	–6.52	–7.17
7		ZINC03831194	5–amino–2–(aminomethyl)–6–(4,6–diamino–2,3–dihydroxycyclohexyl)oxyoxane–3,4–diol	–7.33	–6.61
8		ZINC02388213	[2–(hydroxymethyl)–2,3–dihydro–1H–benzo[f]chromen–1–yl]–methylazanium chloride	–6.03	–6.18
9		ZINC02572146	3–(2–aminoethyl)–2–methyl–1H–indol–5–ol	–6.66	–6.12
10		ZINC04024474		–6.12	–6.07
11		ZINC03869363		–7.39	–5.80
12		ZINC03834078	(1S,2R,5R)–5–(4–aminoimidazo[4,5–c]pyridin–1–yl)–3 hydroxymethyl) cyclopent–3–ene–1,2–diol	–6.50	–5.65
13		ZINC03839685	3–amino–N–(3–amino–3–oxopropyl)–4,5–dihydroxycyclohexene–1–carboxamide	–6.77	–5.60
14		ZINC03683350		–6.07	–5.50

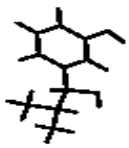
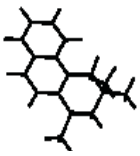
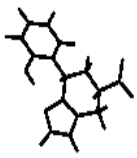
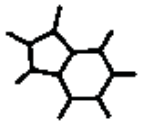
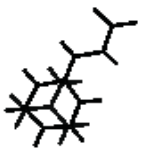
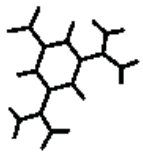
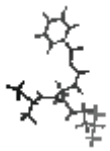
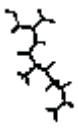
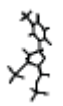
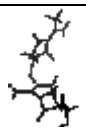
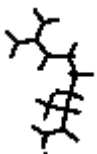
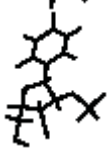
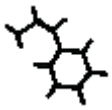
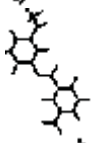
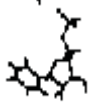
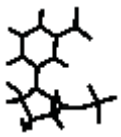
15		ZINC01678601		–6.19	–5.46
16		ZINC02479996	2,2,4–trimethyl–3,4–dihydro–1H–benzo[f] isoquinoline	–6.70	–5.43
17		ZINC03984186	4–(2–hydroxyphenyl)–4,5,6,7–tetrahydro–1H–imidazo[4,5–c]pyridine–6–carboxylic acid	–6.92	–5.27
18		ZINC01532171	Indole–2,3–dione 2,3–Indolinedione	–6.35	–5.23
19		ZINC03162497		–6.61	–5.17
20		ZINC03161158	5–nitrobenzene–1,3–dicarboximidamide	–6.04	–5.13

Table 5. Database ID number, chemical structure, IUPAC names and Glide scores of top scored ligands docked with pilA of *Pseudomonas aeruginosa* using docking program Glide.

No.	Ligand	ZINC ID	Ligand Name	Glide (HTVS) score	Glide (XP) score
1		ZINC03871693		–8.90	–9.08
2		ZINC01888932	(2S)–2–[[(2S)–2,6–diaminohexanoyl]amino]–3–hydroxypropanoic acid	–5.46	–6.89
3		ZINC03786623		–5.77	–6.71
4		ZINC03870884	5–(5–methyl–2,4–dioxopyrimidin–1–yl)–2–(phosphonooxymethyl)oxolan–3–yl	–5.28	–6.24
5		ZINC03870179	5–(2,4–dioxo–1H–pyrimidin–5–yl)–3,4–dihydroxyoxolan–2–yl]methyl dihydrogen phosphate	–4.61	–6.24

6		ZINC03780837	3–(2–methylpiperidin–1–ium–1–yl)propyl 4–(3–methylbutoxy)benzoate chloride	–5.44	–6.22
7		ZINC06483376	2–(diaminomethylideneamino)ethyl methyl hydrogen phosphate	–4.97	–6.19
8		ZINC06487269	4–azanidylpentan–2–ylazanide; dichloroplatinum(2+)	–5.52	–6.03
9		ZINC03871599	Pentanamide, 2–amino–5–[(aminoiminomethyl)amino]–	–5.10	–5.83
10		ZINC03875251	2–(2,4–dioxopyrimidin–1–yl)–4–hydroxy–5–(hydroxymethyl)oxolan–3–yl	–4.83	–5.78
11		ZINC03786627	2–[(2R)–piperidin–2–yl]ethanamine	–4.72	–5.72
12		ZINC00006017	5–chloro–7–(trifluoromethyl)–1,4–dihydroquinoxaline–2,3–dione	–4.79	–5.60
13		ZINC03870260	disodium [5–(2,4–dioxopyrimidin–1–yl)–3,4–dihydroxyoxolan–2–yl]methyl	–4.21	–5.46
14		ZINC00262043	1–[3–[(2–hydroxy–5–nitrophenyl)methylideneamino]phenyl]ethanone	–4.18	–5.35
15		ZINC03870177	5–(2,4–dioxo–1H–pyrimidin–5–yl)–3,4–dihydroxyoxolan–2–yl]methyl dihydrogen phosphate	–5.10	–5.22
16		ZINC03870733	3,4–dihydroxy–5–(phosphonatooxymethyl)oxolan–2–yl]methyl phosphate	–4.57	–5.16
17		ZINC03869332	5–[(2–aminoacetyl)amino]–3,4–dihydroxyoxolan–2–yl]methyl dihydrogen phosphate	–5.75	–5.14
18		ZINC03870114	4–[(4–aminophenyl)sulfonylamino]benzoic acid	–4.36	–5.11
19		ZINC02051031	2–(3–oxo–4H–1,4–benzoxazine–6–carbonyl)benzoate	–4.74	–5.08

4 CONCLUSIONS

The data presented here demonstrate that stepwise prioritization of genome open reading frames using simple biological criteria can be an effective way of rapidly reducing the number of genes of interest to an experimentally manageable number. This process is an efficient way for enriching potential target genes, and for identifying those that are critical for normal cell function. The generation of a comprehensive essential gene list will allow an accelerated genetic dissection of traits such as metabolic flexibility and inherent drug resistance that render *P. aeruginosa* such a tenacious pathogen. Such a strategy will enable us to locate critical pathways and steps in pathogenesis; to target these steps by designing new drugs; and to inhibit the infectious agent of interest with new antimicrobial agents. We propose probable chemical compounds, which could be tested to devise drug molecules to retard the hazardous proliferation of *P. aeruginosa*. The scope of this work could be to use this data to do cost–effective experimental screening. The proposed potential chemical compounds could provide the prime lead for future experimental screening.

5 REFERENCES

[1] F. Humer, Innovation in the Pharmaceutical Industry—Future Prospects. Available: http://www.roche.com/fbh_zvg05_e.pdf. **2005**.
[2] G.C. Terstappen and A. Reggiani, In silico research in drug discovery. *Trends Pharmacol Sci*. **2001**, *22*, 23–26.
[3] Freiberg, C. Novel computational methods in anti–microbial target identification. *Drug Discovery Today* **2001**, *6*, S72–S80.
[4] Wang, S.; Sim, T.B.; Kim, Y.S.; Chang, Y.T. Tools for target identification and validation. *Curr Opin Chem Biol*, **2004**, *8*, 371–377.
[5] Sanseau, P. Impact of human genome sequencing for in silico target discovery. *Drug Discov Today* **2001**, *6*, 316–323.
[6] Roemer, T.; Jiang, B.; Davison, J.; Ketela, T.; Veillette, K. Large–scale essential gene identification in Candida albicans and applications to antifungal drug discovery. *Mol. Microbiol*. **2003**, *50*, 167–181.
[7] Sassetti, C.M.; Boyd, D.H.; Rubin, E.J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol*. **2003**, *48*, 77–84.
[8] Wixon, J.; Kell, D. The Kyoto encyclopedia of genes and genomes—KEGG. *Yeast* **2000**, *17*, 48–55.
[9] Tatusov, R.L.; Koonin, E.V.; Lipman, D.J. A genomic perspective on protein families. *Science* **1997**, *278*, 631–637.
[10] Hopkins, A.L.; Groom, C.R. The druggable genome. *Nat. Rev. Drug Discov*. **2002**, *1*, 727–730.
[11] Robertson, J.G. Mechanistic basis of enzyme–targeted drugs. *Biochemistry* **2005**, *44*, 8918.
[12] Wheeler, D.L.; Church, D.M.; Edgar, R.; Federhen, S.; Helmberg, W.; Madden, T.L.; Pontius, J.U.; Schuler, D.G.; Schriml, L.M.; Sequeira, E.; Suzek, T.O.; Tatusova, T.A.; Wagner, L. Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res*. **2004**, *32*, D35–D40.
[13] Miesel, L.; Greene, J.; Black, T.A. Genetic strategies for antibacterial drug discovery. *Nat. Rev. Genet*. **2003**, *4*, 442–456.
[14] Sakharkar, K.R.; Sakharkar, M.K.; Chow, V.T. 2006. Gene fusion in *Helicobacter pylori*: making the ends meet. *Antonie Van Leeuwenhoek* **2006**, *89*, 169–180.
[15] Sakharkar, K.R.; Sakharkar, M.K.; Chow, V.T.K. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol*. **2004**, *4*, 0028.
[16] Itaya, M. An estimation of minimal genome size required for life. *FEBS Lett*. **1995**, *362*, 257–260.
[17] Koonin, E.V.; Tatusov, R.L.; Galperin, M.Y. 1998. Beyond complete genomes: from sequence to structure and function. *Curr. Opin. Struct. Biol*. **1998**, *8*, 355–363.
[18] Kobayashi, K., et al. Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4678–4683.
[19] Galperin, M.Y.; Koonin, E.V. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res*. **2004**, *32*, 5452–5463.

[20] Huynen, M.; Diaz–Lazcoz, Y.; Bork, P. Differential genome display. *Trends Genet*. **1997**, *13*, 389–390.
[21] Huynen, M.; Dandekar, T.; Bork, P. 1998. Differential genome analysis applied to the species–specific features of *Helicobacter pylori*. *FEBS Lett*. **1998**, *426*, 1–5.
[22] Bruccoleri, R.E.; Dougherty, T.J.; Davison, D.B. 1998. Concordance analysis of microbial genomes. *Nucleic Acids Res*. **1998**, *26*, 4482–4486.
[23] Mushegian, A.R.; Koonin, E.V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **1996**, *93*, 10268–10273.
[24] Zhang, R.; Ou, H.Y.; Zhang, C.T. DEG: a database of essential genes. *Nucleic Acids Res*. **2004**, *32*, D271–D272.
[25] Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Brinkman, F.S.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrock–Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.; Smith, K.; Spencer, D.; Wong, G.K.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.; Lory, S.; Olson, M.V. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **2000**, *406*, 959–964.
[26] Mdluli, K.E.; Witte, P.R.; Kline, T.; Barb, A.W.; Erwin, A.L.; Mansfield, B.E.; McClerren, A.L.; Pirrung, M.C.; Tumey, L.N.; Warrener, P.; Raetz, C.R.H.; Stover, C.K. 2006. Molecular validation of LpxC as an antibacterial drug target in *Pseudomonas aeruginosa*. *Antimicrob*. *Agents Chemother*. **2006**, *50*, 2178–2184.
[27] Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
[28] Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351.
[29] Dutta, A.; Singh, S. K.; Ghosh, P.; Mukherjee, R.; Mitter, S.; Bandyopadhyay, D. *In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *In Silico Biol*. **2006**, *6*, 005.
[30] Jacobs, M.A.; Alwood, A.; Thaipisuttikul, I.; Spencer, D.; Haugen, E.; Ernst, S.; Will, O.; Kaul, R.; Raymond, C.; Levy, R.; Chun–Rong, L.; Guenthner, D.; Bovee, D.; Olson, M.V.; Manoil, C. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **2003**, *100*, 14339–14344.
[31] Todd, W.; Beaman, Kurt; W., Vogel; Dale, G.; Drueckhammer; John, S.; Blanchard; Steven, L.; Roderick. Acyl group specificity at the active site of tetrahydridipicolinate N–succinyltransferase, *Protein Sci.* **2002**, *11(4)*, 974–979.
[32] Hutton, C.A., Southwood, T.J., Turner, J.J. Inhibitors of lysine biosynthesis as antibacterial agents. *Mini Rev Med Chem*. **2003**, *3(2)*, 115–27.
[33] Dmitry, A.; Rodionov; Alexey, G.; Vitreschak; Andrey, A.; Mironov; Mikhail, S.; Gelfand. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acid Res*. **2003**, *31(23)*, 6748–6757.
[34] Johnson, T.L.; Abendroth, J.; Hol, W.G.; Sandkvist, M. Type II secretion: from structure to function. *FEMS Microbial Lett*. **2006**, *255*, 175–186.
[35] Bally, M.; Filloux, A.; Akrim, M.; Ball, G.; Lazdunski, A.; Tommassen, J. Protein secretion in Pseudomonas aeruginosa: characterization of seven xcp genes and processing of secretory apparatus components by prepilin peptidase. *Mol Microbiol*. **1992**, *6(9)*, 1121–1131.
[36] Voulhoux, R.; Ball, G.; Ize, B.; Vasil, M.L.; Lazdunski, A.; Wu, L.F.; Filloux, A. Involvement of the twin–arginine translocation system in protein secretion via the type II pathway. *EMBO J*. **2001**, *20*, 6735–6741.
[37] d'Enfert, C.; Pugsley, A.P. *Klebsiella pneumoniae pulS* gene encodes an outer membrane lipoprotein required for pullulanase secretion. *J Bacteriol*. **1989**, *171*, 3673–3679.
[38] Filloux, A.; Michel, G.; Bally, M.. GSP–dependent protein secretion in gram–negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. *FEMS Microbiol*. *Rev*. **1998**, *22*, 177–198.
[39] Pugsley, A.P. 1993. The complete general secretory pathway in gram–negative bacteria. *Microbiol*. *Rev*. **1993**, *57*, 50–108.
[40] Andro, T.; Chambost, J.P.; Kotoujansky, A.; Cattaneo, J.; Bertheau, Y.; Barras, F.; van Gijsegem, F.; Coleno, A. Mutants of *Erwinia chrysanthemi* defective in secretion of pectinase and cellulase. *J. Bacteriol*. **1984**, *160*, 1199–1203.
[41] Palomaki, T. *Identification of a secretion signal for the type II protein secretion pathway in Erwinia carotovora.* Dissertations Biocentri Viikki Universitatis Helsingiensis, 6/2003.
[42] Vincent, T.; Lee; Roger, S.; Burkhard, S.; Tümmler; Lory, S. Activities of *Pseudomonas aeruginosa* Effectors Secreted by the Type III Secretion System In Vitro and during Infection. *Infection and Immunity* **2005**, 695–1705.
[43] Jeffrey, B.; Lyczaka, B.; Carolyn, L.; Cannonb, C.; Gerald, B.; Piera. 2000. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes and Infection* **2000**, *2*, 1051−1060.
[44] Jones, C.H.; Hruby, D.E.1998. New targets for antibiotic development: biogenesis of surface adherence structures. *Drug Discovery Today* **1998**, *3(11)*, 495–504.
[45] Toby, A.; Jenkins, A. Surface plasmon resonance shows that type IV pili are important in surface attachment by Pseudomonas aeruginosa, Angus Buckling, Marsha McGhee and Richard H. ffrench–Constant. *J. R. Soc*.

Interface **2005**, *2*, 255–259.
[46] Thanassi, J.A.; Hartman–Neumann, S.L.; Dougherty, T.J.; Dougherty, B.A.; Pucci, M.J. Identification of 113 conserved essential genes using a high–throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acid Res*. **2002**, *30*, 3152–3162.
[47] Altschul, S.F.; Thomas, L.M.; Alejandro, A.S.; Jinghui, Z.; Zheng, Z.; Webb, M.; David, J.L. Gapped BLAST and PSI BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **1997**, *25(17)*, 3389–3402.
[48] Kanehisa, M.; Goto, S.; Kawashima, S.; Nakaya, A. The KEGG databases at Genome Net. *Nucleic Acids Res*. **2002**, *30(1)*, 42–46.
[49] Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **1997**, *24*, 4876–4882.
[50] Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **1993**, *26*, 283–291.
[51] Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol*. **1963**, *7*, 95–99.
[52] Eisenberg, D.; Luthy, R.; Bowie, J.U. VERIFY3D: assessment of protein models with three–dimensional profiles. *Methods Enzymol* **1997**, *277*, 396–404.

Biographies

Pradeep K. Naik is assistant professor of bioinformatics at the Jaypee University of Information Technology. After obtaining a Ph.D. degree in Biotechnology from the Sambalpur University, Dr. Naik undertook advanced P.G. Diploma in Bioinformatics at the Jawaharlal Nehru University. More recently, Dr. Naik has undertaken many research projects in bioinformatics and biotechnology sponsored by DST, DRDO, DBT and ICMR, India.

Seneha Santoshi is a B.Tech Bioinformatics student at Jaypee University of Information Technology.

Ashima Birmani is a B.Tech Bioinformatics student at Jaypee University of Information Technology.