COMPUTATIONAL IDENTIFICATION OF UNIQUE THERAPEUTIC DRUG TARGETS IN BACTERIAL PATHOGENS AND DESIGNING LEAD MOLECULES .

By

ASHIMA BIRMANI(051505)
SENEHA SANTOSHI (051522)

MAY-2009

Submitted in partial fulfillment of the Degree of Bachelors of Technology in Bioinformatics.

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY-WAKNAGHAT

ACKNOWLEDGMENT

We take this opportunity to extend our gratitude towards our project coordinator Dr Pradeep Kumar Naik who has stood by us during this entire semester. We appreciate his efforts and his precious time invested in this work. This work has been one of the most important projects in our academic career by far and a lot of exertion has been put into this work. We are also thankful to our Head of Department, Dr. R.S. Chauhan for his constant encouragement and suggestions for motivating us to put in our best efforts.

We are also very grateful to Mr. Sree Krishna, Research Scholar for extending his help whenever needed. We would also like to extend our gratitude to Ms. Somlata, our worthy lab-assistant for providing us the support whenever needed.

Ashima Birmani
051505

Seneha Santoshi
051522

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

BLAST:	Basic Local Alignment Search Tool
DEG:	Database of Essential Genes
DNA:	DeoxyriboNucleic Acid
FASTA:	FAST-All
HGP:	Human Genome Project
HSA:	KEGG ID for Homo sapiens
KEGG:	Kyoto Encyclopedia of Genes and Genomes
LWP:	Library for WWW in PERL
PAP:	KEGG ID for *Pseudomonas aeruginosa* PA7
PSB:	KEGG ID for *Pseudomonas syringea* B728a
PPU:	KEGG ID for *Pseudomonas putida* KT2440
NCBI:	National Center for Biotechnology Information
PERL:	Practical Extraction and Report Language
RNA:	RiboNucleic Acid
SQL:	Structured Query Language

ABSTRACT

Pseudomonas aeruginosa is an ubiquitous human pathogen capable of infecting virtually all tissues while *Pseudomonas syringe* is a common cause of halo blight in Phaseolus Vulgaris. A large variety of virulence factors contribute to their importance their pathogenesis. However, the complete genome sequence of pathogens has provided a plethora of potential drug targets. While this data potentially contains all the determinants of host-pathogen interactions and possible drug targets, computational methods for selecting suitable candidates for further experimental analysis are currently limited. In general, a target should provide adequate selectivity; yielding a drug which is specific or highly selective against the pathogens with respect to their host. Integration of the knowledge generated by subtractive genomics and comparative metabolomics in comparing bacterial genomes with host and symbiotic organism's genome (if any) is predicted to have a tremendous impact on identification of potent therapeutic targets and rational drug discovery against infectious diseases. Moreover, the entire approach is 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. We have predicted 6 and 22 unique essential proteins in *Pseudomonas aeruginosa* using comparative genomics and metabolomics and for *Pseudomonas syringe* we have 8 targets after the entire approach. After critical evaluations of the targets we have finally considered 3 targets for *Pseudomonas aeruginosa* 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (dapD), general secretion pathway protein L (gspL) and type IV pilus assembly protein PilA as the potent targets. 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase is the key enzyme essential for the synthesis of Lysine, general secretion pathway protein L is involved in type II secretion pathway which is the main terminal branch of the general secretory pathway and type IV pilus assembly protein PilA which is the major pilin of type IV pilus involved in several bacteria, which are known to function in attachment and biofilm formation known virulence factors in many known bacteria. For *Pseudomonas syringe* we have taken two targets UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE) and UDP-N-acetylmuramoyl-L-alanyl-D-

glutamate synthetase (murD) both known to be indespensible for the peptidoglycan biosynthesis leading to formation of cell wall. Inhibiting the function of these enzymes could control the infections caused by these pathogens. We modeled the three dimensional structure of the ,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (dapD), general secretion pathway protein L (gspL) and type IV pilus assembly protein PilA proteins, UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE) and UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase (murD).We used templates with identity 56% , 60% and 43% for2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (dapD), general secretion pathway protein L (gspL) and type IV pilus assembly protein PilA respectively and 62% and 46% for UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE) and UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase (murD) respectively to employ homology-modeling approach. Stereo chemical quality of protein structures were validated by protein structure validation program PROCHECK and VERIFY3D. 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 were used for screening process. Along with the high scoring results, the interaction studies provided promising ligands for future experimental screening to inhibit the function of 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (dapD), general secretion pathway protein L (gspL) and type IV pilus assembly protein PilA in *Pseudomonas aeruginosa* and UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE) and UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase (murD) in *Pseudomonas syringe*

CHAPTER 1: INTRODUCTION

Availability of genome sequences of pathogens has provided a tremendous amount of information that can be useful in drug target identification. One of the recently adopted strategies is based on a subtractive genomics approach, in which the subtraction dataset between the host and pathogen genome provides information for a set of genes that are likely to be essential to the pathogen but absent in the host. Further using comparative metabolomics approach results are verified, in which the subtraction dataset between the host and pathogen metabolome provides information for a set of enzymes that are likely to be essential to the pathogen but absent in the host and can serve as therapeutic targets.

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 defence against antibiotic sensitive 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. Moreover, the target should be essential for growth and viability of the pathogen at least under the condition of infection.

The search for potential drug targets has increasingly relied on genomic and metabolomics approaches. The entire approach is 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. At the same time, this target should not have any well-conserved homolog in the host and symbiotic organisms (if any). This would preclude possibilities of unacceptable cross-reactivity that might prove detrimental to the host. The above approach to target identification is essentially subtractive because we use a subtraction dataset while comparing the two genomes and metabolomes under consideration. The focus is on the complement of the dataset of the pathogen that is essential for it but is not present in host.

1.1 WHAT ARE POTENT DRUG TARGETS?

Typically a drug target is a key molecule involved in a particular metabolic or signaling pathway that is specific to a disease condition or pathology, or to the infectivity or survival of a microbial pathogen. Some approaches attempt to inhibit the functioning of the pathway in the diseased state by causing a key molecule to stop functioning. Drugs may be designed that bind to the active region and inhibit this key molecule. Drug targets are membrane or cellular receptors or other molecules that are pivotally involved in disease processes. From a pharmacological viewpoint, a drug target is either inhibited or activated by drug molecules (e.g. small organic molecules, antibodies, therapeutic proteins). Drug molecules can physically attach to a drug target, triggering a cascade of intracellular biochemical reactions, followed by a cellular reaction. Potential drug targets can include genes that are differentially expressed between individuals who are and are not in need of individual is exposed to a drug known to alleviate or exacerbate the symptoms of interest, and genes that are co-expressed with other genes presumed to be involved in the systems and pathways under study. Any gene falling into one of these categories may be a gene for which manipulation of its expression might affect disease or symptom progression (Allison 2002). In summary, good drug targets are potent and specific, that is, they must have strong effects on a specific biological pathway and minimal effects on all other pathways.

1.2 ORGANISM UNDER STUDY:

1.2.1 *Pseudomonas Aeruginosa PA7:*

Pseudomonas aeruginosa, a ubiquitous gram-negative bacterium, has been intensively studied as an opportunistic human pathogen (Britigan et al., 1997) and as the dominant pathogen infecting the lungs of cystic fibrosis patients (Pier et al., 1996). *Pseudomonas aeruginosa* is member of the Gamma Proteobacteria class of Bacteria. It is a Gram-negative, aerobic rod belonging to the bacterial family Pseudomonadaceae. Since the revisionist taxonomy based on conserved macromolecules (e.g. 16S ribosomal RNA) the family includes only members of the genus *Pseudomonas* which are cleaved into eight groups. *Pseudomonas aeruginosa* is the type species of its group which contains 12 other members.

Like other members of the genus, *Pseudomonas aeruginosa* is a free-living bacterium, commonly found in soil and water. However, it occurs regularly on the surfaces of plants and occasionally on the surfaces of animals, *Pseudomonas aeruginosa* has become increasingly recognized as an emerging opportunistic pathogen of clinical relevance. Several different epidemiological studies track its occurrence as a nosocomial pathogen and indicate that antibiotic resistance is increasing in clinical isolates. In fact, *Pseudomonas aeruginosa* is the epitome of an opportunistic pathogen of humans. The bacterium almost never infects uncompromised tissues, yet there is hardly any tissue that it cannot infect if the tissue defenses are compromised in some manner. It causes urinary tract infections, respiratory system infections, dermatitis, soft tissue infections, bacteremia, bone and joint infections, gastrointestinal infections and a variety of systemic infections, particularly in patients with severe burns and in cancer and AIDS patients who are immunosuppressed. *Pseudomonas aeruginosa* infection is a serious problem in patients hospitalized with cancer, cystic fibrosis, and burns. Pseudomonas aeruginosa can cause a variety of skin infections, both localized and diffuse. The common predisposing factors are breakdown of the integument which may result from burns, trauma or dermatitis; high moisture conditions such as those found in the ear of swimmers and the toe webs of athletes, hikers and combat troops, in the perineal region and under diapers of infants, and on the skin of whirlpool and hot tub users. The case fatality rate in these patients is near 50 percent. According to the CDC, the overall incidence of *P. aeruginosa* infections in U.S. hospitals averages about 0.4 percent (4 per 1000 discharges), and the bacterium is the fourth most commonly-isolated nosocomial pathogen accounting for 10.1 percent of all hospital-acquired infections.(*Pseudomonas aeruginosa:* Opportunistic Infections,© 2008).

Pseudomonas aeruginosa is notorious for its resistance to antibiotics and is, therefore, a dangerous and dreaded pathogen. The bacterium is naturally resistant to many antibiotics due to the permeabiliity barrier afforded by its Gram-negative outer membrane. Also, its tendency to colonize surfaces in a biofilm form makes the cells impervious to therapeutic concentrations antibiotics. Since its natural habitat is the soil, living in association with the bacilli, actinomycetes and molds, it has developed resistance to a variety of their naturally-occuring antibiotics. Moreover, *Pseudomonas* maintains antibiotic resistance plasmids, both

R-factors and RTFs, and it is able to transfer these genes by means of the bacterial mechanisms of horizontal gene transfer (HGT), mainly transduction and conjugation.Only a few antibiotics are effective against *Pseudomonas aeruginosa*, including fluoroquinolones, gentamicin and imipenem, and even these antibiotics are not effective against all strains. The futility of treating *Pseudomonas* infections with antibiotics is most dramatically illustrated in cystic fibrosis patients, virtually all of whom eventually become infected with a strain that is so resistant that it cannot be treated.

For an opportunistic pathogen such as *Pseudomonas aeruginosa*, the disease process begins with some alteration or circumvention of normal host defenses. The pathogenesis of *Pseudomonass* infections is multifactorial, as suggested by the number and wide array of virulence determinants possessed by the bacterium. Multiple and diverse determinants of virulence are expected in the wide range of diseases caused, which include septicemia, urinary tract infections, pneumonia, chronic lung infections, endocarditis, dermatitis, and osteochondritis. Most *Pseudomonas* infections are both invasive and toxinogenic. The ultimate Pseudomonas infection may be seen as composed of three distinct stages: (1) bacterial attachment and colonization; (2) local invasion; (3) disseminated systemic disease. However, the disease process may stop at any stage. Particular bacterial determinants of virulence mediate each of these stages and are ultimately responsible for the characteristic syndromes that accompany the disease.

***Pseudomonas syringe* B728a:**

Pseudomonas syringae is a rod shaped, Gram-negative bacterium with polar flagella. It is a member of the *Pseudomonas* genus, and based on 16S rRNA analysis, P. syringae has been placed in the *P. syringae* group (Beattie, G.A., and S.E. Lindow. 1994). It is a plant pathogen which can infect a wide range of plant species, and exists as over 50 different pathovars, *P. syringae* pv. syringae strain B728a is the cause of bacterial brown spot of bean (Beattie, G.A., and S.E. Lindow. 1994)It is a highly fit colonist of bean with predominant leaf surface localization (Beattie, G.A., and S.E. Lindow. 1994.,Wilson, M., S.S. Hirano, and S.E. Lindow. 1999.)It is also active in ice nucleation at warm subfreezing temperatures (Lindow, S.E. 1993, Gurian-Sherman, D., and S.E. Lindow. 1993) It is known to be acause of: Bacterial

brown spot. Brown spot is a recurring problem that can cause serious losses in snap beans. The initial foliar symptom of bacterial brown spot is small water-soaked spots that develop into distinctive necrotic brown spots about 3-8mm in diameter, often with a narrow, diffuse yellow margin. Sunken brown spots can form on the pods. If infection occurs early in pod development, the pod may become bent or twisted at the infection site. Halo blight has caused serious losses in both dry and snap beans. Symptoms of halo blight initially appear as small watersoaked spots on the underside of the leaflets, eventually developing into numerous small, reddish-brown lesions on the leaves.

***Pseudomonas putida*(used as a non pathogenic reference for comparison)**

Pseudomonas putida is a rod-shaped, Gram-negative, bacteria officially discovered in the mid 1900's. It is a very common bug that lives in soil and freshwater environments all over the world where it moves about by way of one or more flagella close to the surface.It is the first patented organism in the world.It plays a very important role in the decomposition that drives the carbon cycle. This has caused microbiologists and environmentalists all around the globe to take notice. *P. putida* has been shown to have the ability to break down many organic toxins including Atrazine, the worlds most widely used herbicide, into carbon dioxide and water. Atrazine is toxic to wildlife and is also thought to be a carcinogen. Until recently the only way to rid the environment of Atrazine laced earth was to relocate the poisoned dirt to a landfill. Now scientists are developing a system to inoculate the contaminated ground with high concentrations of *P. putida* in order to rid it of the toxic chemical. *P. putida* is so effective that it lessens the half-life of Atrazine (8 years) to just 5 ½ hours.

The diverse metabolism of *P. putida* may be exploited for bioremediation; for example, it is used as a soil inoculant to remedy naphthalene contaminated soil (Newton C.M. Gomes 2005). *P. putida* is capable of converting styrene oil into the biodegradable plastic PHA(Ward PG, Goff M, Donner M, Kaminsky W, O'Connor KE. 2006). This may be of use in the effective recycling of Polystyrene foam, otherwise thought to be non-biodegradable. It may be useful for the decomposition of *m*-toluate in industrial waste containing both 3-chlorobenzoate and *m*-toluate.(XinCai Chen, 2006)

P. putida has demonstrated potential biocontrol properties, as an effective antagonist of damping off diseases such as *Pythium*(Amer GA, Utkhede RS. 2000) and *Fusarium*(Validov S, Kamilova F , 2007).

P. putida demonstrates high capacities for biosorption of Cu and Zn. Since *P. putida* CZ1 could grow in the presence of significant concentrations of metals and because of its high metal uptake capacity in aerobic conditions, this bacterium may be potentially applicable in bioreactors or in situ bioremediation of heavy-metal-contaminated aqueous or soil.

2.3 Approaches for drug target identification

Traditionally micro arrays have been well utilized in genomics/proteomics approaches for gene/protein expression profiling and tissue/cell-scale target identification and validation. Chemical genomics and proteomics are emerging tools for generating phenotype changes, thus leading to target and hit identifications. NMR-based screening, as well as activity-based protein profiling, are trying to meet the requirement of high-throughput target identification. Besides being used high-throughput experiments, bioinformatics can contribute to the processes of target identification and validation by providing functional information of target candidates and positioning information on the biological networks.

There are many *in silico* approaches for finding drug targets in pathogenic bacteria. Bruccoleri et al. (1998) has developed a simple and computational tool that 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, but the availability of this approach as an automated tool is limited. An automated tool, T-iDT developed by Singh et al. (2006) predict highly conserved genes, which are essential for pathogenic bacteria with no similarities with the host genes as potential drug targets. This and other existing tools use only human genome sequence as a template for comparison against pathogens. However, comparison with the symbiotic organisms living within the human body cannot be ruled out for successful drug development. Fortunately the genome sequences of all these symbiotic bacteria are available and can be used as template for comparison with pathogen bacteria. Although bioinformatics tools and resources can be used to identify putative drug targets, validating these targets is again very essential.

Traditionally it requires an understanding of the role of the gene or protein in the disease process and is heavily dependent on laboratory-based work.

Besides being used high-throughput experiments, bioinformatics can contribute to the processes of target identification and validation by providing functional information of target candidates and positioning information on the biological networks. There are many *in silico* approaches for finding drug targets in pathogenic bacteria. Bruccoleri et al. (1998) has developed a simple computational tool that 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, but the availability of this approach as an automated tool is limited. However, comparison with the symbiotic organisms living within the human body cannot be ruled out for successful drug development. Fortunately the genome sequences of all these symbiotic bacteria are available and can be used as template for comparison with pathogen bacteria. Although bioinformatics tools and resources can be used to identify putative drug targets, validating these targets is again very essential. Traditionally it requires an understanding of the role of the gene or protein in the disease process and is heavily dependent on laboratory-based work.

Genomics and proteomics technologies have created a paradigm shift in the drug discovery process, with bioinformatics having a key role in the exploitation of genomic, transcriptomic, and proteomic data to gain insights into the molecular mechanisms that underlie disease and to identify potential drug targets. 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/) and the comparative and subtractive genomic 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.

Drug target identification involves acquiring a molecular level understanding of a specific disease state and includes analysis of gene sequences, protein structures, protein interactions and metabolic pathways the ultimate goal of the process is to discover a suitable

target whose biological activity can be directly linked to a pathological process. In the age of genomics, discovery of novel drug targets needs to incorporate and integrate different sources of data including gene expression data, gene sequence data, and gene polymorphism data and so on. Many public biological databases are warehousing and providing a great amount of functional information for drug discovery. Yet one of the most important information is the annotation of human genome itself and its associated symbiotic organisms. In addition, the publicly available tools are as important as the data and include algorithms for gene prediction, sequence homology searching, prediction of function and so on integrating existing data (essential genes) from public databases and piping the tools for gene prediction, BLAST search as subtractive genomic tool and automated validation tools create systematic analysis which will be helpful for predicting and inferring the therapeutic targets.

OBJECTIVE:

The objectives of our project are-

- Searching for potential drug targets in microbial genomes taking into consideration specifically *Pseudomonas* species whose biological activity can be linked to a pathological process using comparative genomics and metabolomics.

- Analyzing the biological data being generated on potential drug targets and providing insights to understand the biological regulatory mechanisms in diseases, which has been playing an increasingly important role in searching for novel drug targets from the information contained in genomes.

- Followed by structure modeling of the predicted therapeutic targets as well as designing of potent lead compounds.

CHAPTER 2:
DRUG TARGET IDENTIFICATION

2.1 COMPARATIVE GENOMICS

The availability of genome-scale sequenced data of microbes and the human genome has revolutionized the field of drug-discovery against threatening human pathogens (Lander et al. 2001; Venter et al. 2001). The strategies for drug design and development are progressively shifting from the genetic approach to the genomic approach (Galperin and Koonin 1999). Comparative genomics is the study of the relationship of genome structure and function across different biological species or strains. Comparative genomics is an attempt to take advantage of the information provided by the signatures of selection to understand the function and evolutionary processes that act on genomes. The sheer amount of information contained in modern genomes (750 megabytes in the case of humans) necessitates that the methods of comparative genomics are automated. Gene finding is an important application of comparative genomics, as is discovery of new, non-coding functional elements of the genome.

Comparative genomics exploits both similarities and differences in the proteins, RNA, and regulatory regions of different organisms to infer how selection has acted upon these elements. Those elements that are responsible for similarities between different species should be conserved through time (stabilizing selection), while those elements responsible for differences among species should be divergent (positive selection). Finally, those elements that are unimportant to the evolutionary success of the organism will be unconserved (selection is neutral).

Having come a long way from its initial use of finding functional proteins, comparative genomics is now concentrating on finding regulatory regions and siRNA molecules. Recently, it has been discovered that distantly related species often share long conserved stretches of DNA that do not appear to code for any protein. It is unknown at this time what function such ultra-conserved regions serve. Using computer-based analysis to zero in on the genomic features that have been preserved in multiple organisms over millions of years, researchers will be able to pinpoint the signals that control gene function, which in turn

should translate into innovative approaches for treating human disease and improving human health. In addition to its implications for human health and well-being, comparative genomics may benefit the animal world as well. As sequencing technology grows easier and less expensive, it will likely find wide applications in agriculture, biotechnology and zoology as a tool to tease apart the often-subtle differences among animal species. Such efforts might also possibly lead to the rearrangement of our understanding of some branches on the evolutionary tree, as well as point to new strategies for conserving rare and endangered species. The work has been effectively complemented with the compilation of the Database of Essential Genes (DEG) for a number of pathogenic microorganims (Zhang et al. 2004).

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) (Ref: http://tubic.tju.edu.cn/deg/) and the comparative genomic 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.

One of the applications is the prediction of essential genes based on homologous sequence search against DEG. The functions encoded by essential genes are considered to be generally essential for all cells. It is even believed that some basic functions and principles are common to all cellular life on this planet. Therefore, if the query sequences compared using BLAST have homologous genes in DEG, it is likely that the queried genes are also essential. In addition, by performing the BLAST search against DEG for all the protein-coding genes in a genome, it is possible to define the putative essential genes for the proteomes of newly sequenced genomes. However, caution must be taken in interpreting the BLAST results, since many essential genes are essential only in given growth conditions, such as in rich or minimal medium.

2.1.1 Essential genes

Essential genes are genes that are indispensable to support cellular life. These genes constitute a minimal gene set required for a living cell. Therefore, the functions encoded by

this gene set are essential and could be considered as a foundation of life itself (Kobayashi 2003; Itaya 1995). The definition of the minimal gene set needed to sustain a living cell is of considerable interest not only because it represents a fundamental question in biology, but also because it has much significance in practical use. For example, since most antibiotics target essential cellular processes, essential gene products of microbial cells are promising new targets for antibacterial drugs (Judson and Mekalanos 2000).

2.1.2 COMPARATIVE METABOLOMICS

Metabolomics is a relatively new member to the '-omics' family of systems biology technologies (Bino 2004). The term 'metabolome' was coined in 1998 and was used to describe the metabolite complement of living tissues (Oliver et al. 1998), metabolomics as a field of study is now firmly established as a functional genetics approach to understanding the molecular complexity of life (Wagner et al. 2003.). Metabolomics is the comprehensive, qualitative, and quantitative study of all the small molecules in an organism (Oliver et al.1998)

Since the structure and dynamics of metabolic pathways in pathogens can suggest vital processes that may serve as targets for inactivation, one application area of this tool is in drug design.Computational pathway analysis also facilitates the identification of enzymes that participate in several pathways. The disruption of such enzymes may be particularly harmful to the pathogen. Comparative metabolomics approach is the comparison of pathways present in host and in microbial pathogen that may 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 .

We have worked on how the analysis of metabolic pathways can provide support for the determination of gene functions, for comparisons of cellular networks and for the design of novel antimicrobial drugs. The networks of interconnected biochemical reactions in a cell are graphically represented as metabolic pathway maps depicting the relationships between enzymes and the chemical compounds they transform. Our objective for using comparative metabolomics is the comparison of the pathways present in the host cell and in a microbial pathogen that may reveal enzymes that are unique for the pathogen and thus suitable as drug targets. Computational pathway analysis also facilitates the identification of enzymes that participate in several pathways. The

disruption of such enzymes may be particularly harmful to the pathogen. Metabolic databases also enable identification of enzymes that have a "back-up" of similar enzymes that are able to perform the same function if the original enzyme is disrupted. Such redundant enzymes are probably less suitable as drug targets.

The complete genome sequence of *Pseudomonas* provides an opportunity for a more focused and planned approach towards the identification of new drug targets.In this study, we have adopted a strategy of comparative metabolic pathway analysis. The enzymes in the pathways of *Pseudomonas*, which do not show similarity to any protein from the host, represent attractive potential drug targets. The elimination of pseudo drug targets is essential since the cost involved in the investigation of drug targets is prohibitive. 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.

2.2 Materials and Methods

2.2.1 *Database of Essential genes*

The database of essential genes of bacteria (DEG version 5.0) was downloaded from http://tubic.tju.edu.cn/deg/ using a program, LWP module in PERL and manually compiled to use as a stand-alone database for the BLAST program. It consists of a collection of 9948 essential genes from 14 completely annotated bacterial genomes(*Acinetobacter baylyi* (ADP1):499; *Bacillus subtilis* 168:271; *Escherichia coli* (MG1655):615; *Francisella novicida* (U112):392; *Haemophilus influenzae* Rd:642; *Helicobacter pylori (*26695):323; *Helicobacter pylori*(J99):312; *Mycobacterium tuberculosis* (H37Rv):614; *Mycoplasma genitalium* (G37):381; *Mycoplasma pulmonis* (UAB CTIP):310; *Salmonella typhimurium* (LT2):230; *Staphylococcus aureus* (N315):302; *Streptococcus pneumoniae:244; Vibrio cholerae*:5)and 8 eukaryotes(Arabidopsis thaliana:777; Caenorhabditis elegans:294; Danio rerio:288; Drosophila melanogaster:339; Homo sapiens:118; Mus musculus:2114; Saccharomyces cerevisiae:878. The program used for downloading the sequences from the DEG is specified in appendix. Each entry of essential genes has a unique DEG identification number, gene reference number, gene function and sequence. All information is stored and operated by using an open-source

database management system, MySQL. Users can browse and extract all the records of these entries. In addition, users can also search DEG by gene function or name. Furthermore, there is a BLAST program locally. Therefore, users can BLAST the query sequences against all the essential gene sequences in DEG.

One of the applications is the prediction of essential genes based on homologous sequence search against DEG. The functions encoded by essential genes are considered to be generally essential for all cells (2). It is even believed that some basic functions and principles are common to all cellular life on this planet (5). Therefore, if the query sequences compared using BLAST have homologous genes in DEG, it is likely that the queried genes are also essential. In addition, by performing the BLAST search against DEG for all the protein-coding genes in a genome, it is possible to define the putative essential genes for the proteomes of newly sequenced genomes. However, caution must be taken in interpreting the BLAST results, since many essential genes are essential only in given growth conditions, such as in rich or minimal medium.

Another application is that by analyzing all the essential genes in DEG, some principles or regulations could be found to answer the question of what are the basic functions necessary to support cellular life. Those principles could lead to the development of new algorithms to predict essential genes. Some functions encoded by essential genes are expected, such as DNA replication, gene transcription, protein synthesis, energy production and cell division. Some essential genes, however, are somewhat unexpected, such as Embden–Meyerhof–Parnas pathway genes and a purine biosynthesis gene (1). Analysis of DEG, which has all essential genes among different organisms, could help to classify those 'unexpected' essential genes.

2.2.2 KEGG Database

KEGG is a database of biological systems, consisting of genetic building blocks of genes and proteins (KEGG GENES), chemical building blocks of both endogenous and exogenous substances (KEGG LIGAND), molecular wiring diagrams of interaction and reaction networks (KEGG PATHWAY), and hierarchies and relationships of various biological objects (KEGG BRITE). KEGG provides a reference knowledge base for linking genomes to

biological systems and also to environments by the processes.There are approximately 100 reference metabolic pathway maps in KEGG, each showing a biochemical pathway such as glycolysis or peptidoglycan biosynthesis. Enzymes are represented by rectangles, marked with enzyme commission (EC) numbers (see below), and circles represent metabolites and intermediates. The pathway maps are interlinked through the compounds that are present in several pathways. KEGG also contains organism-specific pathway maps for each sequenced organism. These are generated from the reference maps by coloring the enzymes that are present in the organism's genome. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a bioinformatics resource for understanding higher-order functional meanings and utilities of the cell or the organism from its genome information. It is an integrated resource consisting of three types of database for genomic, chemical and network information, and associated software, which are all developed by the Kanehisa Laboratory (now part of the Bioinformatics Center) in the Institute for Chemical Research, Kyoto University. While KEGG has cross-references to numerous outside databases, it is intended to be a self-sufficient system for linking genomes to life at the cellular level, containing a complete set of building blocks (genes and molecules) and wiring diagrams (interaction networks) for cellular functions (Table 3.1). The key linking the enzymes in the pathway maps to the genes is the EC number. EC numbers constitute a systematic classification scheme for enzymes, maintained by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB).

Here we have made use of PATHWAY database. The protein network in KEGG is an abstract network of gene products, representing not only the pathway or the complex resulting from direct protein–protein interactions, but also the metabolic network viewed as a network of enzymes, and the gene regulatory network viewed as a network of transcription factors and target products. The PATHWAY database is a collection of manually drawn diagrams called the KEGG reference pathway diagrams (maps), each corresponding to a known network of functional significance. The PATHWAY database also contains organism-specific pathways, which are automatically generated by superimposing genes in given organisms. The database currently contains 13 457 entries including 235 reference pathway diagrams.

Table 2.1: Detailed Information about KEGG Database.

Database	Content	Source
PATHWAY	Molecular interaction and reaction networks for metabolism, various cellular processes, and human diseases	Manually entered from published materials
BRITE	Functional hierarchies representing our knowledge on various aspects of biological systems	Manually entered from published materials
GENES	KEGG ORTHOLOGY (KO): Ortholog groups based on PATHWAY and BRITE	Manually defined by KEGG
	GENES: Gene catalogs of complete genomes with manual annotation	Generated from RefSeq and other public resources with reannotation by KEGG
	DGENES: Gene catalogs of draft genomes with automatic annotation	
	EGENES: Gene catalogs (consensus contigs) of EST data with automatic annotation	
	GENOME: Genome maps and organism information	
	SSDB: Sequence similarities with best-hit information for identifying ortholog/paralog clusters and conserved gene clusters	Computationally derived from GENES by pairwise genome comparisons of all protein-coding genes
LIGAND	COMPOUND: Chemical compounds	Manually entered from published materials
	DRUG: Drugs approved in the U.S. and Japan	
	GLYCAN: Glycans	
	REACTION: Chemical reactions	
	RPAIR: Chemical structure transformation patterns	
	ENZYME: Enzyme nomenclature	Generated from ExplorEnz enzyme database with annotation by KEGG

2.2.3 BLAST

The standalone BLAST executables including BLASTn, BLASTp and BLASTx were downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/) and installed locally. In bioinformatics, Basic Local Alignment Search Tool, or BLAST, is an algorithm for

comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. Standard protein-protein BLAST (blastp) is used for both identifying a query amino acid sequence and for finding similar sequences in protein databases. Like other BLAST programs, blastp is designed to find local regions of similarity. When sequence similarity spans the whole sequence, blastp will also report a global alignment, which is the preferred result for protein identification purposes. Blastx compares translational products of the nucleotide query sequence to a protein database. Because blastx translates the query sequence in all six reading frames and provides combined significance statistics for hits to different frames, it is particularly useful when the reading frame of the query sequence is unknown or it contains errors that may lead to frame shifts or other coding errors

NCBI has been used mainly to download the following genome and proteome sequences.

1. *Homo Sapiens* Whole Proteome
2. *Pseudomonas aeruginosa* PA7 Whole Genome
3. *Pseudomonas syringae* B728a Whole Genome
4. *Pseudomonas putida* KT2440 Whole Genome
5. ESTof Phaseolus Vulgaris
6. Proteomes of symbiotic organisms
    1. *Bacteroides thetaiotaomicron*
    2. *Escherichia coli,*
    3. *Lactobacillus acidophilus*
    4. *Lactobacillus johnsonii*

2.2.4 Target identification

2.2.4.1 *Pseudomonas aeruginosa PA7*

➢ **Comparative Genomics:** The protocol used for indentification of therapeutic targets of *P. aeruginosa* based on comparative genomics is represented in figure 2.1.

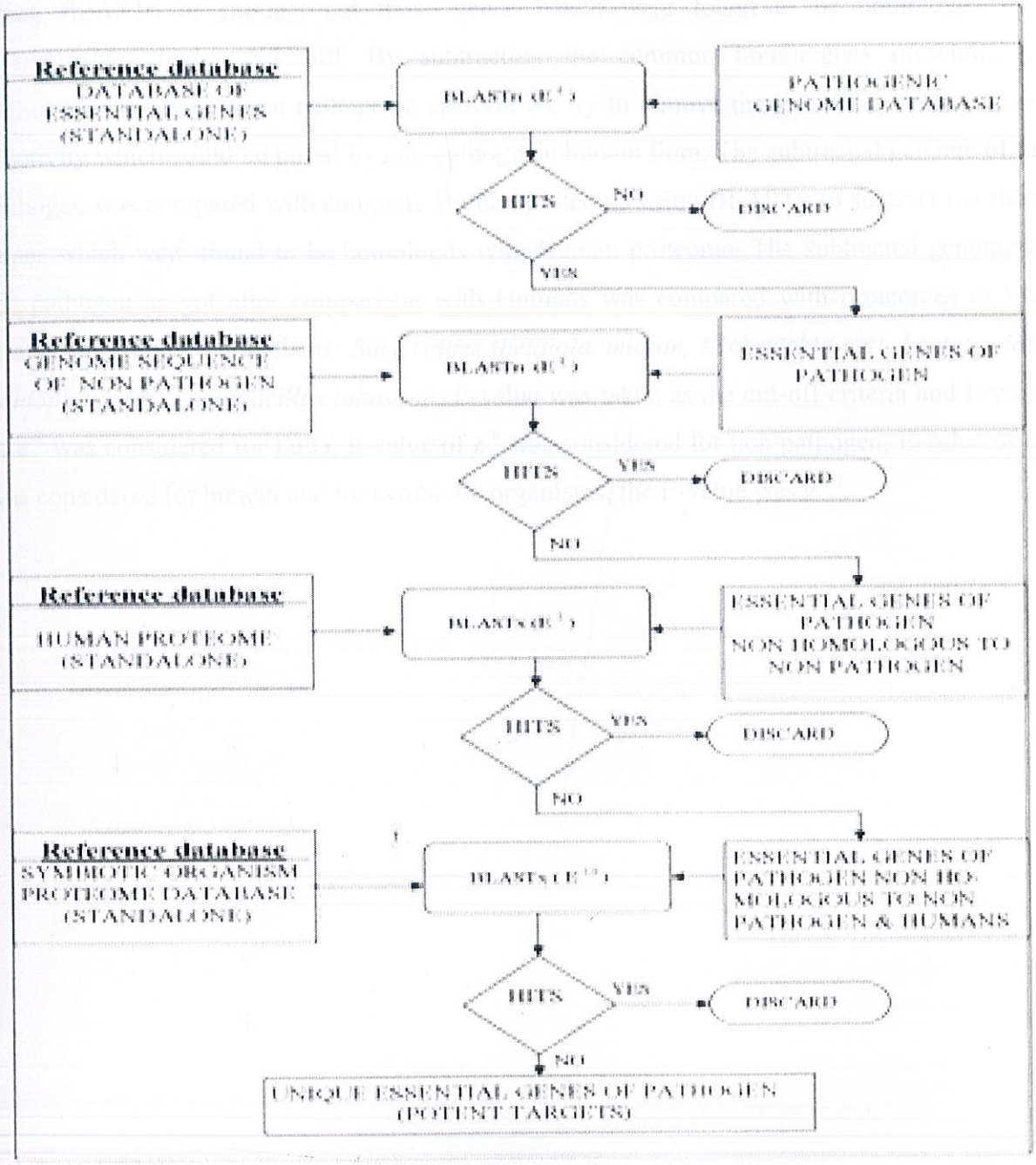


FIG 2.1: Protocol for comparative genomics of *Pseudomonas aeruginosa* PA7

The set of essential genes in *P. aeruginosa* have been predicted by searching with DEG. We have used BLASTn to compare the genome of the pathogen with the data set of 9948 genes from DEG. Those genes which were found non homologus to the genes of the database used

were subtracted out of the pathogen genome. The subtracted genome of the pathogen was compared with complete genome of non pathogenic bacteria (*Pseudomonas putida* KT2440) using BLASTn to subtract out those genes which were found to be homologus with *Pseudomonas putida* KT2440. By subtracting the common homologues presenting in pathogenic bacteria of non-pathogenic bacteria we try to remove the possibility of any cross-reactivity which could be posed by non-pathogenic human flora. The subtracted genome of the pathogen was compared with complete Human proteome using BLASTx to subtract out those genes which were found to be homologus with Human proteome. The subtracted genome of the pathogen as got after comparison with Humans was compared with proteomes of four strains of symbiotic organisms: *Bacteroides thetaiotaomicron, Escherichia coli, Lactobacillus acidophilus* and *Lactobacillus johnsonii*. E-value was taken as the cut-off criteria and E-value of e^{-8} was considered for DEG, E-value of e^{-8} was considered for non pathogen, E-value of e^{-5} was considered for human and for symbiotic organisms, the E-value was e^{-10}.

➢ Comparative metabolomics

The protocol used for indentification of therapeutic targets of *P. aeruginosa* based on comparative metabolomics is represented in figure 2.2.

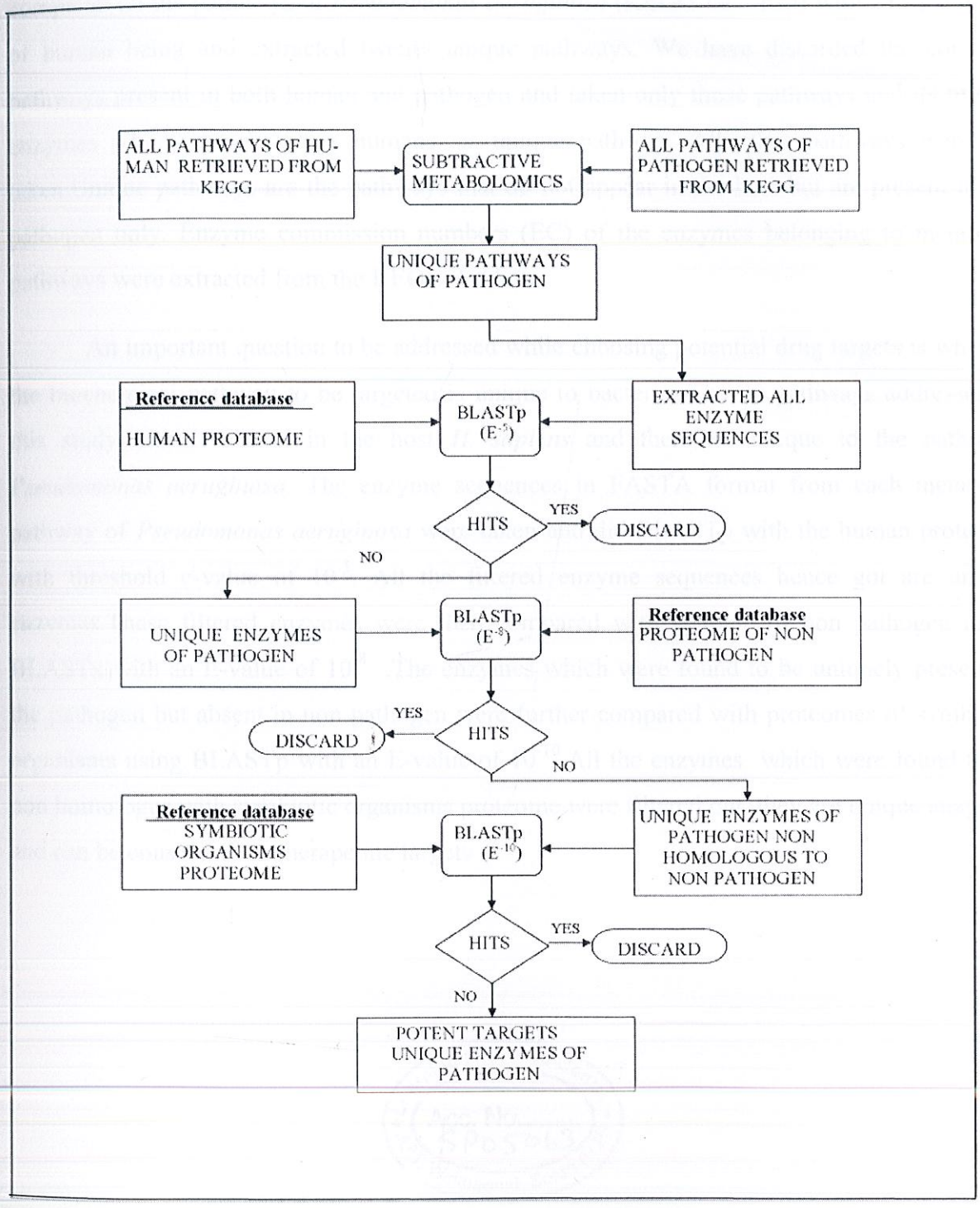


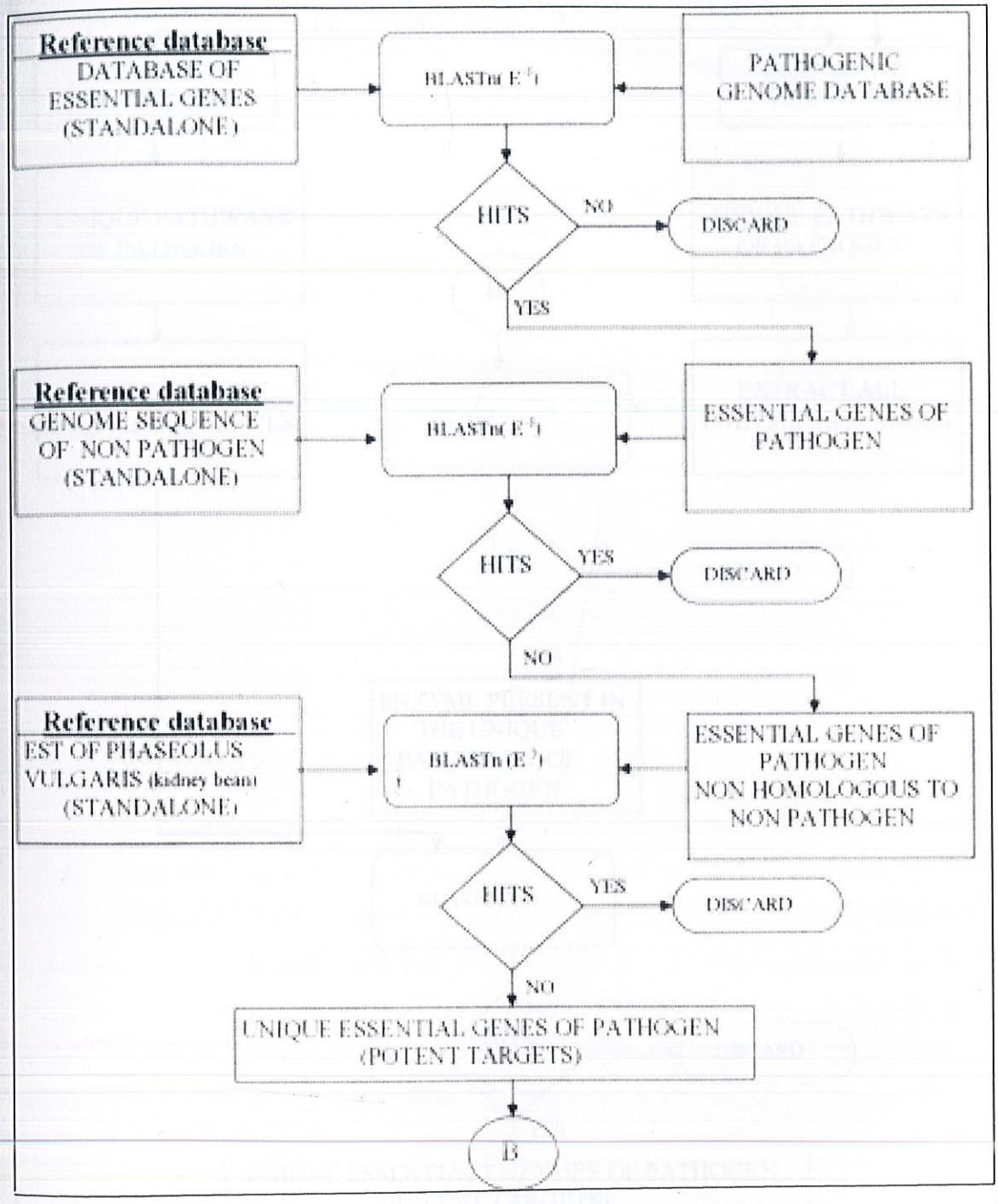
FIG 2.2: Protocol for comparative metabolomics of *Pseudomonas aeruginosa* PA7

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 [Kanehisa et al., 2002]. We have compared all the pathways of *Pseudomonas aeruginosa* (KEGG i.d. –pap) with 213 pathways of human being and extracted twenty unique pathways. We have discarded the common pathways present in both human and pathogen and taken only those pathways and its related enzymes which are absent in humans, as unique pathway. All other pathways were not taken.Unique pathways are the pathways that do not appear in the host but are present in the pathogen only. Enzyme commission numbers (EC) of the enzymes belonging to metabolic pathways were extracted from the KEGG database.

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}. All the filtered enzyme sequences hence got are unique enzymes.These filtered enzymes were then compared with genome of non pathogen using BLASTx 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 homologus with symbiotic organisms proteome were filtered out ,they are unique enzymes and can be considered as therapeutic targets .

2.2.4.2 *Pseudomonas syringae B728a*

- The protocol used for indentification of therapeutic targets of *P. syringea* based on comparative genomics and metabolomics is represented in figure 2.3.



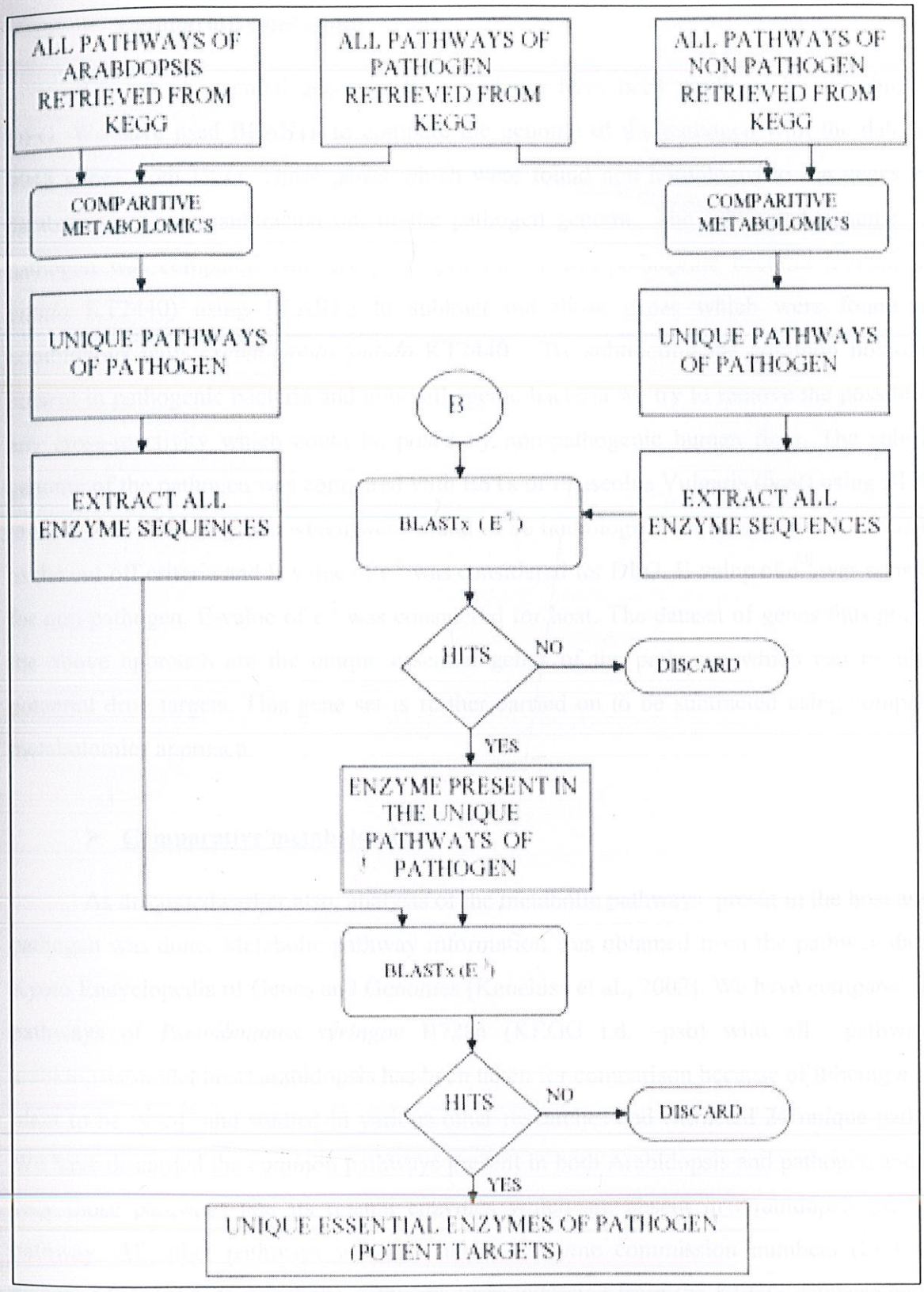


FIG 2.3: Protocol used for *Pseudomonas syringea B728*

➢ Comparative genomics

The set of essential genes in *P. aeruginosa* have been predicted by searching with DEG. We have used BLASTn to compare the genome of the pathogen with the data set of 9948 genes from DEG. Those genes which were found non homologus to the genes of the database used were subtracted out of the pathogen genome. The subtracted genome of the pathogen was compared with complete genome of non pathogenic bacteria (*Pseudomonas putida* KT2440) using BLASTn to subtract out those genes which were found to be homologous with *Pseudomonas putida* KT2440. . By subtracting the common homologues present in pathogenic bacteria and non-pathogenic bacteria we try to remove the possibility of any cross-reactivity which could be posed by non-pathogenic human flora. The subtracted genome of the pathogen was compared with ESTs of Phaseolus Vulgaris (host) using BLASTn to subtract out those genes which were found to be homologus with host . E-value was taken as the cut-off criteria and E-value of e^{-8} was considered for DEG, E-value of e^{-8} was considered for non pathogen, E-value of e^{-3} was considered for host. The dataset of genes thus got from the above approach are the unique essential genes of the pathogen which can be used as potential drug targets. This gene set is further carried on to be subtracted using comparative metabolomics approach.

➢ Comparative metabolomics

As discussed earlier also, analysis of 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 [Kanehisa et al., 2002]. We have compared all the pathways of *Pseudomonas syringae* B728a (KEGG i.d. –psb) with all pathways of arabidopsis(model host),arabidopsis has been taken for comparison because of it being a model plant to be used and studied in various other researches and extracted 24 unique pathways. We have discarded the common pathways present in both Arabidopsis and pathogen and taken only those pathways and its related enzymes which are absent in Arabidopsis ,as unique pathway. All other pathways were not taken. Enzyme commission numbers (EC) of the enzymes belonging to metabolic pathways were extracted from the KEGG database and with the help of those the enzyme sequences in FASTA format from each of the above got

metabolic pathway of *Pseudomonas syringae* B728a were extracted.we hav also compared all the pathways of our pathogen and the non pathogenic bacteria (*Pseudomonas putida*KT2440) and found out pathways unique to our pathogen. Enzyme commission numbers (EC) of the enzymes belonging to metabolic pathways were extracted from the KEGG database and with the help of those the enzyme sequences in FASTA format from each of the above got metabolic pathway of *Pseudomonas syringae* B728a were extracted. Using BLASTx unique enzymes as compare to *Pseudomonas putida* KT2440 were compared with gene dataset got after genomics comparisons. E-value of e^{-5} was taken as cutoff. Those enzymes which matched with genomics result dataset were taken as result dataset for this step. These are the essential unique enzymes present in the unique pathways of the pathogen. These were further compared with the enzyme dataset got after metabolic comparison of Arabidopsis and *Pseudomonas syringae* B728a using BLASTx taking an E-value of e^{-3} as cutoff. The enzymes which matched with the resultant enzymes of metabolic comparison of Arabidopsis and *Pseudomonas syringae* B728a were taken as final resultant dataset. These enzymes are the best potential drug targets essential and uniquely present in the metabolic pathways of the pathogen.

2.3. Results and discussions

Comparative genomics:

Although experimental and computational methods has been previously employed for the identification of probable drug targets by predicting only the essential genes in pathogen bacteria (Dutta et al., 2006; Sakharkar et al., 2004); these works are confined to selective bacterial species. Moreover, they 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. The therapeutic target genes should be essential to the concerned pathogenic bacteria, i.e. any disruption in the functioning of those genes will lead to death of bacteria. All such essential genes can be potential drug targets but including those genes whose products have sequence similarities with any host and its symbiotic bacterial species if any may lead to drug reactions with the host and, thus, lad 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.

Comaparitive metabolomics:

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 (Galperin and Koonin, 1999).

2.3.1 *Pseudomonas aeruginosa:*

The essential genes of pathogenic bacteria were predicted by comparing with DEG using BLASTn .We identified 183 unique essential genes in bacterial species.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).

The predicted essential genes of pathogenic bacteria were compared with the predicted essential genes of non-pathogenic bacteria using BLASTn. It is revealed that out of 183 essential genes only 43 genes are uniquely present in *P.aeruginosa,* which could be related with pathogenicity. The essential genes obtained after first subtraction were then compared with human proteome (BLASTx),only 33 essential genes were uniquely identified from *P.aeruginosa.*

The predicted probable therapeutic targets after comparing with human proteomes were further validated by comparing (BLASTn) with four symbiotic organisms (residing within human being). This results in further reduction in the number of unique therapeutic targets.Finally only 6 unique therapeutic target genes were identified in *P.aeruginosa*

following four level genomic subtractions; first with DEG then with non-pathogenic bacteria, thirdly with human being and finally with symbiotic organisms as shown in table 2.2. Following figure, fig 2.4 shows Subsequent Decrease in number of genes in *P. aeruginosa* after different comaparisons .

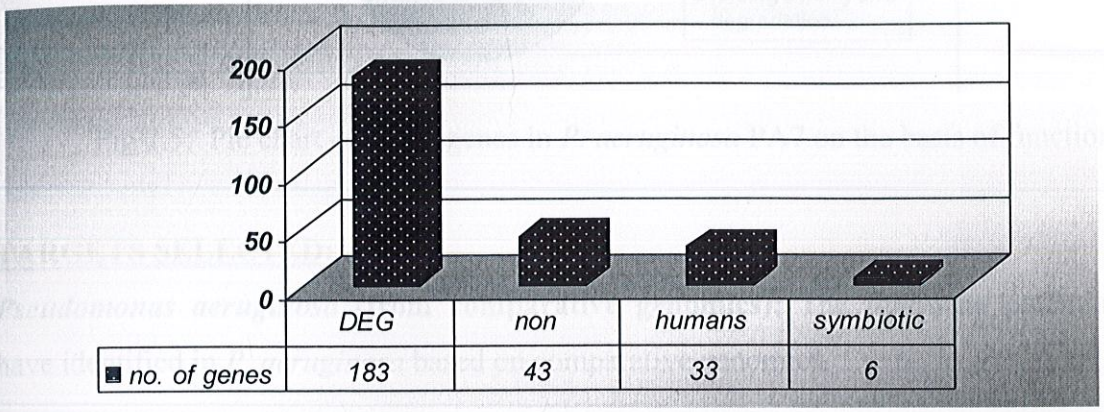


Fig 2.4 : Subsequent Decrease in number of genes in *P. aeruginosa* after different comaparisons.

Table 2.2 : unqiue targets indentified in *Pseudomonas aeruginosa* PA7 after four levels of subtraction.

S. No.	Gene Id	Gene Name	Protein name	pathway
1	PSPA7_1473	dapD	tetrahydrodipicolinate succinylase	Lysine biosynthesis (pap00300)
2	PSPA7_1235	narH	nitrate reductase, beta subunit	Nitrogen metabolism (pap00910) Environmental Information Processing; Signal Transduction (pap02020)
3	PSPA7_1556	metE	5methyltetrahydropteroyl-triglutamate 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)

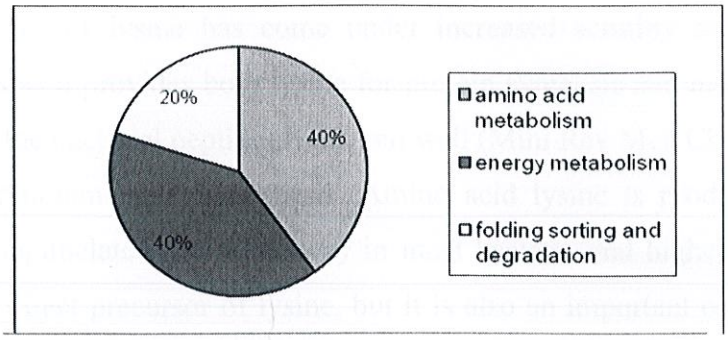


Fig 2.5: Pie chart of target genes in *P. aeruginosa* PA7 on the basis of function

TARGETS SELECTED:

***Pseudomonas aeruginosa* (from comparative genomics):** The following potential targets have identified in *P. aeruginosa* based on comparative genomics.

(a) 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 shown in fig:2.6. 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 (Todd W. Beaman 2002).

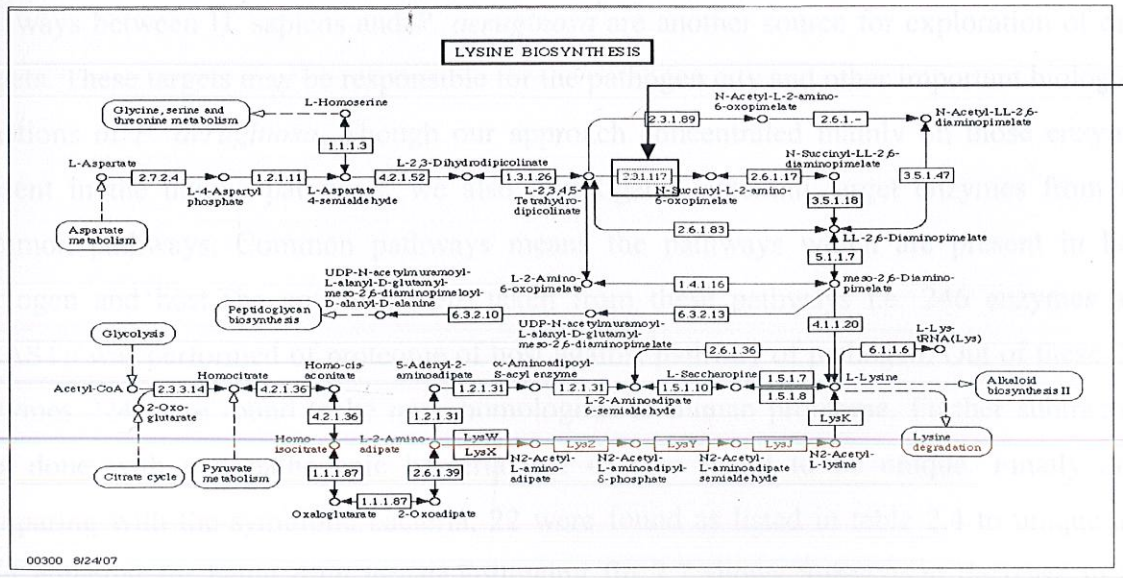


Fig 2.6: Location of the enzyme Tetrahydrodipicolinate succinylase(dapD) in lysine biosynthesis

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 (Mini Rev Med Chem. 2003). It is very important for diaminopimelate biosynthesis. 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 (Dmitry A. Rodionov 2003.). 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.

Cell wall —*Drug*→ Weakened; loss of mechanical support; cell lysis.

Enzymatic reaction of tetrahydrodipicolinate succinylase (Dmitry A. Rodionov 2003**)**

H_2O + tetrahydrodipicolinate + succinyl-CoA <=> *N*-succinyl-2-amino-6-ketopimelate + coenzyme A

COMPARATIVE METABOLOMICS

These enzymes present uniquely in the *P. aeruginosa* (unique metabolic pathways) were given first priority and considered most potent therapeutic targets. A total of 20 unique metabolic pathways were identified as shown in table 2.3. Unique enzymes in common pathways between H. sapiens and *P. aeruginosa* are another source for exploration of drug targets. These targets may be responsible for the pathogen city and other important biological functions of *P. aeruginosa*. Though our approach concentrated mainly on those enzymes present in the unique pathways, we also investigated potential target enzymes from the common pathways. Common pathways means the pathways which are present in both pathogen and host.The enzymes were taken from these pathways i.e. 246 enzymes and BLASTp was performed of proteome of host against pathway of pathogen. Out of these 246 enzymes 224 were found to be non- homologous to human proteome. Further subtraction was done with non pathogenic bacteria and 41were found to be unique. Finally after comparing with the symbiotic bacteria, 22 were found as listed in table 2.4 to unique and hold potential for being drug targets.Following fig 2.7 shows Subsequent decrease in the number proteins in *P. aeruginosa* after different comparison.

Table 2.3: Unique pathways in *P.aeruginosa* when compared with humans.

	UNIQUE PATHWAYS WHEN COMPARED TO HUMAN
	Carbohydrate Metabolism
1	C5-Branched dibasic acid metabolism
	Metabolism of Other Amino Acids
1	D-Alanine metabolism
	Glycan Biosysnthesis and metabolism
1	Lipopolysaccharide biosynthesis
	Biosynthesis of Polyketides and Nonribosomal Peptides
1	Polyketide sugar unit biosynthesis
2	Biosynthesis of siderophore group non ribosomal peptides
	Biosynthesis of Secondary Metabolites
1	Flavonoid Biosynthesis
	Xenobiotics Biodegradation and Metabolism
1	Toluene and xylene degradation
2	2,4-Dichlorobenzoate degradation
3	1,2-Dichloroethane degradation
4	Napthalene and Anthracene Degradation
5	Carbazole Degradation
6	Trinitrotoluene degradation
7	Benzoate degradation via hydroxylation
	Folding, Sorting and Degradation
1	Type II secretion system
2	Type III secretion system
3	Type IV secretion system
	Membrane Transport
1	Phosphotransferase system (PTS)
	Signal Transduction
1	Two-component system
	Cell Motility
1	Bacterial chemotaxis
2	Flagellar assembly

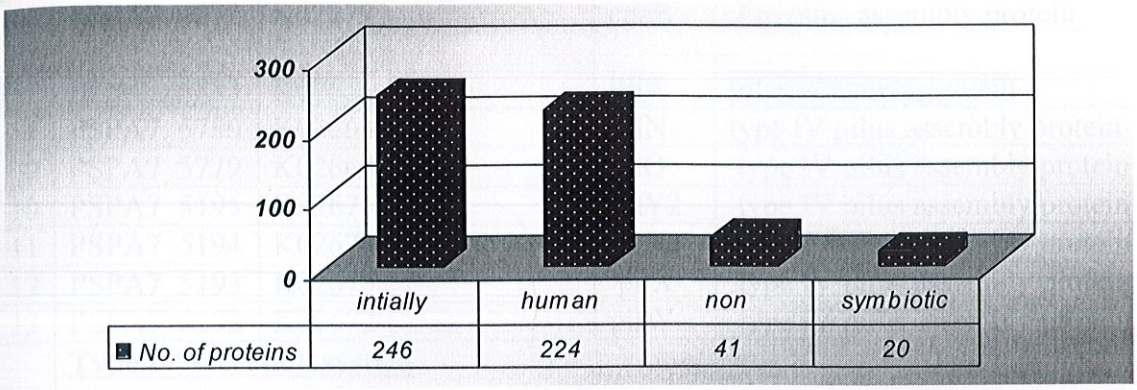


Fig 2.7: Subsequent decrease in the number proteins in *P. aeruginosa* after different comparison.

Table2.4 :22 unique potential targets in *P. aeruginosa* PA7 after metabolomics

	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	K02651		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

Following targets have been selected from *Pseudomonas aeruginosa* base on comparative metabolomics

1. **GspL** :GspL stands for general secretary pathway protein L.

It is involved in T2SS, i.e. type II secretion pathway. Gram-negative type IV pili use a modified version of the type II system for their biogenesis, and in some cases certain proteins are shared between a pilus complex and type II system within a single bacterial species. Type II protein secretion system (T2SS) is widely distributed among most Gram-negative bacteria for secretion of extracellular degradative enzymes and toxins (reviewed by Johnson et al., 2006ased on experimental data with P. aeruginosa T2SS component proteins (Bally et al., 1992; Voulhoux et al., 2001), a type II secretion apparatus model formed by the multi-protein complex that spans the entire cell envelope has been proposed .

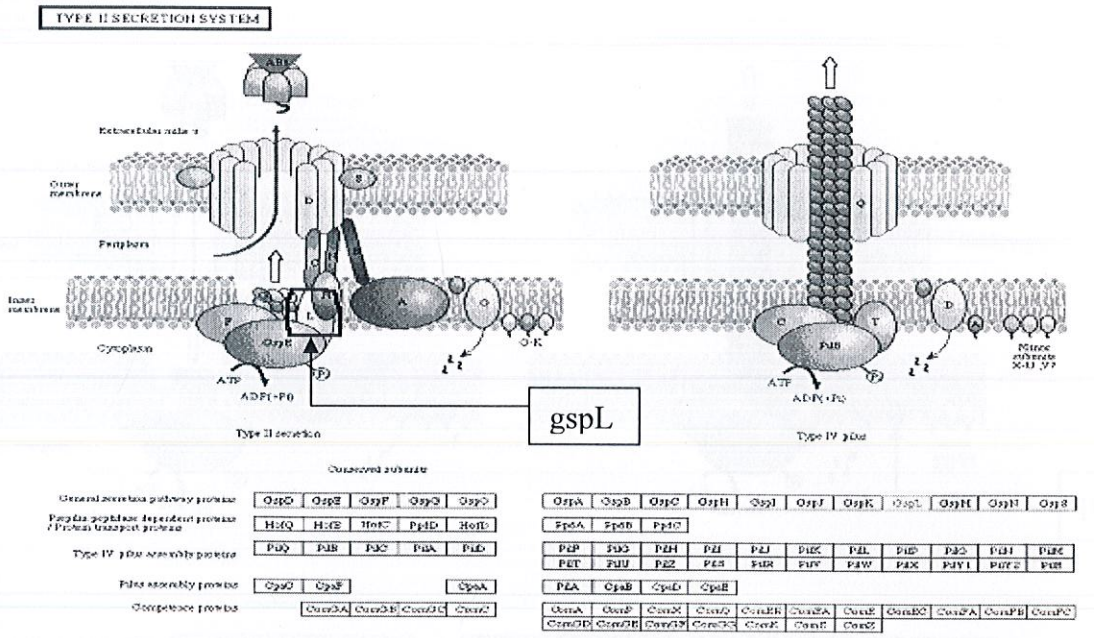


Fig 2.8 : Location of the enzyme General secretary pathway protein L in type II secretion system

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.

2. pilA

Pili are extracellular filaments, found on a wide variety of bacteria, that play an important role in adhesion of pathogenic bacteria to their host, biofilm formation, conjugative DNA transfer, non-flagellar motility and bacteriophage infection (Clewell, 1993; O'Toole and Kolter, 1998; Soto and Hultgren, 1999; Wall and Kaiser, 1999). Some bacteria have pili localized to the poles of the cells, and others have pili covering the entire cell surface. These structures are formed by the polymerization of pilin subunits at the base of the growing filament (Lowe et al., 1987). In Gram-negative bacteria, the pilin subunit must be secreted across both inner and outer membranes before being assembled into an extracellular filament.

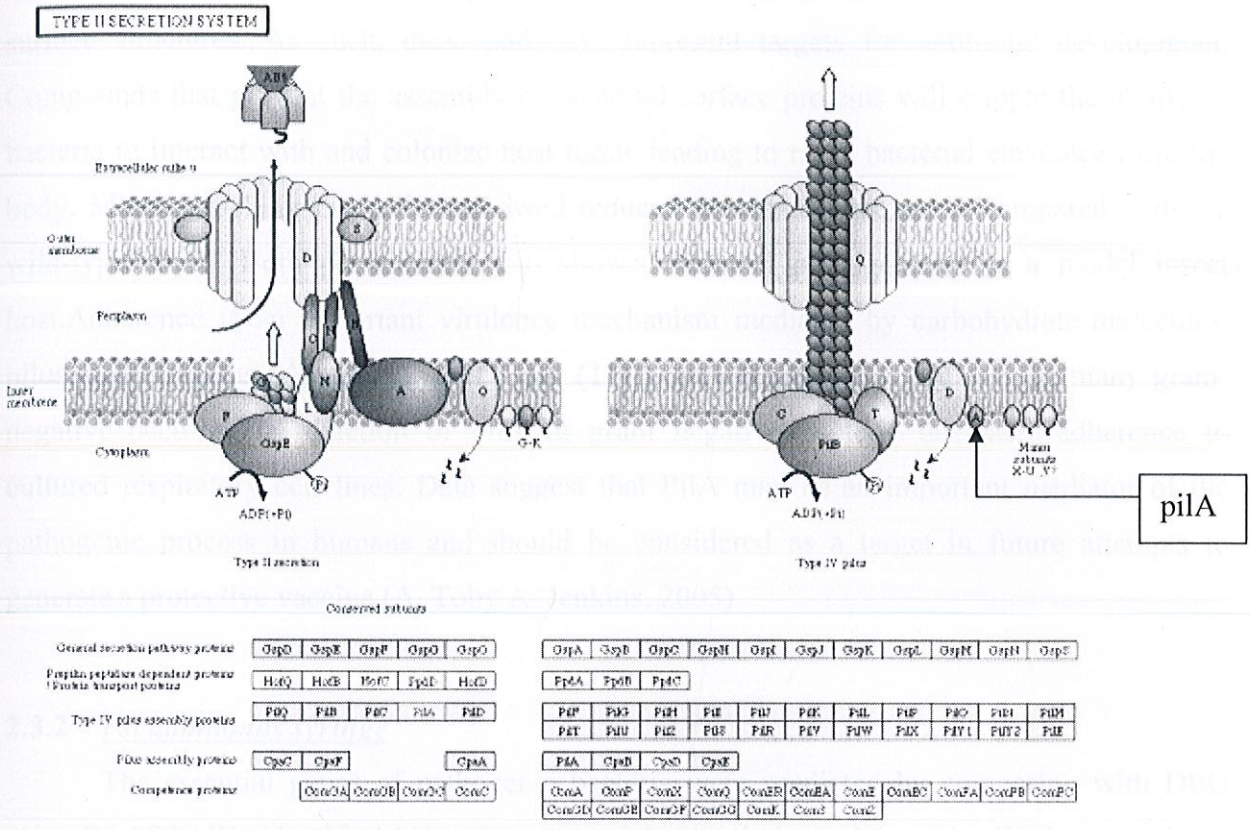


Fig 2.13 : Location of the enzyme pilA in type II secretion system.

pilA stands pilus assembly protein A.*P. aeruginosa* virulence factors such as proteases, flagella, pili and LPS O side chains. *P. aeruginosa* elaborates a multitude of factors which affect the initiation and the course of infection. Among these are: glycocalyx, endotoxin, exotoxin, protease, flagella and pili. *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 secretion system (Jeffrey B. Lyczaka,b, 2000). 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. Mutants lacking pili (pilA), showed reduced measured attachment compared with the wild-type strain. Both pil mutants also showed reduced pathogenicity in a model insect host.Adherence is an important virulence mechanism mediated by carbohydrate molecules, pilus, and nonpilus adhesins .Type IV pili (TFP) are important for virulence in many gram-negative bacteria.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 (A. Toby A. Jenkins, 2005)

2.3.2 ***Pseudomonas syringe***

The essential genes of pathogenic bacteria were predicted by comparing with DEG using BLASTn We identified 160 unique essential genes in bacterial species.We have used two step processes and further metabolic analysis 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).

The predicted essential genes of pathogenic bacteria were compared with the predicted essential genes of non-pathogenic bacteria using BLASTn. It is revealed that out of 160 essential genes only 43 genes are uniquely present in *P.syringea,* which could be related with pathogenicity. The essential genes obtained after first subtraction were then compared with EST of phaseolus vulgaris (BLASTn), 42 essential genes were uniquely identified from *P.syringea.*

The predicted probable therapeutic targets after comparing with host EST were taken further for comparative metabolomics.. This results in further reduction in the number of unique therapeutic targets as listed in table 2.5.

These enzymes present uniquely in the *P. syringea* (unique metabolic pathways) were given first priority and considered most potent therapeutic targets. A total of 24 unique metabolic pathways were identified as listed in table 2.6 when comapared with Arabdopsis(model plant).

We have also compared all the pathways of our pathogen and the non pathogenic bacteria (*Pseudomonas putida*KT2440) and found out 2 pathways unique to our pathogen as listed in table 2.7.Enzyme commission numbers (EC) of the enzymes belonging to metabolic pathways were extracted from the KEGG database and with the help of those the enzyme sequences in FASTA format from each of the above got metabolic pathway of *Pseudomonas syringae* B728a were extracted. Using BLASTx unique enzymes as compare to *Pseudomonas putida* KT2440 were compared with gene dataset got after genomics comparisons. E-value of e^{-5} was taken as cutoff. Those enzymes which matched with genomics result dataset was taken as result dataset for this step. These are the essential unique enzymes present in the unique pathways of the pathogen. This step eliminated none of the enzymes. These were further compared with the enzyme dataset got after metabolic comparison of Arabidopsis and *Pseudomonas syringae* B728a using BLASTx taking an E-value of e^{-3} as cutoff. The enzymes which matched with the resultant enzymes of metabolic comparison of Arabidopsis and *Pseudomonas syringae* B728a were taken as final resultant dataset that 8 in numberas listed in table 2.8 and their classification based on their function is shown in figure 2.11.These enzymes are the best potential drug targets essential and uniquely present in the metabolic pathways of the pathogen.Figure 2.10 shows the subsequent decrease in the number of genes after different comparisons.

Table 2.5: Unique essential genes of *P. syringe* after comaparitive genomics.

	GENE ID	NAME	DEG ID	SCORE	PATHWAY
	CARBOHYDRATE METABOLISM				
1	Psyr_0517		DEG10040028	66	Glycolysis / Gluconeogenesis
2	Psyr_0517		DEG10040028	66	Pyruvate metabolism
3	Psyr_1120		DEG10040316	62	Pentose phosphate pathway
4	Psyr_2176		DEG10110051	70	Propanoate metabolism
5	Psyr_2988		DEG10100164	68	Pentose phosphate pathway
	ENERGY METABOLISM				
1	Psyr_1238		DEG10040392	70	Nitrogen metabolism
2	Psyr_3483		DEG10040067	70	Nitrogen metabolism
	LIPID METABOLISM				
1	Psyr_0500		DEG10100527	74	Fatty acid biosynthesis
2	Psyr_1349		DEG10040043	62	Biosynthesis of steroids
3	Psyr_1647	fabG	DEG10040181	72	Fatty acid biosynthesis
4	Psyr_1647	fabG	DEG10040181	72	Biosynthesis of unsaturated fatty acids
5	Psyr_1754		DEG10110097	62	Fatty acid biosynthesis
6	Psyr_2176		DEG10110051	70	Fatty acid metabolism
	NUCLEOTIDE METABOLISM				
1	Psyr_1269		DEG10130197	72	Purine metabolism
	AMINO ACID METABOLISM				
1	Psyr_0012	glyQ	DEG10040536	230	Glycine, serine and threonine metabolism
2	Psyr_0183	dapF	DEG10040559	64	Lysine biosynthesis
3	Psyr_0487		DEG10040446	66	Glutamate metabolism
4	Psyr_0517		DEG10040028	66	Alanine and aspartate metabolism
5	Psyr_1735	cysS	DEG10040100	74	Cysteine metabolism
6	Psyr_1985		DEG10100480	66	Valine, leucine and isoleucine biosynthesis
7	Psyr_2176		DEG10110051	70	Valine, leucine and isoleucine degradation
8	Psyr_4107	murE	DEG10110006	64	Lysine biosynthesis
9	Psyr_4893		DEG10100263	66	Histidine metabolism
10	Psyr_4104	murD	DEG10110006	52	D-Glutamine and D-glutamate metabolism
	METABOLISM OF OTHER AMINO ACID METABOLISM				

1	Psyr_0487		DEG10040446	66	Glutathione metabolism
2	Psyr_1120		DEG10040316	62	Glutathione metabolism
3	Psyr_2176		DEG10110051	70	beta-Alanine metabolism
4	Psyr_2988		DEG10100164	68	Glutathione metabolism
	GLYCAN BIOSYNTHESIS AND METABOLISM				
1	Psyr_4107	murE	DEG10110006	64	Peptidoglycan biosynthesis
2	Psyr_4104	murD	DEG10110006	52	Peptidoglycan biosynthesis
	METABOLISM OF COFACTORS AND VITAMINS				
1	Psyr_0387		DEG10040563	96	Ubiquinone biosynthesis
	TRANSLATION				
1	Psyr_0012	glyQ	DEG10040536	230	Aminoacyl-tRNA biosynthesis
2	Psyr_1735	cysS	DEG10040100	74	Aminoacyl-tRNA biosynthesis
3	Psyr_2165	rplT	DEG10010205	64	Ribosome
4	Psyr_4120	rplM	DEG10040487	86	Ribosome
5	Psyr_4544	rpsS	DEG10040508	58	Ribosome
6	Psyr_4639	rpsU	DEG10130100	60	Ribosome
	MEMBRANE TRANSPORT				
1	Psyr_0759		DEG10100446	60	ABC transporters
2	Psyr_3076		DEG10110057	62	ABC transporters

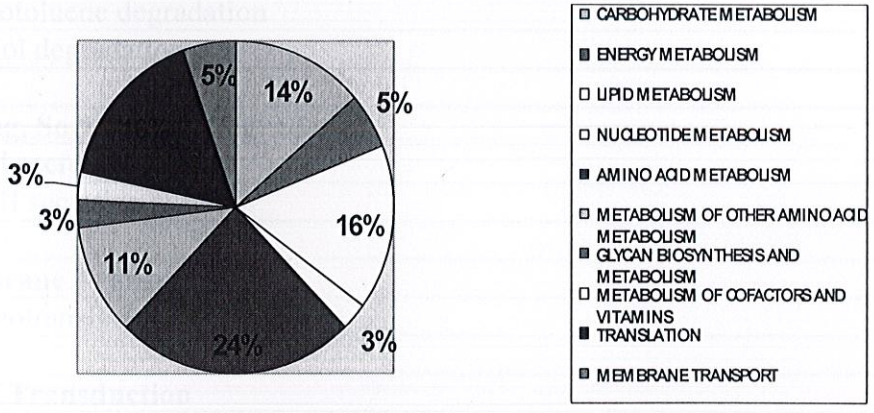


Fig 2.9 : Classification of unique essential genes *P. syringe* after comaparitive genomics on the basis of function.

	Carbohydrate Metabolism
1	C5-Branched dibasic acid metabolism
	Lipid Metabolism
1	C21-steroid hormone metabolism
2	Androgen and estrogen metabolism
	Metabolism of Other Amino Acids
1	D-Glutamine and D-glutamate metabolism
2	D-Alanine metabolism
	Biosynthesis of Polyketides and Nonribosomal Peptides
1	Polyketide sugar unit biosynthesis
	Biosynthesis of Secondary Metabolites
1	Streptomycin biosynthesis
2	Novobiocin biosynthesis
	Xenobiotics Biodegradation and Metabolism
1	Caprolactam degradation
2	Toluene and xylene degradation
3	2,4-Dichlorobenzoate degradation
4	1,2-Dichloroethane degradation
5	Ethylbenzene degradation
6	Fluorene degradation
7	Atrazine degradation
8	Bisphenol A degradation
9	Trinitrotoluene degradation
10	Geraniol degradation
	Folding, Sorting and Degradation
1	Type II secretion system
2	Type III secretion system
	Membrane Transport
1	Phosphotransferase system (PTS)
	Signal Transduction
1	Two-component system
	Cell Motility
1	Bacterial chemotaxis
2	Flagellar assembly

Table 2.6 : Unique pathways present in *P.syrngea* when compared with Arabdopsis

	UNIQUE PATHWAYS AS COMPARED TO PUTIDA
	Lipid Metabolism
1	C21-steroid hormone metabolism
2	Androgen and estrogen metabolism

Table 2.7: Unique pathways of *P.syringea* when compared against pathways of *P.putida* KT2440

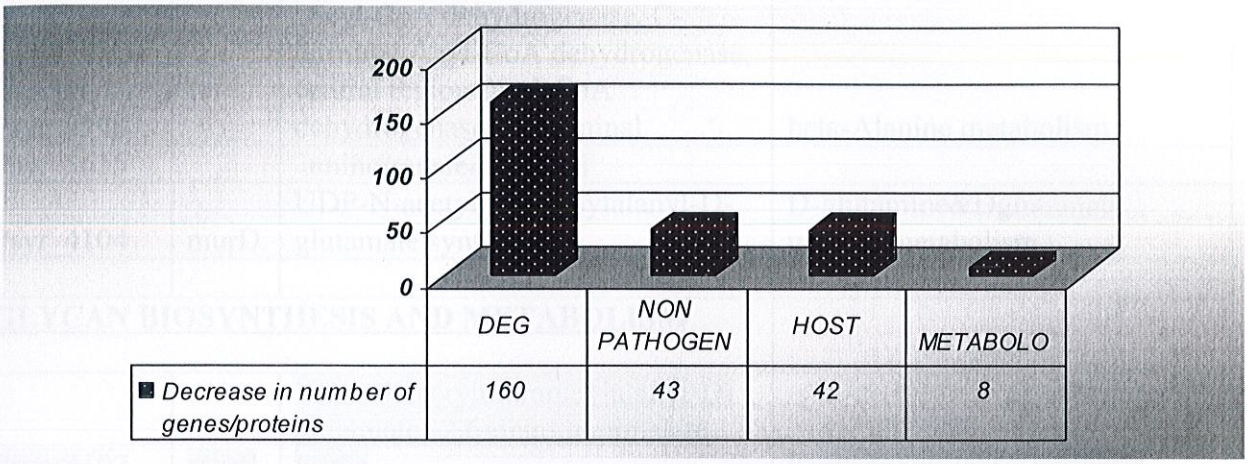


fig 2.10 : Subsquent decrease in the number of genes/proteins after complete genomics and metabolomics analysis

METABOLISM OF COFACTORS AND VITAMINS			
GENE ID	Name	Protein name	pathway
Psyr_0387		2-polyprenylphenol 6-hydroxylase	Ubiquinone biosynthesis
LIPID METABOLISM			
Psyr_2176		Acyl-CoA dehydrogenase, C-terminal:Acyl-CoA dehydrogenase, central region:Acyl-CoA dehydrogenase, N-terminal	Fatty acid metabolism
AMINO ACID METABOLISM			
Psyr_2177		Acyl-CoA dehydrogenase, C-terminal:Acyl-CoA dehydrogenase, central region:Acyl-CoA dehydrogenase, N-terminal	Valine, leucine and isoleucine degradation
Psyr_4107	murE	UDP-N-acetylmuramoylalanyl-D-	Lysine biosynthesis

		glutamate--2,6-diaminopimelate ligase	
CARBOHYDRATE METABOLISM			
Psyr_2178		Acyl-CoA dehydrogenase, C-terminal:Acyl-CoA dehydrogenase, central region:Acyl-CoA dehydrogenase, N-terminal	Propanoate metabolism
OTHER AMINO ACID METABOLISM			
Psyr_2178		Acyl-CoA dehydrogenase, C-terminal:Acyl-CoA dehydrogenase, central region:Acyl-CoA dehydrogenase, N-terminal	beta-Alanine metabolism
Psyr_3615		aminotransferase AlaT	
Psyr_4104	murD	UDP-N-acetylmuramoylalanyl-D-glutamate synthetase	D-glutamine&Dglutamate metabo metabolism
GLYCAN BIOSYNTHESIS AND METABOLISM			
Psyr_4107	murE	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	Peptidoglycan biosynthesis
Psyr_4104	murD	UDP-N-acetylmuramoylalanyl-D-glutamate synthetase	Peptidoglycan biosynthesis

Table 2.8 : Potential targets present in *P.syrngea* after the complete genomic and metabolomic comparisons

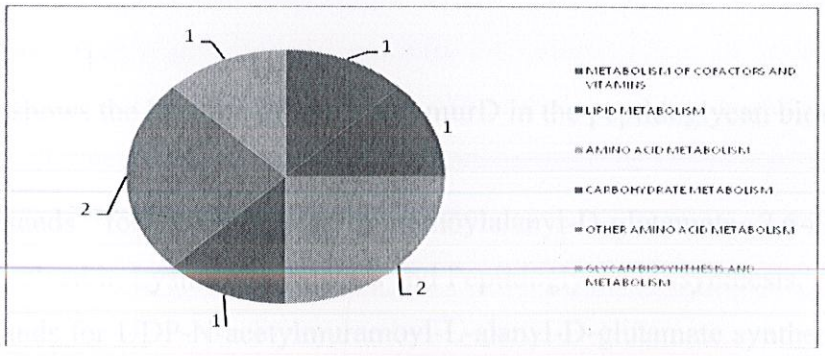


Fig 2.11 : Classification of potent targets present in *P.syrngea* after the complete genomics and metabolomic comparisons on the basis of function.

TARGETS SELECTED:

murD and murE

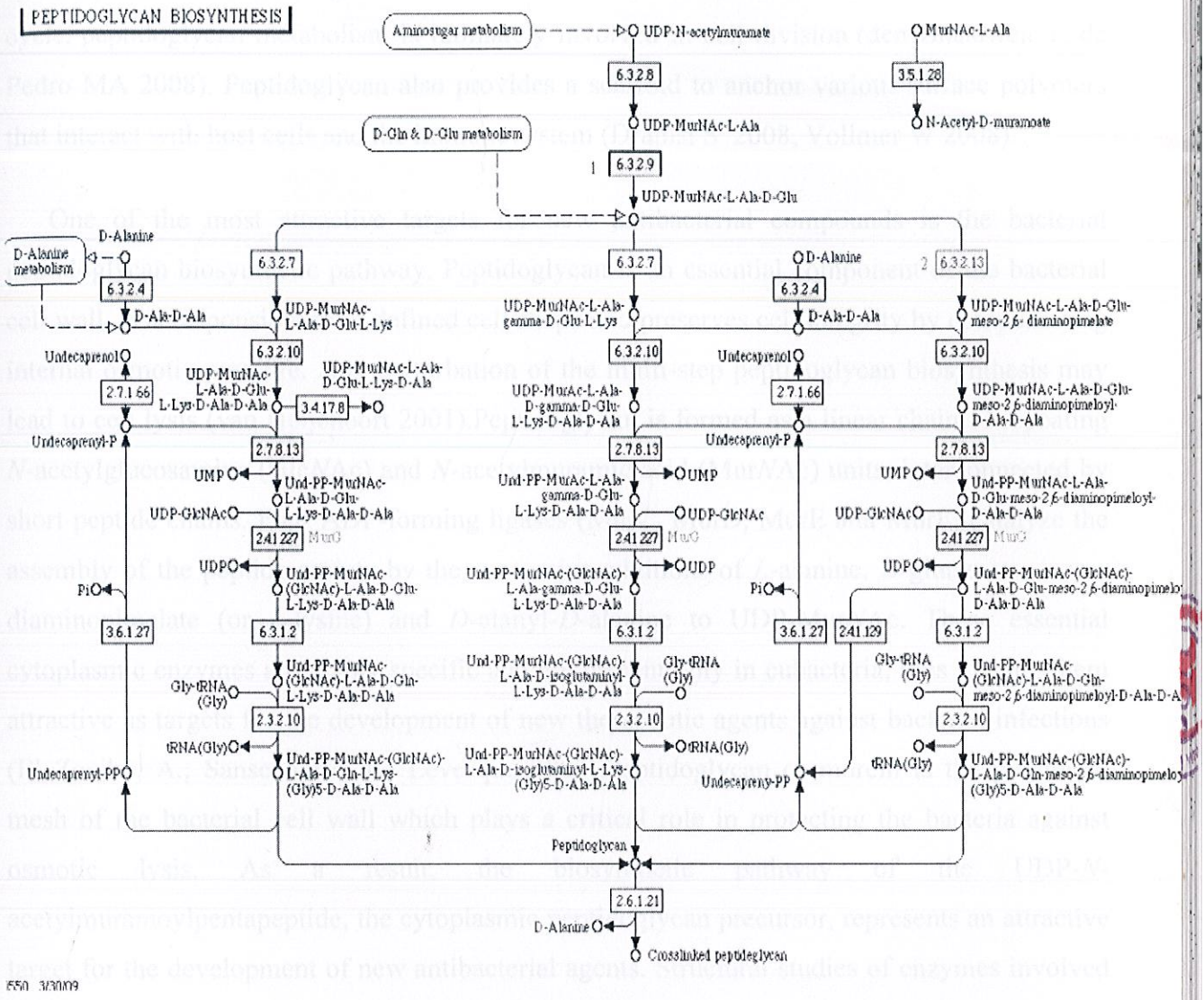


Fig 2.12: shows the location of murE and murD in the peptidoglycan biosynthesis

1. murE stands for UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase involved in Lysine biosynthesis and Peptidoglycan biosynthesis.
2. murD stands for UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase involved in D-Glutamine and D-glutamate metabolism and also in Peptidoglycan biosynthesis.

Peptidoglycan is a giant macromolecule, that completely surrounds the cytoplasmic membrane and thereby provides a mechanical protection against the turgor pressure of the cytoplasm. Since the osmoprotective function is required in continuity throughout the cell cycle, peptidoglycan metabolism is intimately involved in cell division (den Blaauwen T, de Pedro MA 2008). Peptidoglycan also provides a scaffold to anchor various surface polymers that interact with host cells and the immune system (Dramsi S 2008, Vollmer W 2008).

One of the most attractive targets for new antibacterial compounds is the bacterial peptidoglycan biosynthetic pathway. Peptidoglycan is an essential component of the bacterial cell wall. It is responsible for a defined cell shape and preserves cell integrity by compensating internal osmotic pressure. Any perturbation of the multi-step peptidoglycan biosynthesis may lead to cell lysis (van Heijenoort 2001).Peptidoglycan is formed as a linear chain of repeating *N*-acetylglucosamine (Glc*N*Ac) and *N*-acetylmuramic acid (Mur*N*Ac) units, interconnected by short peptide chains. Four ADP-forming ligases (MurC, MurD, MurE and MurF) catalyze the assembly of the peptide moiety by the successive additions of *L*-alanine, *D*-glutamate, *meso*-diaminopimelate (or *L*-lysine) and *D*-alanyl-*D*-alanine to UDP-Mur*N*Ac. These essential cytoplasmic enzymes are highly specific and are present only in eubacteria, thus making them attractive as targets for the development of new therapeutic agents against bacterial infections (El Zoeiby, A.; Sanschagrin, F.; Levesque 2003).Peptidoglycan or murein is the polymeric mesh of the bacterial cell wall which plays a critical role in protecting the bacteria against osmotic lysis. As a result, the biosynthetic pathway of the UDP-*N*-acetylmuramoylpentapeptide, the cytoplasmic peptidoglycan precursor, represents an attractive target for the development of new antibacterial agents. Structural studies of enzymes involved in the pathway have already provided strategies for the rational design of novel inhibitors.In an attempt to discover broad spectrum antibiotics, the inhibition of the D-glutamic acid adding enzyme UDP-N-acetylmuramoyl-L-alanine:D glutamate ligase or MurD was chosen as a target in studies. The MurD enzyme is involved in the biosynthesis of bacterial peptidoglycan in both gram-negative and gram-positive bacteria, therefore, inhibitors of this enzyme should also be broad-spectrum antibiotics (Jay A. Bertrand 1997)

The bacterial cell wall is a polymer--a single molecule composed of peptidoglycan--that defines the boundary and shape of the cell. Of the enzymes involved MurD and MurE mediates

the formation of a peptide bond between the g- carboxylate of D-glutamate and the amino group of L-lysine. Presumably these structures render the exposed peptidoglycan resistant to the action of proteases, but they also imply that the active sites of the enzymes must have unusual structures in order to handle the somewhat uncommon substrates. These unusual active sites are targets to bind novel inhibitors that can have antimicrobial activity. Among these potential enzyme targets is MurD and murE.

Chapter 3
LEAD DESIGNING

3.1 Introduction

The mechanism of action of any drug is very important in drug development. Generally, the drug compound binds with a specific target, a receptor, to mediate its effects. Therefore, suitable drug–receptor interactions are required for high activity. Understanding the nature of these interactions is very significant and theoretical calculations, in particular the molecular docking method, seem to be a proper tool for gaining such understanding. The docking results obtained will give information on how the chemical structure of the drug should be modified to achieve suitable interactions and for the rapid prediction.

3.2 Pseudomonas aeruginosa

Pseudomonas aeruginosa, a ubiquitous gram-negative bacterium, has been intensively studied as an opportunistic human pathogen (Britigan et al., 1997) and as the dominant pathogen infecting the lungs of cystic fibrosis patients (Pier et al., 1996). *Pseudomonas aeruginosa* is an opportunistic pathogen, meaning that it exploits some break in the host defenses to initiate an infection. In fact, Pseudomonas aeruginosa is the epitome of an opportunistic pathogen of humans. The bacterium almost never infects uncompromised tissues, yet there is hardly any tissue that it cannot infect if the tissue defenses are compromised in some manner. It causes urinary tract infections, respiratory system infections, dermatitis, soft tissue infections, bacteremia, bone and joint infections, gastrointestinal infections and a variety of systemic infections, particularly in patients with severe burns and in cancer and AIDS patients who are immunosuppressed. Its resistance towards exsisting antibiotics urges us to look for new and new drug targets.

The potential targets used for lead designing includes:

1. dapD
2. gspL
3. pilA

DapD encodes for Tetrahydrodipicolinate succinylase. Tetrahydrodipicolinate N-succinyltransferase 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 (Todd and Beaman 2002) 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 (Hutton et al. 2003).

Pseudomonas aeruginosa utilizes a number of distinct pathways to secrete proteins that play various roles during infection .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 (Vincent and Lee 2005).

Pili are extracellular filaments, found on a wide variety of bacteria, that play an important role in adhesion of pathogenic bacteria to their host, biofilm formation, conjugative DNA transfer, non-flagellar motility and bacteriophage infection (Clewell, 1993; O'Toole and Kolter, 1998; Soto and Hultgren, 1999; Wall and Kaiser, 1999). .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 (Toby and Jenkins 2005).

3.2.1 Materials and methods

3.2.1.1 Sequence analysis

The protein sequence of Tetrahydrodipicolinate succinylase (dapD) , General Secretory pathway protein L (gspL) and type IV pilus assembly protein A (pilA) of the organism *Pseudomonas aerugimosa* PA7 was obtained from the NCBI database (accession number PSPA7_1473, PSPA7_2037, PSPA7_4873 respectively). Even though dapD, gspL and pilA homologues are not known to be present in humans, their sequences were subjected to various sequence analysis programs as explained below to examine their potential as a drug target. Pairwise sequence alignment of dapD, gspL and pilA from *Pseudomonas aeruginosa* was carried out against human database using BLASTx as well as against the *Pseudomonas putida* and symbiotic organism without any significant hit. Blastp against PDB databank gave significant hits. Template was selected on the basis of identities, score, gaps and e-value. These proteins had given a hit with 1EFD_N (with dapD, score 30, 37% identity, 56% positive scores and 4% gaps with an e-value of 0.94), 1W97_L (with gspL, score 32.3, 34% identity, 60% positive scores and 6% gaps with an e-value of .22) ; 2H12_A (with pilA , score 46.6 , 35% identity, 43% positive scores and 4% gaps with an e-value of 1.7e-06).

3.2.1.1 Homology model construction

The three dimensional structure of the protein was constructed using Prime .To select the templates, a basic local alignment search tool (BLAST) search against the PDB was conducted using our protein sequence as a query (Fig: 3.1). Only the structures with a Blast Bit-score above 100 were kept. The following criteria were then used to discard structures: (a) no ligand bound (apart from one structure, see below), (b) missing residues in the binding site, (c) inactive conformation, and (d) structural similarity of the binding site (for the structures sharing a very similar structural binding site, only one was kept to limit the number of homology models). Furthermore, the homology models were built using Prime accessible through the Maestro interface . All water molecules were removed and the bound ligand was kept. The alignment of the binding site residues was always unambiguous, since there was no gap in the sequences for these residues. 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. Known binders and database for high throughput docking 1,25,000 compounds were randomly selected from Pubchem Database. This database includes both the 2D coordinates (Mol file) and 3D coordinates (mol2 file) of the ligands. The compounds with 3D coordinates in mol2 format were download and prepared using Ligprep module of Schrodinger. The compound structures were ionized assuming a pH of 7.0 with the ionizer module supplied by Schrödinger. Compounds were finally subject to a minimization based on the OPLS-AA (2005) force-field.

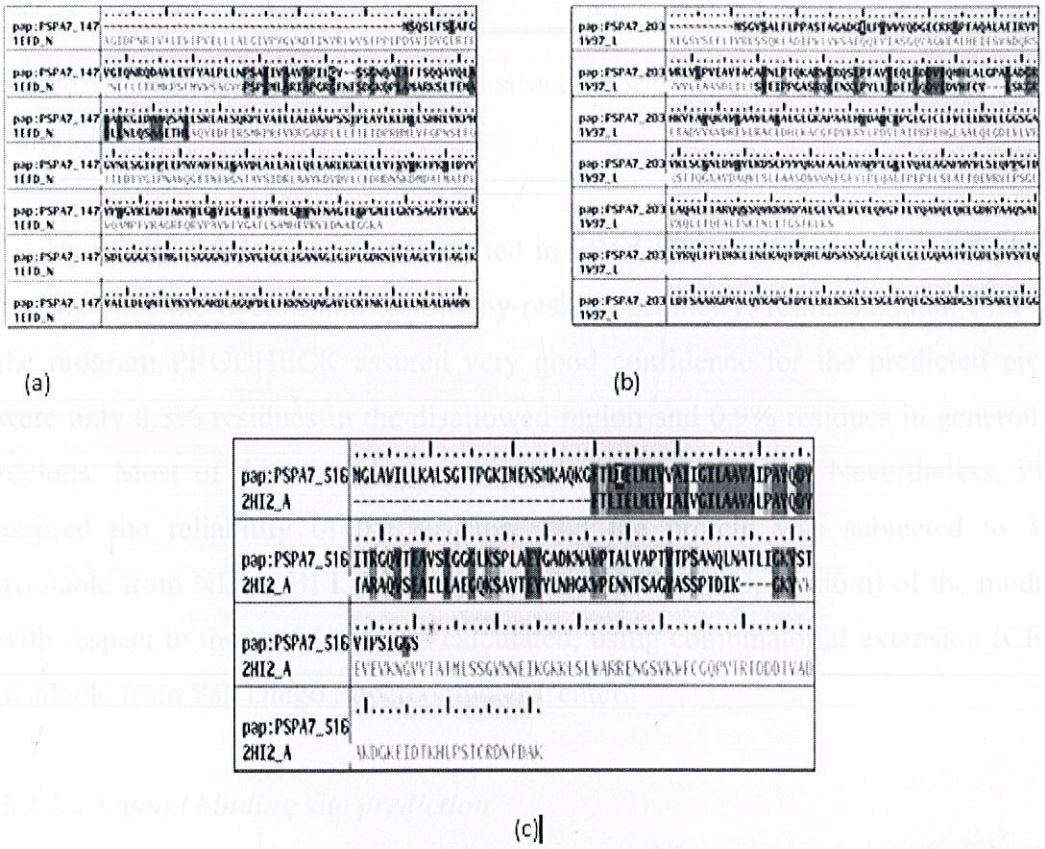


Figure 3.1 : Sequence Alignment of protein dapD(a), gspL(b) and pilA (c) of *P.aeruginosa* with the templates

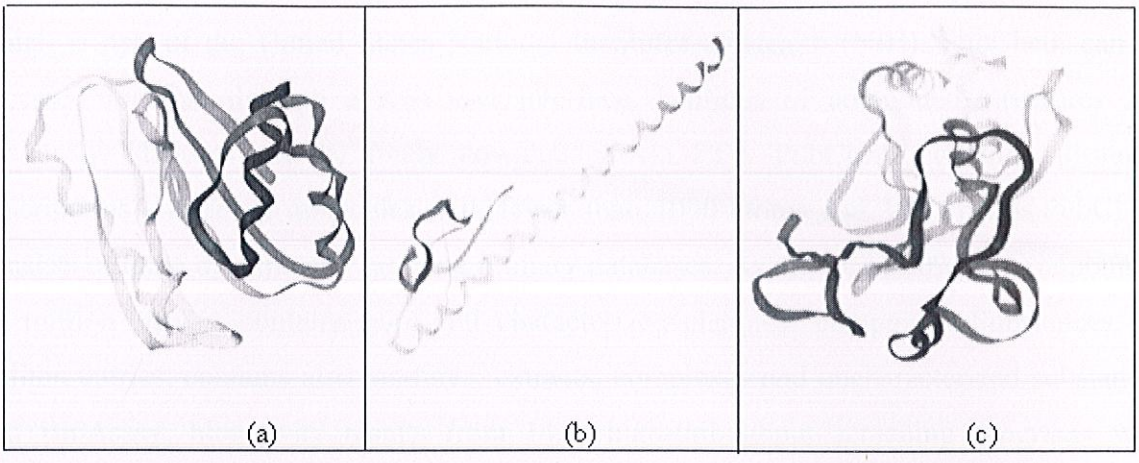


Figure 3.2 Modeled three-dimensional structure of gspL(a) , pilA(b) and dapD (c) of *P.aeruginosa*

Model evaluation was performed in PROCHECK v3.4.4 producing plots that were analyzed for the overall and residue-by-residue geometry. Ramachandran Plot 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. Most of these residues were in C-terminal domain. Nevertheless, PROCHECK assured the reliability of the structure and the protein was subjected to VERIFY3D, available from NIH MBI Laboratory Servers. RMSD (c-alpha atom) of the modeled protein with respect to the templates was calculated, using combinatorial extension (CE) algorithm available from San Diego Supercomputer Center.

3.2.1.3 Ligand binding site prediction

Binding sites from the modeled structure of the targets were predicted using Sitemap (Schrodinger). A number of binding sites were predicted. The site with maximum score was finally selected for virtual screening.

3.2.1.4 Virtual library ligand dataset

The virtual library created for screening consisted of 1,25,000 compounds compiled together to form a stabndalone library taken from pubchem drug databse. **PubChem** is a database of chemical molecules. The system is maintained by the National Center for

Biotechnology Information (NCBI), a component of the National Library of Medicine, which is part of the United States National Institutes of Health (NIH). PubChem can be accessed for free through a web user interface. Millions of compound structures and descriptive datasets can be freely downloaded via FTP. PubChem contains substance descriptions and small molecules with fewer than 1000 atoms and 1000 bonds.PubChem consists of three dynamically growing primary databases. As of April 2009 i.e. Compounds, 19 million entries, contains pure and characterized chemical compounds,Substances, 48 million entries, contains also mixtures, extracts, complexes and uncharacterized substances and BioAssay, bioactivity results from 1459 high-throughput screening programs with several million values

3.2.1.5 Ligand Preparation

All protein structures were prepared according to the standard procedure described in the Glide user manual. The default input parameters implemented in Maestro were used for the generation of the command files for the docking of the small-molecule database against the modelled crystal structures and homology models. For the generation of the scoring grid, the van der Waals radii of the non-polar protein atoms were not scaled. Compounds were docked using Glide in single-precision mode. The best pose for each compound, as evaluated by the Glide scoring function, was written out. The Schrodinger Glide program version 4.0 has been used for docking (Friesner et al., 2004; Halgren, 2004). The best 10 poses and corresponding scores have been evaluated using Glide in single precision mode (Glide SP) for each ligand from the virtual library of podophyllotoxin. For each screened ligand, the pose with the lowest Glide SP 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 van der Waals radii of receptor atoms by 1.00 with partial atomic charge 0.25.

3.2.1.6 Docking procedure

LigPrep (Schrodinger 2007)was used for final preparation of ligands from libraries for docking. LigPrep is a utility of Schrodinger software suit that combines tools for generating 3D structures from 1D (Smiles) and 2D (SDF) representation, searching for

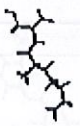
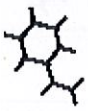
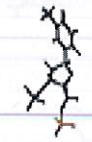
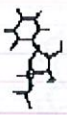
were minimized by means of Molecular Mechanics Force Fields (OPLS-2005) with default setting.

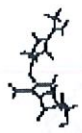
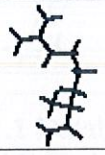
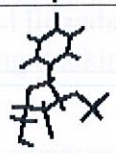
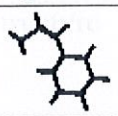
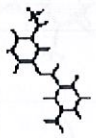
3.2.2 RESULTS AND DISCUSSIONS

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. Our initial work in sequence analysis and domain study suggested to pick up dapD, the key enzyme in lysine biosynthesis,gsp L the key enzyme in type II secretory system and pilA , a key type IV pilus assembly protein in *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.

The atomic coordinates of dapD, gspL and pilA for the organism *P. aeruginosa* were not available in Protein Data Bank, which necessitated developing a protein model. Homology modeling protocol was employed to predict the 3D structure of the protein. An identity of 37% with well-studied protein of E.coli (Resolution—1.9) for dapD and an identity of 34% with well-studied protein of Vibrio Cholerae (Resolution—2.7) for gspL and an identity of 35% with well-studied protein of Acetobacter aceti (Resolution—1.85) for pilA provided a great strength for modeling the protein. Three dimensional structure prediction by comparative modeling was done by Prime. Macromodel module of schrodinger used in energy minimization embraces a range of force fields. We used both PROCHECK and the VERIFY3D softwares to check the quality of the modeled protein. We carried out in silico studies to find the active sites, using sitemap and ac-site. The output from the sitemap program showed coherent active sites for the target protein as from the ac-site. The best ligands show in Table 3.1 were chosen based on the glide scores obtained from the gilde docking and the consensus scoring program Glide-XP. 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. The Glide Score function consisted of four components: (a) protein–ligand hydrogen bond energy (external H-bond); (b) protein–ligand van der Waals (vdw) energy (external vdw); (c)ligand internal vdw energy (internal vdw); (d) ligand torsional strain energy (internal torsion). On the other hand, the scoring schema used in the software Glide-XP computes a binding score for a given protein–ligand complex structure, and this binding score correlates to experimental binding constants well. It takes into account van der Waals interactions, hydrogen bonding, deformation penalty, and hydrophobic effects between the receptor and the ligand. The best 20 ligands as shown in Table 3.1 chosen with Glide score proved their reliability in the Glide-XP. Further in the paper, name of the ligands are indicated by the ligand ID's given in the database. 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.

S.NO	Structure	ZINC ID	IUPAC Name	Glide score	GlideXP score	Glide energy
1		ZINC01888932	(2S)-2-[[(2S)-2,6-diaminohexanoyl]amino]-3-hydroxypropanoic acid	-6.894	-5.456	-27.408
2		ZINC03786623		-6.709	-5.769	-14.769
3		ZINC03870884	5-(5-methyl-2,4-dioxopyrimidin-1-yl)-2-(phosphonooxymethyl) oxolan-3-yl	-6.242	-5.281	-30.034
4		ZINC03870179	5-(2,4-dioxo-1Hpyrimidin-5-yl)-3,4-dihydroxyoxolan-2-yl]methyl dihydrogen phosphate	-6.237	-4.607	-29.120

5		ZINC03780837	3-(2-methylpiperidin-1-ium-1-yl)propyl 4-(3-methylbutoxy)benzoate chloride	-6.224	-5.444	-26.323
6		ZINC06483376	2-(diaminomethylideneamino)ethyl methyl hydrogen phosphate	-6.191	-4.970	-17.704
7		ZINC06487269	4-azanidylpentan-2-ylazanide; dichloroplatinum(2+)	-6.026	-5.522	-22.010
8		ZINC03871599	Pentanamide, 2-amino-5-(aminoiminomethyl)amino]-	-5.834	-5.104	-20.869
9		ZINC03875251	2-(2,4-dioxopyrimidin-1-yl)-4-hydroxy-5-(hydroxymethyl)oxolan-3-yl	-5.777	-4.83	-25.790
10		ZINC03786627	2-[(2R)-piperidin-2-yl]ethanamine	-5.715	-4.721	-16.128
11		ZINC00006017	5-chloro-7-(trifluoromethyl)-1,4-dihydroquinoxaline-2,3-dione	-5.598	-4.787	-14.294
12		ZINC03870260	disodium [5-(2,4-dioxopyrimidin-1-yl)-3,4-dihydroxyoxolan-2-yl]methyl	-5.462	-4.206	-27.253
13		ZINC00262043	1-[3-[(2-hydroxy-5-nitrophenyl)methylideneamino]phenyl]ethanone	-5.350	-4.183	-24.786
14		ZINC03581167	2-azanidylcyclohexyl) azanide;oxonio(oxoniocarbonyl)phosphinate;platinum(4+)	-5.348	-6.234	-15.795
15		ZINC03870177	5-(2,4-dioxo-1Hpyrimidin-5-yl)-3,4-dihydroxyoxolan-2-yl]methyl dihydrogen phosphate	-5.216	-5.098	-31.002
16		ZINC03870733	3,4-dihydroxy-5-(phosphonatooxymethyl)oxolan-2-yl]methyl	-5.162	-4.574	-29.744

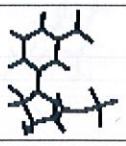
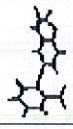
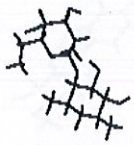
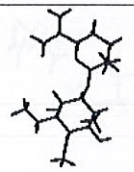
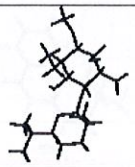
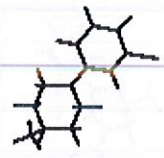
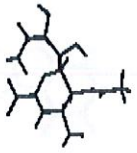
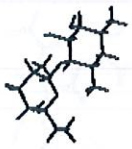
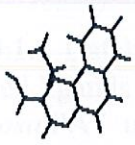
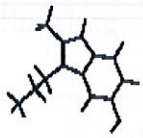
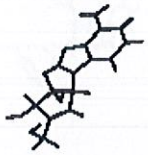
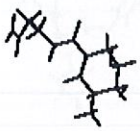
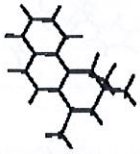
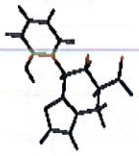
			phosphate			
17		ZINC03869332	5-[(2-aminoacetyl)amino]-3,4-dihydroxyoxolan-2-yl]methyl dihydrogen phosphate	-5.138	-5.753	-25.695
18		ZINC03870114	4-[(4-aminophenyl)sulfonylamino]benzoic acid	-5.114	-4.357	-27.960
19		ZINC02051031	2-(3-oxo-4H-1,4-benzoxazine-6-carbonyl)benzoate	-5.082	-4.738	-24.158

Table 3.1.1 Database ID number, chemical structure, IUPAC names and fitness scores of top scored ligands docked Tetrahydrodipicolinate succinylase of *Pseudomonas aeruginosa* PA7 using docking program Glide

S. No.	Structure	ZINC ID	IUPAC Name	Glide score	GlideXP score	Glide energ
1		ZINC03831196	5-amino-2-(aminomethyl)-6-(4,6-diamino-2,3-dihydroxycyclohexyl)oxyoxane-3,4-diol	-7.4536	-10.542	-52.
2		ZINC03830242	2-amino-3-[3-amino-6-(1-aminoethyl)oxan-2-yl]oxy-6-methoxy-5-(methylamino)cyclohexane-1,4-diol	-6.859	-7.826	-42.
3		ZINC03830241	2-amino-3-[3-amino-6-(1-aminoethyl)oxan-2-yl]oxy-6-methoxy-5-(methylamino)cyclohexane-1,4-diol	-6.91	-7.485	-42.
6		ZINC02383344	1-(6-chloropyridin-2-yl)piperidin-4-amine	-6.862	-7.242	-35.

7		ZINC03870976	3-acetamido-4-amino-2-(3-amino-1,2-dihydroxypropyl)-3,4-dihydro-2H-pyran-6-carboxylic acid	-6.517	-7.172	-41.
8		ZINC03831194	5-amino-2-(aminomethyl)-6-(4,6-diamino-2,3-dihydroxycyclohexyl)oxy oxane-3,4-diol	-7.332	-6.610	-45.
9		ZINC02388213	[2-(hydroxymethyl)-2,3-dihydro-1H-benzo[f]chromen-1-yl]-methylazanium chloride	-6.034	-6.183	-31
10		ZINC02572146	3-(2-aminoethyl)-2-methyl-1H-indol-5-ol	-6.665	-6.116	-30.
11		ZINC03834078	(1S,2R,5R)-5-(4-aminoimidazo[4,5-c]pyridin-1-yl)-3-(hydroxymethyl)cyclopent-3-ene-1,2-diol	-6.497	-5.653	-27.
12		ZINC03839685	3-amino-N-(3-amino-3-oxopropyl)-4,5-dihydroxycyclohexene-1-carboxamide	-6.771	-5.597	-36.
13		ZINC02479996	2,2,4-trimethyl-3,4-dihydro-1H-benzo[f]isoquinoline	-6.699	-5.433	-32.
14		ZINC03984186	4-(2-hydroxyphenyl)-4,5,6,7-tetrahydro-1H-imidazo[4,5-c]pyridine-6-carboxylic acid	-6.921	-5.2669	-30.

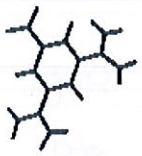
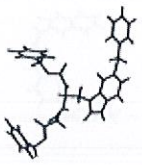
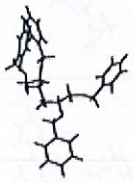
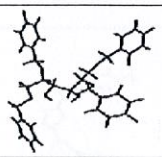
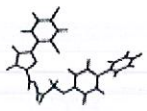
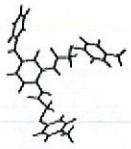
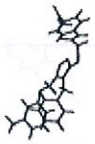
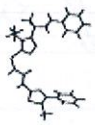
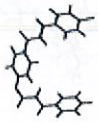
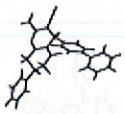
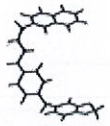
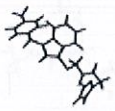
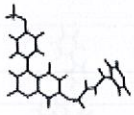
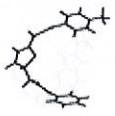
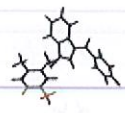
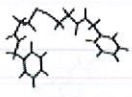
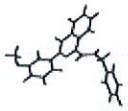
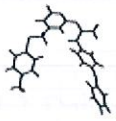
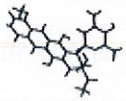
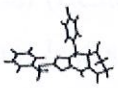
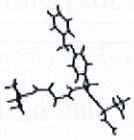
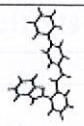
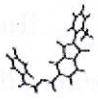
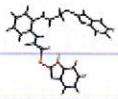
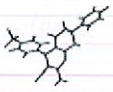
15		ZINC01532171	Isatin Indole -2,3- dione 2,3-Indolinedione	-6.352	-5.232	-21.
16		ZINC03161158	5-nitrobenzene-1,3-dicarboximidamide	-6.035	-5.133	-29.

Table 3.1.2 Database ID number, chemical structure, IUPAC names and fitness scores of top scored ligands docked with General secretory pathway protein L(gspL) of *Pseudomonas aeruginosa* PA7 using docking program Glide .

S. No.	Structure	ZINC ID	IUPAC Name	Glide score	GlideXP score	Gli ene
1		ZINC04014946	phenylmethyl 2-[[2-(phenylmethoxycarbonylamino)-3-[5-(phenylmethoxy)-1H-indol-3-yl] propanoyl]amino]acetate	-7.006	-15.649	-59
2		ZINC03994289	(2R,3S,4S,5R)-1,6-bis[(2,6-difluorophenyl)methoxy]-2,5-bis(phenylmethoxy)hexane-3,4-diol	-7.791	-15.192	-50
3		ZINC03994290	(2R,3S,4S,5R)-1,6-bis[(2-fluorophenyl)methoxy]-2,5-bis[(4fluorophenyl) methoxy]hexane-3,4-diol	-8.733	-14.767	-55
4		ZINC03146249	N-[[5-(2,4-dichlorophenyl)furan-2-yl]methylideneamino]-2-(4-phenylphenoxy)acetamide	-7.565	-13.559	-43
5		ZINC02154335	N-[5-(benzoyl)-2-[[2-(4-methylphenyl)sulfanylacetyl]amino]phenyl]-2-(4-methylphenyl)sulfanylacetamide	-7.706	-13.504	-55

6		ZINC00916256	N-(4-acetylphenyl)-2-[[5-[(3-methylphenoxy)methyl]-4-(phenylmethyl)-1,2,4-triazol-3-yl]sulfanyl]acetamide	-7.019	-13.398	-56
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	-7.654	-13.22	-47
8		ZINC03985979	[4-[3-(4-chlorophenyl)prop-2-enoyl]phenyl] 3-phenylprop-2-enoate	-7.129	-13.108	-43
9		ZINC00728134	2-amino-5-oxo-7-phenyl-4-(4-phenylphenyl)-4,6,7,8-tetrahydrochromene-3-carbonitrile	-7.498	-12.991	-42
10		ZINC02952205	3-[4-[4-(1,3-benzodioxol-5-ylmethyl)piperazin-4-ium-1-yl]-4-oxobutyl]-1H-quinazoline-2,4-dione	-7.859	-12.954	-50
11		ZINC04059724	N-(furan-2-ylmethyl)-2-[1-[(4-methylphenyl)methyl]indol-3-yl]	-7.147	-12.924	-42
12		ZINC04000222	2-[4-(4-methoxyphenyl)-2-oxochromen-7-yl]oxy-N-(pyridin-4-ylmethyl)acetamide	-7.197	-12.895	-43
13		ZINC03592328	N-(4-fluorophenyl)-2-[(2Z,5R)-2-[2-[(4-methylphenyl)amino]-2-oxoethylidene]-4-oxo-1,3-thiazolidin-5-yl]acetamide	-7.985	-12.823	-45
14		ZINC04059664	3-[(2,5-dimethylphenyl)methylsulfanyl]-1-[(4-fluorophenyl)methyl]indole	-7.331	-12.766	-37
15		ZINC04001913	3-[3-oxo-3-(phenylmethylamino)propyl]disulfanyl-N-(phenylmethyl)propanamide	-7.164	-12.736	-47

16		ZINC02941742	N-[(4-fluorophenyl)methyl]-3-(3-methoxyphenyl)-4-oxophthalazine-1-carboxamide	-7.476	-12.732	-41
17		ZINC03869146	N-(6-{[(4-methylphenyl)thio]methyl}-4-oxo-1,4-dihydro-2-pyrimidinyl)-N''-(4-phenoxyphenyl)guanidine	-7.177	-12.685	-51
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	-7.063	-12.666	-46
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	-7.113	-12.652	-38
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	-7.101	-12.63	-5
21		ZINC04031659	N-[2-(1H-benzimidazol-2-yl)phenyl]-4-phenylbenzamide	-7.016	-12.594	-42
22		ZINC00629889	[2-(4-bromophenyl)-2-oxoethyl]2-(2-methylphenyl)-1,3-dioxoisoindole-5-carboxylate	-7.329	-12.538	-49
23		ZINC02809032	2-(1H-benzimidazol-3-ium-2-ylsulfanyl)-N-[2-[[2-(1H-benzimidazol-3-ium-2-ylsulfanyl)acetyl]amino]phenyl]acetamide	-8.335	-12.485	-54
24		ZINC00752690	2-amino-5-oxo-7-phenyl-4-[4-trifluoromethyl)phenyl]-4,6,7,8-	-7.212	-12.476	-37

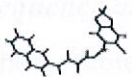
			tetrahydrochromene-3-carbonitrile			
25		ZINC01054177	N-[(6-chloro-1,3-benzodioxol-5-yl)methylideneamino]-2-(naphthalen-2-ylamino)acetamide	-7.136	-12.464	-37

Table 3.1.2 Database ID number, chemical structure, IUPAC names and fitness scores of top scored ligands docked with type IVpilus assembly protein A(pilA) of *Pseudomonas aeruginosa* PA7 using docking program Glide .

3.2 Pseudomonas syringe:

It is a plant pathogen which can infect a wide range of plant species, and exists as over 50 different pathovars, P. syringae pv. syringae strain B728a is the cause of bacterial brown spot of bean (Beattie, G.A., and S.E. Lindow. 1994)It is a highly fit colonist of bean with predominant leaf surface localization (Beattie, G.A., and S.E. Lindow. 1994, Wilson, M., S.S. Hirano, and S.E. Lindow. 1999.)It is also active in ice nucleation at warm subfreezing temperatures (Lindow, S.E. 1993, Gurian-Sherman, D., and S.E. Lindow. 1993) Brown spot is a recurring problem that can cause serious losses in snap beans. Halo blight has caused serious losses in both dry and snap beans. The losses caused by this pathogen in agriculture are urging to call for new and more potent drugs to get rid of the pathogenic hazards. Our study has reveale two very potent targets , drugs against which can help control the hazaroyus effects of the pathogen.

The targets are

1. murD
2. murE

One of the most attractive targets for new antibacterial compounds is the bacterial peptidoglycan biosynthetic pathway. Peptidoglycan is an essential component of the bacterial cell wall. It is responsible for a defined cell shape and preserves cell integrity by compensating internal osmotic pressure. Any perturbation of the multi-step peptidoglycan biosynthesis may lead to cell lysis (Van Heijenoort, J. 2001) The murD and murE enzyme is involved in the biosynthesis of bacterial peptidoglycan in both gram-negative and gram-positive bacteria, therefore, inhibitors of this enzyme should also be broad-spectrum antibiotics.

3.2.1 Materials and methods

3.2.1.1 *Sequence analysis*

The protein sequence of UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE) involved in Lysine biosynthesis and Peptidoglycan biosynthesis. UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase(murD) involved in D-Glutamine and D-glutamate metabolism and also in Peptidoglycan biosynthesis of the organism *Pseudomonas aerugimosa* PA7 was obtained from the NCBI database (accession number PSYR_4107, PSYR_4104 respectively). Even though murE and murD homologues are not known to be present in humans , their sequences were subjected to various sequence analysis programs as explained below to examine their potential as a drug target. Pairwise sequence alignment of murE and murD from *Pseudomonas syringe* was carried out against ESTs of phaseolus vulgaris and arabidopsios in BLAST as well as against the *Pseudomonas putida* without any significant hit. Blastp against PDB databank gave significant hits .Template was selected on the basis of identities, score, gaps and e value. These proteins had given a hit with 4UAG (with murE , score 34.3 , 41% identity, 62% positive scores and 0% gaps with an expect value of 0.074), 1EFH_A (with murD , score 311.2 , 46% identity, 54% positive scores and 4% gaps with an expect value of 2.7e-85)

3.2.1.2 Homology model construction

The three dimensional structure of the protein was constructed using Prime using the same methodology as decribed to *P.aeruginosa* as shown in fig 3.4.

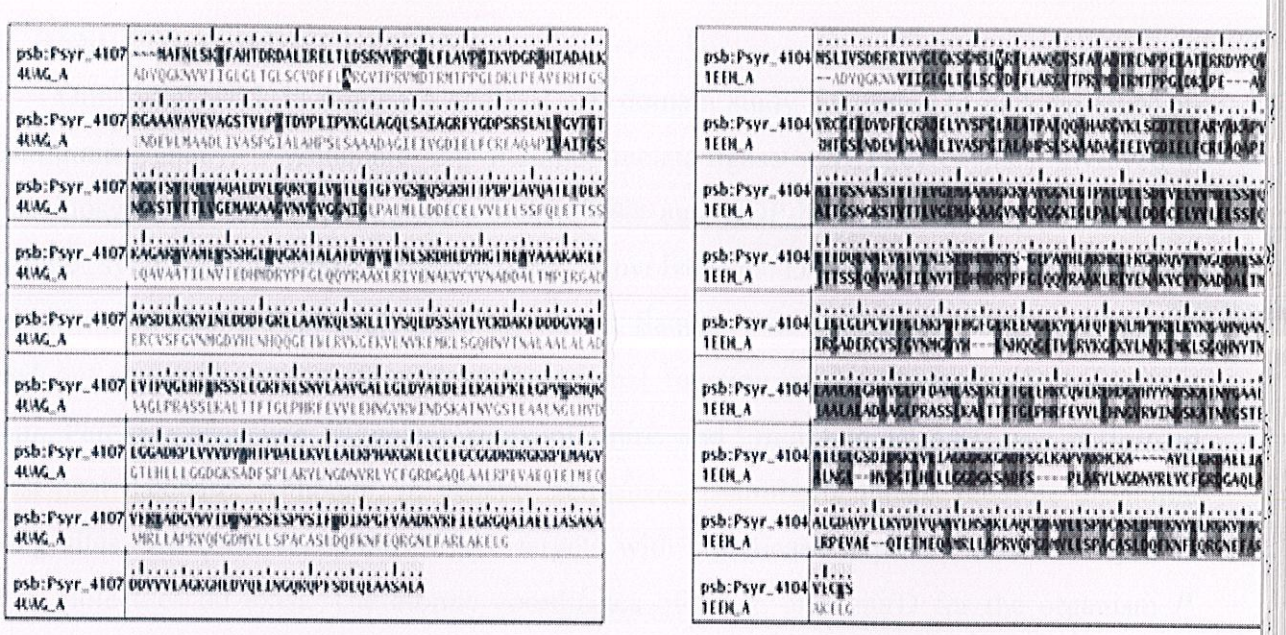


Figure 3.3 : Sequence Alignment of protein murE and murD of *P.syringea* with the templates

(a) (b)

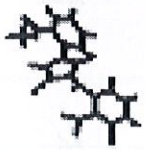
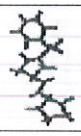
Figure 3.4 Modeled three-dimensional structure of murD(a) and murE(b) *P.syringea*

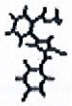
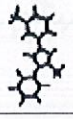
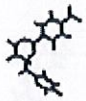
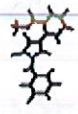
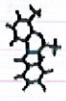
The methods for homology modeling, binding site prediction and virtual screening for designing lead molecules are similar as mentioned above.

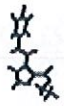
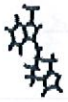
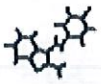
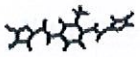
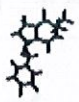
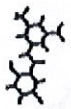
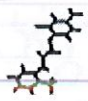
3.2.2 RESULTS AND DISCUSSIONS:

Our initial work in sequence analysis and domain study suggested to pick up UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase(murE) key ezymes involved in Lysine biosynthesis and Peptidoglycan biosynthesis and UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase (murD) key enzyme involved in D-Glutamine and D-glutamate metabolism and also in Peptidoglycan biosynthesis. Sequence as well as structural analysis was carried out to avoid any interference in the normal function of the Phaseolus Vulgaris host system. Pairwise sequence alignment study of murE and murD in BLAST was performed against host database. Although, in general, sequence dissimilarity means possible dissimilar binding sites, however, exceptions where targets with significant sequence dissimilarity bind similar molecules do exist. The atomic coordinates of murE and murD for the organism *P. syringea* were not available in Protein Data Bank, which necessitated developing a protein model. Homology modeling protocol was employed to predict the model of the protein. An identity of 46% with well-studied protein of E.coli (Resolution—1.9) for murD and an identity of 41% with well-studied protein of E.coli ((Resolution—1.6) and provided a great strength for modeling the protein. Three dimensional structure prediction by comparative modeling was done by Prime. Impact Module of schrodinger used in energy minimization embraces a range of force fields. Imapct module is a molecular mechanics simulation environment offering energy minimization, dynamics and conformational search on molecular, aggregate or periodic systems. We used both PROCHECK and the VERIFY3D softwares to check the quality of the modeled protein. We carried out in silico studies to find the active sites, using sitemap and escite algorithm. The output from the sitemap program showed coherent active sites for the target protein as from the ascite. The best ligands shows in 3.2 were chosen based on the scores obtained from the genetic docking algorithm Gilde and the consensus scoring program Glide-XP. The standard methodology accepted for virtual screening with the program Prime 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. Fifty genetic algorithm runs with standard default parameter settings were performed without early termination, after which 10 best solutions were kept for each ligand. 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. The Glide fitness function consisted of four components: (a) protein–ligand hydrogen bond energy (external H-bond); (b) protein–ligand van der Waals (vdw) energy (external vdw); (c)ligand internal vdw energy (internal vdw); (d) ligand torsional strain energy (internal torsion). On the other hand, the scoring schema used in the software Glide-XP computes a binding score for a given protein–ligand complex structure, and this binding score correlates to experimental binding constants well. It takes into account van der Waals interactions, hydrogen bonding, deformation penalty, and hydrophobic effects between the receptor and the ligand. Glide-XP does not perform the molecular docking process and it is applied in combination with docking programs such as Glide. The best 20 ligands as shown in table 3.2 chosen with Glide Fitness score proved their reliability in the Glide-XP. Further in the paper, name of the ligands are indicated by the ligand ID's given in the database. 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.

S.NO	Structure	ZINC ID	IUPAC Name	Glide score	GlideXP score	Glide energy
1		ZINC00999929	methyl 4-methyl-3-[(4,5,6,7-tetrahydro-1-benzothien-3-ylcarbonyl)amino]benzoate	-9.361	-6.268	-33.492
2		ZINC01938126	methyl 2-[3-chloro-4-(3-chloro-2-methylanilino)-2,5-dioxopyrrol-1-yl]benzoate	-9.349	-6.636	-41.03(
3		ZINC01031540	3-[(5-amino-2-chlorophenyl)amino] -1-(2,4,6-trichlorophenyl)-1H-pyrazol-5-ol	-9.172	-6.591	-40.757
4		ZINC00219356	N-cyclopentyl-5-oxo-1-(1-phenylethyl)pyrrolidine-	-9.169	-6.046	-35.368

			3-carboxamide			
5		ZINC00720035	(6R,8R,9R)-6-(2-fluorophenyl)-9-methyl-7-oxo-5,6,8,9,10,11-hexahydrobenzo[c][1,5]benzodiazepine-8-carboxylate	-9.138	-6.061	-35.800
6		ZINC02391940	8-(2-hydroxyethylamino)-1,3-dimethyl-7-(naphthalen-1-ylmethyl)-9H-purin-7-ium-2,6-dione	-9.093	-6.874	-37.143
7		ZINC01317711	(2R)-2-[(3,5-dioxo-2H-1,2,4-triazin-6-yl)amino]propanehydrazide	-9.092	-6.861	-30.211
8		ZINC01938091	methyl 2-[3-chloro-4-(4-chloroanilino)-2,5-dioxopyrrol-1-yl]benzoate	-8.954	-6.674	-40.617
9		ZINC03133798	5-methyl-4-phenyl-2-[3-(trifluoromethyl)phenyl]-1H-pyrazol-3-one	-8.895	-6.889	-33.05
10		ZINC01748535	4-[[3-(3,5-dioxo-2H-1,2,4-triazin-6-yl)propanoylhydrazinylidene]methyl]benzoate	-8.819	-6.385	-31.68
11		ZINC00704941	[5-acetyloxy-2-[(4R)-4,6,6-trimethyl-2-sulfanylidene-1,3-diazinan-4-yl]phenyl] acetate	-8.814	-6.207	-38.32
12		ZINC00753445	2-amino-4-(3,4-difluorophenyl)-5-oxo-4,6-dihydropyrano[3,2-c]quinoline-3-carbonitrile	-8.752	-6.067	-31.89
13		ZINC01938089	methyl 2-[3-chloro-4-(3-chloroanilino)-2,5-dioxopyrrol-1-yl]benzoate	-8.730	-6.501	-41.796
14		ZINC00112329	N-(4-chlorophenyl)-2,8-dimethyl-4-quinolinamine	-8.727	-6.200	-31.59

15		ZINC00608429	ethyl 4-[(3-chlorophenyl)carbamoyl] -1H-imidazole-5-carboxylate	-8.695	-6.672	-39.001
16		ZINC03873275	(1-methylindol-3-yl)-(4,5,6,7-tetrahydro-3H-benzimidazol-5-yl)methanone	-8.666	-6.921	-35.679
17		ZINC00666142	2-amino-N-(4-bromophenyl)-4,5,6, 7-tetrahydro-1-benzothiophene-3-carboxamide	-8.651	-6.114	-33.406
18		ZINC01795351	(6S,7S)-7-(3-hydroxyphenyl)-2-(3-hydroxypropyl)-5-methylidene-6, 7-dihydro-4H-[1,2,4]triazolo[1,5-a]pyrimidine-6-carboxamide	-8.642	-6.645	-38.267
19		ZINC00884397	5-bromo-N-[4-(furan-2-carbonylamino)-3-methylphenyl]furan-2-carboxamide	-8.61	-6.048	-45.31
20		ZINC03839604	8-[(4-fluorophenyl)methylsulfanyl]-3,3-dimethyl-6,7,8, 8a-tetrahydro-2H-pyrrolo[1,2-a]pyrazine-1,4-dione	-8.59	-6.634	-33.766
21		ZINC03986227	N-[3,5-bis(trifluoromethyl)phenyl]-5-chloro-2-hydroxybenzamide	-8.582	-6.214	-36.989
22		ZINC03332308	3H-benzimidazol-1-ium-5-yl(2,3-dihydroindol-1-yl)methanone	-8.579	-6.742	-33.732
23		ZINC00611426	2-(2-bromo-4-formylphenoxy)-N-(2-methylphenyl)acetamide	-8.577	-6.144	-37.897
24		ZINC03157427	N-(3-bromophenyl)-3-(2,2-dichloroethenyl)-2, 2-dimethylcyclopropane-1-carboxamide	-8.544	-6.171	-35.625

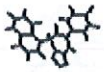
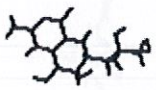
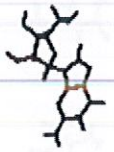
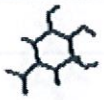
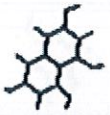
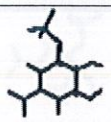
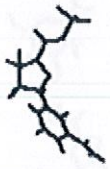
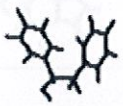
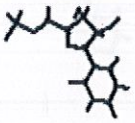
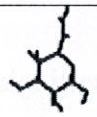
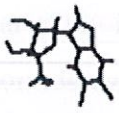
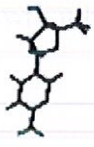
25		ZINC03211261	5-(4-chlorophenyl)-7-naphthalen-1-yl-4,7-dihydrotetrazolo[1,5-a]pyrimidine	-8.520	-6.208	-38.758
28		ZINC03809181	4-nitro-2-(1,1,2,2-tetrafluoroethyl)-6-(trifluoromethyl)-1H-benzimidazole	-8.507	-6.422	-26.889
29		ZINC00562947	2-[(4-hydroxyphenyl)methylamino]-4,5,6,7-tetrahydro-1-benzothiophene-3-carbonitrile	-8.414	-6.288	-35.822

Table 3.2.1 Database ID number, chemical structure, IUPAC names and fitness scores of top scored ligands docked UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase (murD) of *Pseudomonas syringe* B728a using docking program Glide

S.NO	Structure	ZINC ID	IUPAC Name	Glide score	GlideXP score	Glide energy
1		ZINC03869824	1,3,4-trihydroxy-5-oxocyclohexane-1-carboxylic acid	-6.266	-7.504	-30.128
2		ZINC03869200	2-amino-7,8-dihydroxy-6-(1,2,3-trihydroxypropyl)-1,5,6,7-tetrahydropteridin-4-one	-6.258	-7.377	-34.707
3		ZINC03870192	1,3,4,5-tetrahydroxycyclohexanecarboxylic acid	-6.664	-7.37	-30.868
4		ZINC01532637	2-amino-9-[(2S,3S,4R,5S)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]-3H-purin-6-one	-6.134	-7.292	-41.929
5		ZINC03869823	1,3,4-trihydroxy-5-oxocyclohexane-1-carboxylic acid	-6.968	-7.269	-30.559

6		ZINC04095578	(2S,3S,4S,5R,6R)-3,4,5-trihydroxy-6-phosphonooxyoxane-2-carboxylic acid	-6.001	-7.227	-34.292
7		ZINC03870234	3,4,5-trihydroxycyclohexene-1-carboxylic acid	-6.242	-6.995	-29.701
8		ZINC00901903	naphthalene-1,3,6,8-tetrol	-6.471	-6.94	-32.219
9		ZINC03870238	4,5-dihydroxy-3-phosphonooxycyclohexene-1-carboxylic acid	-6.355	-6.922	-35.482
10		ZINC03870107	[5-(3-carbamoylpyridin-1-ium-1-yl)-3,4-dihydroxyoxolan-2-yl]methyl dihydrogen phosphate	-6.055	-6.902	-41.116
11		ZINC03852703	1,2-bis(4-fluorophenyl)ethane-1,2-diol	-6.419	-6.779	-34.246
12		ZINC03870178	[5-(2,4-dioxo-1H pyrimidin-5-yl)-3,4-dihydroxyoxolan-2 yl] Methyl dihydrogen phosphate	-6.471	-6.753	-40.723
13		ZINC03872700	3,4,5,6-tetrahydroxyoxane-2-carboxamide	-6.499	-6.740	-33.253
14		ZINC03872698	3,4,5,6-tetrahydroxyoxane-2-carboxamide	-6.375	-6.666	-33.701
15		ZINC03873418	(2R,3R,4S,5R)-2-(5,6-dichlorobenzimidazol-1-yl)-5-(hydroxymethyl)oxolane-3,4-diol	-6.478	-6.623	-36.234
16		ZINC03800994	4-amino-1-[(2R,3S,4S,5S)-5-(chloromethyl)-3,4-dihydroxyoxolan-2-yl]pyrimidin-2-one	-6.391	-6.553	-36.153

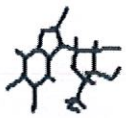
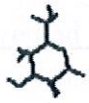
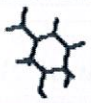
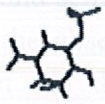
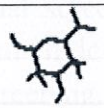
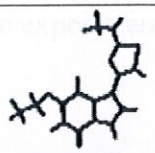
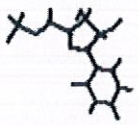
17		ZINC03797914	(2R,3S,4R,5R)-2-(hydroxymethyl)-5-(2,5,6-trichlorobenzimidazol-1-yl)oxolane-3,4-diol	-6.0632	-6.48	-34.601
18		ZINC03869937	3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-one	-6.01	-6.464	-25.413
19		ZINC03869253	3,4-dihydroxy-5-oxocyclohexene-1-carboxylic acid	-6.273	-6.463	-28.35
20		ZINC03870239	4,5-dihydroxy-3-phosphonooxycyclohexene-1-carboxylic acid	-6.446	-6.456	-34.710
21		ZINC03869799	3,4,5,6-tetrahydroxyoxane-2-carboxylic acid	-6.588	-6.429	-31.449
22		ZINC02330112	4-(5-ethoxy-1H-indol-3-yl)-N-methyl-1,3-thiazol-2-amine	-6.184	-6.418	-35.533
		ZINC03870178	[5-(2,4-dioxo-1H pyrimidin-5-yl)-3,4-dihydroxyoxolan-2 yl] Methyl dihydrogen phosphate	-6.471	-6.753	-40.723

Table 3.2.2 Database ID number, chemical structure, IUPAC names and fitness scores of top scored ligands docked with UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE) of *Pseudomonas syringe* B728a using docking program Glide .

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 or pilA in *P. aeruginosa* . We propose that a similar study in silico could be designed for the pathogens within, reated to *P. aeruginosa* with high confidence interval, to assess the feasibility of using the inhibitors defined here against their dapD. Further, a detailed experimental study would be interesting to validate the utility of the proposed growth inhibitors as drug molecules for the human pathogen, *P. aeruginosa* .

Conclusion

With more and more bacterial strains becoming resistant to single or multiple antibiotics there is an ever increasing urge to find and look for new targets and development of drugs against them to escape the hazardous effect of these strains.In this work, we propose probable chemical compounds, which could be tested to devise drug molecules to retard the hazardous proliferation of *P. aeruginosa* and *P. syringea* outbreaks in various parts of the world. Extensive literature study and sequence analyses were carried out to project the need to find new potential drug target. In the absence of crystal structure, we used homology-modeling protocol , using prime,to predict the three dimensional structure of the protein with very good template structures. Virtual screening was carried out, using Glide program, followed by a consensus scoring algorithm Glide-XP. 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.

REFERENCES

1. D.B. Allison. Statistical methods for microarray research for drug target identification 2002. *Proceedings of the American Statistical Association*, Biopharmaceutical Section Alexandria, (VA): American Statistical Association, 2002
2. Pier GB, Grout M, Zaidi TS, Olsen JC, Johnson LG, Yankaskas JR, Goldberg JB. Role of mutant CFTR in hypersusceptibility of cystic fibrosis patients to lung infections. Science. 1996;271:64–67.
3. Britigan BE, Rasmussen GT, Cox CD. Augmentation of oxidant injury to human pulmonary epithelial cells by the Pseudomonas aeruginosa siderophore pyochelin. Infect Immun. 1997;63:1071–1076. [
4. The Microbial World ,University of Wisconsin – Madison, *Pseudomonas aeruginosa* Opportunistic Infections,© 2008 Kenneth Todar University of Wisconsin-Madison Department of Bacteriology
5. Beattie, G.A., and S.E. Lindow. 1994. Comparison of the behavior of epiphytic fitness mutants of Pseudomonas syringae under controlled and field conditions. Appl. Environ. Microbiol. 60:3799-3808
6. Wilson, M., S.S. Hirano, and S.E. Lindow. 1999. Location and survival of leaf-associated bacteria in relation to pathogenicity and potential for growth within the leaf. Appl. Environ. Microbiol. 65:1435-1443
7. Lindow, S.E. 1993. Novel method for identifying bacterial mutants with reduced epiphytic fitness. Appl. Environ. Microbiol. 59:1586-1592.,
8. Gurian-Sherman, D., and S.E. Lindow. 1993. Bacterial ice nucleation: significance and molecular basis. FASEB Journal 9:1338-1343.
9. Newton C.M. Gomes, Irina A. Kosheleva, Wolf-Rainer Abraham, Kornelia Smalla (2005) Effects of the inoculant strain *Pseudomonas putida* KT2442 (pNF142) and of naphthalene contamination on the soil bacterial community. FEMS Microbiology Ecology 54 (1), 21–33
10. Immortal Polystyrene Foam Meets its Enemy | LiveScience
11. Ward PG, Goff M, Donner M, Kaminsky W, O'Connor KE. (2006) A two step chemo-biotechnological conversion of polystyrene to a biodegradable thermoplastic. Environmental Science and Technology 40(7):2433-7. PMID 16649270

12. XinCai Chen, JiYan Shi, YingXu Chen, XiangHua Xu, ShengYou Xu, and YuanPeng Wang Can. J. Tolerance and biosorption of copper and zinc by *Pseudomonas putida* CZ1 isolated from metal-polluted soil Microbiol. 52(4): 308–316 (2006)
13. Amer GA, Utkhede RS. (2000) Development of formulations of biological agents for management of root rot of lettuce and cucumber. Can J Microbiol. 46(9):809-16. PMID 11006841
14. Validov S, Kamilova F, Qi S, Stephan D, Wang JJ, Makarova N, Lugtenberg B. (2007) Selection of bacteria able to control *Fusarium oxysporum* f. sp. *radicis-lycopersici* in stonewool substrate. J Appl Microbiol. 102(2):461-71
15. Bruccoleri, R. E., Dougherty, T. J. and Davison, D. B. (1998). Concordance analysis of microbial genomes. Nucleic Acids Res. 26, 4482-4486.
16. T-iDT: Tool for Identification of Drug Target in Bacteria and Validation by Mycobacterium Tuberculosis, Nitesh Kumar Singh[1], S. Mahalaxmi Selvam[1], Paulsharma Chakravarthy[2],1386-6338,ISB vol 6,number6/2006
17. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.
18. Galperin, M. Y. and Koonin, E. V. (1999). Searching for drug targets in microbial genomes. Curr. Opin. Biotechnol. 10, 571-578.
19. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P. et al. (2003) Essential Bacillus subtilis genes. Proc. Natl Acad. Sci. USA, 100, 4678–4683…..
20. Itaya,M. (1995) An estimation of minimal genome size required for life. FEBS Lett., 362, 257–260
21. Judson,N. and Mekalanos,J.J. (2000) TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. Nat. Biotechnol., 18, 740–745
22. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW. 2004. Potential of metabolomics as a functional genomics tool. *Trends in Plant Science 9*: 418-425
23. Oliver SG, Winson MK, Kell DB, Baganz F. 1998. Systematic functional analysis of the yeast genome. *Trends in Biotechnology 16*: 373-378.). Despite its relative youth (in comparison to genomics and proteomics

24. Wagner C, Sefkow M, Kopka J. 2003. Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry 62*: 887-900
25. Zhang, R., Ou, H. Y. and Zhang, C. T. (2004). DEG: a database of essential genes. Nucleic Acids Res. 32, D271-D272.
26. Sakharkar, K. R., Sakharkar, M. K. and Chow, V. T. K. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. In Silico Biol. 4, 0028.
27. Dutta, A., Singh, S. K., Ghosh, P., Mukherjee, R., Mitter, S. and Bandyopadhyay, D. (2006). *In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. In Silico Biol. 6, 0005
28. Galperin MY, Koonin EV,National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. galperin@ncbi.nlm.nih.gov, Curr Opin Biotechnol. 1999 Dec;10(6):571-8
29. Acyl group specificity at the active site of tetrahydridipicolinate N-succinyltransferase, Todd W. Beaman,[1] Kurt W. Vogel,[2,4] Dale G. Drueckhammer,[3] John S. Blanchard,[1] and Steven L. Roderick[1] Protein Sci. 2002 April; 11(4): 974–979.
30. Mini Rev Med Chem. 2003 Mar;3(2):115-27 Inhibitors of lysine biosynthesis as antibacterial agents Hutton CA, Southwood TJ, Turner JJ
31. Dmitry A. Rodionov,* Alexey G. Vitreschak,[1] Andrey A. Mironov, and Mikhail S. Gelfand "Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch" 2003 December 1; 31(23): 6748–6757
32. Vincent T. Lee,[1] Roger S. Smith,[1] Burkhard Tümmler,[2] and Stephen Lory[1] Activities of *Pseudomonas aeruginosa* Effectors Secreted by the Type III Secretion System In Vitro and during Infection
33. Bally, M., Filloux, A., Akrim, M., Ball, G., Lazdunski, A., and Tommassen, J. 1992. Protein secretion in *Pseudomonas aeruginosa*: Characterization of seven *xcp* genes and processing of secretory apparatus components by prepilin peptidase. *Mol. Microbiol.* 6: 1121–1131.
34. Voulhoux, R., Ball, G., Ize, B., Vasil, M.L., Lazdunski, A., Wu, L.F., and Filloux, A. 2001. Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J.* 20: 6735–6741.

35. Pugsley, A. P. 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. 57:50-108.

36. O'Toole G.A. and Kolter,R. (1998) Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol. Microbiol.*, 30:, 295–304.

37. Soto G.E. and Hultgren,S.J. (1999) Bacterial adhesins: common themes and variations in architecture and assembly. *J. Bacteriol.*, 181:, 1059–1071.

38. Lowe M.A., Holt,S.C. and Eisenstein,B.I. (1987) Immunoelectron microscopic analysis of elongation of type 1 fimbriae in Escherichia coli.*J. Bacteriol.*, 169:, 157–163

39. Clewell D.B. (1993) *Bacterial Conjugation*. Plenum Press, New York, NY.

40. Wall D. and Kaiser,D. (1999) Type IV pili and cell motility. *Mol. Microbiol.*, 32:, 1–10

41. Christie P.J. (1997) Agrobacterium tumefaciens T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. *J. Bacteriol.*, 179:, 3085–3094

42. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist Jeffrey B. Lyczaka,b, Carolyn L. Cannonb,c, Gerald B. Piera, Microbes and Infection, 2, 2000, 1051–1060)

43. Surface plasmon resonance shows that type IV pili are important in surface attachment by Pseudomonas aeruginosa A. Toby A. Jenkins, Angus Buckling, Marsha McGhee and Richard H. ffrench-Constant , J. R. Soc. Interface (2005) 2, 255–259

44. den Blaauwen T, de Pedro MA, Nguyen-Distèche M, Ayala JA. Morphogenesis of rod-shaped sacculi. FEMS Microbiol. Rev. (2008) 32:321–344

45. Dramsi S, Magnet S, Davison S, Arthur M. Covalent attachment of proteins to peptidoglycan. FEMS Microbiol. Rev. (2008) 32:307–320.

46. Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiol. Rev. (2008) 32:287–306

47. van Heijenoort, J. Recent advances in the formation of the bacterial peptidoglycan monomer unit.*Nat. Prod. Rep.* 2001, *18*, 503–519

48. El Zoeiby, A.; Sanschagrin, F.; Levesque, R.C. Structure and function of the Mur enzymes: development of novel inhibitors. *Mol. Microbiol.* 2003, *47*, 1-12

49. New targets for antibiotic development: biogenesis of surface adherence structures, C. Hal Jones and Dennis E. Hruby, Drug Discovery Today, Volume 3, Issue 11, 1998

50. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325

51. Crystal structure of UD{-N-acetylmuramoyl-L-alanine : D glutamate ligase from E. coli, Jay A. Bertrand ,Geneviève Auger Eric Fanchon, Lydie Martin, Didier Blanot, Jean van Heijenoort and Otto Dideberg *The EMBO Journal* (1997) 16, 3416–3425

APPENDIX

TABLE 4.1 : RESULTS OF BLAST OF *Pseudomonas aeruginosa* PA7 WITH DEG.

DEG a	GENE ID	NAME	DEG ID	SCORE	PATHWAY
	lipid metabolism				
1	PSPA7_1311	ispG	DEG10040387	84	Biosynthesis of steroids
2	PSPA7_1500	accA	DEG10040052	68	Fatty acid biosynthesis
3	PSPA7_3490		DEG10040227	66	Fatty acid biosynthesis
4	PSPA7_3664	fabA	DEG10040157	94	Fatty acid biosynthesis
5	PSPA7_3665	fabB	DEG10040368	82	Fatty acid biosynthesis
6	PSPA7_5268	accC2	DEG10100527	62	Fatty acid biosynthesis
	carbohydrate metabolism				
1	PSPA7_0531	tkt	DEG10040447	192	Pentose phosphate pathway
2	PSPA7_3729	leuA2	DEG10100510	82	Fatty acid metabolism
3	PSPA7_1523		DEG10100384	78	Pyruvate metabolism
4	PSPA7_3730		DEG10100509	80	Bile acid biosynthesis
5	PSPA7_1500	accA	DEG10040052	68	Pyruvate metabolism
4	PSPA7_1500	accA	DEG10040052	68	Propanoate metabolism
5	PSPA7_1504	eno	DEG10110158	176	Glycolysis / Gluconeogenesis
6	PSPA7_2117	leuA1	DEG10100584	72	Pyruvate metabolism
	nucleotide metabolism				
1	PSPA7_0436	ihyA	DEG10040433	62	Pyrimidine metabolism
2	PSPA7_2381	lcd	DEG10110073	94	Citrate cycle (TCA cycle)
3	PSPA7_0830	rpoB	DEG10040577	363	Purine metabolism
4	PSPA7_3688	sucB	DEG10040126	178	Citrate cycle (TCA cycle)
5	PSPA7_0830	rpoB	DEG10040577	363	Pyrimidine metabolism
6	PSPA7_3689	sucA	DEG10040125	172	Citrate cycle (TCA cycle)
7	PSPA7_0831	rpoC	DEG10110213	472	Purine metabolism
8	PSPA7_3768	acnA	DEG10100241	147	Citrate cycle (TCA cycle)
9	PSPA7_0831	rpoC	DEG10110213	472	Pyrimidine metabolism
10	PSPA7_3768	acnA	DEG10100241	147	Glyoxylate and dicarboxylate metabolism
11	PSPA7_0862	rpoA	DEG10110177	90	Purine metabolism
12	PSPA7_4724	acnD	DEG10100247	66	Citrate cycle (TCA cycle)
13	PSPA7_0862	rpoA	DEG10110177	90	Pyrimidine metabolism
14	PSPA7_4724	acnD	DEG10100247	66	Glyoxylate and dicarboxylate metabolism
15	PSPA7_1326	guaB	DEG10100526	60	Purine metabolism
16	PSPA7_3681	fumC	DEG10100163	67	Citrate cycle (TCA cycle)
17	PSPA7_1327	guaA	DEG10100534	94	Purine metabolism
18	PSPA7_5119	pgi	DEG10100140	72	Glycolysis / Gluconeogenesis
19	PSPA7_1485	pyrH	DEG10110017	68	Pyrimidine metabolism
20	PSPA7_5119	pgi	DEG10100140	72	Pentose phosphate pathway
21	PSPA7_1498	dnaE	DEG10040051	193	Purine metabolism
22	PSPA7_5119	pgi	DEG10100140	72	Starch and sucrose metabolism
23	PSPA7_1499	dnaE	DEG10040051	193	Pyrimidine metabolism
24	PSPA7_5368	accC2	DEG10100527	62	Pyruvate metabolism
25	PSPA7_1502	pyrG	DEG10100294	111	Pyrimidine metabolism
26	PSPA7_5368	accC2	DEG10100527	62	Propanoate metabolism
27	PSPA7_1966	cmk	DEG10100298	188	Pyrimidine metabolism
28	PSPA7_5729		DEG10100510	82	Propanoate metabolism
29	PSPA7_3801		DEG10040089	82	Purine metabolism
30	PSPA7_3732	aceF	DEG10040028	84	Glycolysis / Gluconeogenesis
31	PSPA7_3801		DEG10040089	82	Pyrimidine metabolism
32	PSPA7_3732	aceF	DEG10040028	84	Pyruvate metabolism
33	PSPA7_4224		DEG10130065	80	Purine metabolism
34	PSPA7_3902	rfbA	DEG10100033	78	Nucleotide sugars metabolism
35	PSPA7_4224		DEG10130065	80	Pyrimidine metabolism
36	PSPA7_6350	glmS	DEG10040531	78	Aminosugars metabolism
37	PSPA7_4225	nrdB	DEG10130064	115	Purine metabolism
38	PSPA7_4225	nrdB	DEG10130064	115	Pyrimidine metabolism
	energy metabolism				
1	PSPA7_5014	cysN	DEG10100190	111	Purine metabolism
2	PSPA7_0781	ctaD	DEG10100408	64	Oxidative phosphorylation
3	PSPA7_0437	tkt	DEG10040473	192	Carbon fixation
4	PSPA7_5157	pnp	DEG10040475	127	Purine metabolism
5	PSPA7_3671	ppa	DEG10040601	72	Pyrimidine metabolism
6	PSPA7_5476	carB	DEG10130216	109	Oxidative phosphorylation
7	PSPA7_1233	narH	DEG10110163	377	Pyrimidine metabolism
8	PSPA7_3389	ureC	DEG10100302	76	Nitrogen metabolism
9	PSPA7_2327		DEG10100033	92	Purine metabolism
10	PSPA7_3667	purA	DEG10100041	76	Carbon fixation
11	PSPA7_2381	lcd	DEG10110073	94	Purine metabolism
12	PSPA7_6087	dut	DEG10040545	68	Reductive carboxylate cycle (CO2 fixation)
13	PSPA7_2807	glyA2	DEG10040327	94	Pyrimidine metabolism
14	PSPA7_6293	polA	DEG10100274	70	Methane metabolism
15	PSPA7_3768	acnA	DEG10100241	147	Purine metabolism
16	PSPA7_6293	polA	DEG10100274	70	Reductive carboxylate cycle (CO2 fixation)
17	PSPA7_3847	ccmF	DEG10030396	62	Pyrimidine metabolism
					Metabolism; Energy Metabolism; Nitrogen metabolism
18	PSPA7_4724	acnD	DEG10100247	66	Reductive carboxylate cycle (CO2 fixation)
	amino acid metabolism				
1	PSPA7_5014	cysN	DEG10100190	111	Sulfur metabolism
2	PSPA7_0009	glyQ	DEG10040536	254	Glycine, serine and threonine metabolism
3	PSPA7_3681	fumC1	DEG10100163	67	Reductive carboxylate cycle (CO2 fixation)
4	PSPA7_0532	ahcY	DEG10100518	78	Methionine metabolism
5	PSPA7_2743	glyA3	DEG10040397	157	Methane metabolism
6	PSPA7_0649	hmgK	DEG10100226	149	Methionine metabolism
7	PSPA7_5773	gltB2	DEG10100609	72	Nitrogen metabolism
8	PSPA7_1279	valS	DEG10040605	107	Valine, leucine and isoleucine biosynthesis
9	PSPA7_5837	nuoH1	DEG10110078	161	Nitrogen metabolism
10	PSPA7_1523	leuA2	DEG10100384	78	Valine, leucine and isoleucine biosynthesis
11	PSPA7_5835	glnA	DEG10110208	180	Nitrogen metabolism
12	PSPA7_1327	guaA	DEG10100534	94	Glutathate metabolism
13	PSPA7_6203	glyA1	DEG10040397	113	Methane metabolism
14	PSPA7_1438	lysS	DEG10110164	94	Lysine biosynthesis
15	PSPA7_6356	atpD	DEG10100207	133	Oxidative phosphorylation
16	PSPA7_1473	dapD	DEG10100177	64	Lysine biosynthesis
17	PSPA7_6358	atpA	DEG10130015	117	Oxidative phosphorylation

9	PSPA7_0556	metK	DEG10100067	72		Methionine metabolism
10	PSPA7_1002	mfd	DEG10100126	78	90	Pyrimidine metabolism
11	PSPA7_2015	leuB	DEG10100480	62		Valine, leucine and isoleucine biosynthesis
12	PSPA7_2016	serA	DEG10040525	62		Glycine, serine and threonine metabolism
13	PSPA7_2036	dapA	DEG10040533	62	62	Lysine biosynthesis
14	PSPA7_2057	ilvA1	DEG10100284	78	72	Valine, leucine and isoleucine biosynthesis
15	PSPA7_2263	hemL	DEG10100046	101	76	Porphyrin and chlorophyll metabolism
16	PSPA7_2807	glyA2	DEG10040397	94	92	One carbon pool by folate
17	PSPA7_2823	glyA3	DEG10040397	157	92	One carbon pool by folate
18	PSPA7_2908	ilvC	DEG10100085	113	92	Pantothenate and CoA biosynthesis
19	PSPA7_2922	thiC	DEG10100055	234	92	Thiamine metabolism
20	PSPA7_2971	hemE	DEG10100086	60	92	Porphyrin and chlorophyll metabolism
21	PSPA7_2982	ubiE	DEG10100063	68	92	Ubiquinone biosynthesis
22	PSPA7_2984	ubiB	DEG10100063	90	92	Ubiquinone biosynthesis
23	PSPA7_3008	glyS1	DEG10040398	115	157	One carbon pool by folate
24	PSPA7_2805	gcvP2	DEG10100306	155		Glycine, serine and threonine metabolism
25	PSPA7_2807	glyA2	DEG10040397	94		Glycine, serine and threonine metabolism
26	PSPA7_3508	cysS	DEG10040036	92	64	Novobiocin biosynthesis
27	PSPA7_3509	glnS	DEG10040038	92	62	Alkaloid biosynthesis I
28	PSPA7_3008	sfcA	DEG10040326	78	78	Lysine degradation
29	PSPA7_3689	sucA	DEG10040125	172		Lysine degradation
30	PSPA7_3689	kauB	DEG10040125	172		Tryptophan metabolism
31	PSPA7_4546	gspS	DEG10130056	60	68	D-Arginine and D-ornithine metabolism
32	PSPA7_4552	gnaA	DEG10040054	64	90	Drug metabolism - other enzymes
33	PSPA7_4682	alaC	DEG10100367	76	33	Alanine degradation
34	PSPA7_5020	aruC	DEG10100289	80	72	Urea cycle and metabolism of amino groups
35	PSPA7_5090	gatA	DEG10100489	80	84	1- and 2-Methylnaphthalene degradation
36	PSPA7_5097	gatA	DEG10100487	84		Alanine and aspartate metabolism
37	PSPA7_5243	glyA3	DEG10040397	157		Glycine, serine and threonine metabolism
38	PSPA7_0408	rpoH	DEG10100583	165	113	Transcription factors
39	PSPA7_0776	rpoD	DEG10130260	377	109	Glutamate metabolism
40	PSPA7_0880	rpoB	DEG10040307	363	76	RNA polymerase
41	PSPA7_0881	rpoC	DEG10100243	472	76	Alanine and aspartate metabolism
42	PSPA7_0882	poA	DEG10100570	90	82	Valine, leucine and isoleucine degradation
43	PSPA7_5752	aceF	DEG10040028	64		Alanine and aspartate metabolism
44	PSPA7_5773	gltB	DEG10100607	72		Glutamate metabolism
45	PSPA7_0788	glyQ	DEG10040326	254	84	Aminoacyl-tRNA biosynthesis
46	PSPA7_0826	hpt1	DEG10100089	94	61	Histidine metabolism
47	PSPA7_0832	glnA	DEG10040208	196	80	Glutamate metabolism
48	PSPA7_0838	gcvP1	DEG10100306	74	86	Glycine, serine and threonine metabolism
49	PSPA7_0206	glyA1	DEG10040397	139	115	Glycine, serine and threonine metabolism
6	PSPA7_0837	rplC	DEG10110189	72		Ribosome
507	PSPA7_0840	glnS	DEG10040509	125	78	Glutamate metabolism
8	PSPA7_0844	rplP	DEG10120060	62		Ribosome
9	PSPA7_0847	rpsC	DEG10040506	64		Ribosome
10	PSPA7_0849	rplE	DEG10100588	60	78	Selenoamino acid metabolism
11	PSPA7_0852	rplK	DEG10100284	78	49	Selenoamino acid metabolism
12	PSPA7_0854	rpsG	DEG10040506	100	90	Selenoamino acid metabolism
13	PSPA7_0859	rpsM	DEG10040497	168	94	Glutathione metabolism
14	PSPA7_0807	glyA2	DEG10040398	72	94	Selenoamino acid metabolism
15	PSPA7_4289	murC	DEG10040008	107	76	D-Glutamine and D-glutamate metabolism
16	PSPA7_5018	cysS	DEG10100100	94	111	Aminoacyl-tRNA biosynthesis
17	PSPA7_1483	glyA3	DEG10040399	149	157	Selenoamino acid metabolism
18	PSPA7_1629	metG	DEG10100520	90	82	Aminoacyl-tRNA biosynthesis
19	PSPA7_6304	glyS1	DEG10040398	157	115	Aminoacyl-tRNA biosynthesis
20	PSPA7_2507	rplT	DEG10040292	101		Ribosome
21	PSPA7_1698	gltX	DEG10040100	64		Aminoacyl-tRNA biosynthesis
22	PSPA7_3869	glnA	DEG10040208	62	80	Aminoacyl-tRNA biosynthesis
23	PSPA7_4545	aspS	DEG10130053	68		Aminoacyl-tRNA biosynthesis

Biosynthesis of Polyketides and non-ribosomal peptides

metabolism of cofactors and vitamins

metabolism of secondary metabolism

xenobiotics biodgradation and metabolism

transcription

translation

metabolism of other amino acids

glycan biosynthesis and metabolism

24	PSPA7_4552	proS	DEG10040054	90	Aminoacyl-tRNA biosynthesis
25	PSPA7_4612	alaS	DEG10110153	133	Aminoacyl-tRNA biosynthesis
26	PSPA7_5004	rpsI	DEG10040486	76	Ribosome
27	PSPA7_5661	rpsR	DEG10050172	58	Ribosome
28	PSPA7_5662		DEG10040597	113	Ribosome
29	PSPA7_5788	argS	DEG10040323	84	Aminoacyl-tRNA biosynthesis
	folding,Sorting and Degradation				
1	PSPA7_0468		DEG10110193	66	Protein export
2	PSPA7_0857	secY	DEG10110180	88	Protein export
3	PSPA7_1373	ffh	DEG10040412	167	Protein export
4	PSPA7_4751	lepB1	DEG10040403	60	Protein export
5	PSPA7_4974	secA	DEG10100514	94	Protein export
	Replication and repair				
1	PSPA7_1499	dnaE	DEG10040051	192	DNA replication
2	PSPA7_1499	dnaE	DEG10040051	192	Mismatch repair
3	PSPA7_1499	dnaE	DEG10040051	192	Homologous recombination
4	PSPA7_1519	mutS	DEG10110156	66	Mismatch repair
5	PSPA7_3801		DEG10040089	82	DNA replication
6	PSPA7_3801		DEG10040089	82	Mismatch repair
7	PSPA7_3801		DEG10040089	82	Homologous recombination
8	PSPA7_5658	dnaB1	DEG10040586	78	DNA replication
9	PSPA7_6293	polA	DEG10100274	70	DNA replication
10	PSPA7_6293	polA	DEG10100274	70	Base excision repair
11	PSPA7_6293	polA	DEG10100274	70	Nucleotide excision repair
12	PSPA7_6293	polA	DEG10100274	70	PATH: pap03440 Homologous recombination
	membrane transport				
1	PSPA7_0371	cysA	DEG10100381	78	ABC transporters
2	PSPA7_0526	mexB	DEG10110035	339	
3	PSPA7_1170		DEG10100381	72	ABC transporters
4	PSPA7_2922		DEG10100446	96	ABC transporters
5	PSPA7_4046		DEG10040115	105	ABC transporters
	signal transduction				
1	PSPA7_1235	narH	DEG10110103	377	Two-component system
2	PSPA7_5853	glnA	DEG10110208	180	Two-component system

TABLE 4.2: RESULTS OF BLAST OF *Pseudomonas aeruginosa* PA7 WITH *Pseudomonas putida* KT2440

CARBOHYDRATE METABOLISM				
1	PSPA7_5084	fumC1	hypothetical protein	Citrate cycle (TCA cycle)
2	PSPA7_5568	accC2	acetyl-CoA carboxylase biotin carboxylase subunit	Pyruvate metabolism
3	PSPA7_5729		putative acyl-CoA dehydrogenase	Propanoate metabolism
4	PSPA7_5752	aceF	dihydrolipoamide acetyltransferase	Glycolysis / Gluconeogenesis
5	PSPA7_6350	glmS	D-fructose-6-phosphate amidotransferase	Aminosugars metabolism
ENERGY METABOLISM				
1	PSPA7_0181	ctaD	cytochrome c oxidase, subunit I	Oxidative phosphorylation
2	PSPA7_1235	narH	nitrate reductase, beta subunit	Nitrogen metabolism
3	PSPA7_2327		aminotransferase AlaT	Carbon fixation
4	PSPA7_3847	ccmF	cytochrome C-type biogenesis	Nitrogen metabolism
5	PSPA7_5773	gltB	glutamate synthase subunit alpha	Nitrogen metabolism
LIPID METABOLISM				
1	PSPA7_3490		NADH-dependent enoyl-ACP reductase	Fatty acid biosynthesis
2	PSPA7_5730		putative acyl-CoA dehydrogenase	Fatty acid metabolism

NUCLEOTIDE METABOLISM				
1	PSPA7_1346	guaB	inositol-5-monophosphate dehydrogenase	Purine metabolism
2	PSPA7_1347	guaA	bifunctional GMP synthase	Purine metabolism
3	PSPA7_1966	cmk	cytidylate kinase	Pyrimidine metabolism
4	PSPA7_6293	polA	DNA polymerase I	Purine metabolism
AMINO ACID METABOLISM				
1	PSPA7_0009	glyQ	glycyl-tRNA synthetase subunit alpha	Glycine, serine and threonine metabolism
2	PSPA7_0532	ahcY	S-adenosyl-L-homocysteine hydrolase	Methionine metabolism
3	PSPA7_1473	dapD	tetrahydrodipicolinate succinylase	Lysine biosynthesis
4	PSPA7_1556	metE	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	Methionine metabolism
5	PSPA7_2015	leuB	3-isopropylmalate dehydrogenase	Valine, leucine and isoleucine biosynthesis
6	PSPA7_2263		putative biotin carboxylase/biotin carboxyl carrier protein	Valine, leucine and isoleucine degradation
7	PSPA7_3508	cysS	cysteinyl-tRNA synthetase	Cysteine metabolism
8	PSPA7_5731		putative acyl-CoA dehydrogenase	Valine, leucine and isoleucine degradation
OTHER AMINO ACID METABOLISM				
1	PSPA7_5732		putative acyl-CoA dehydrogenase	beta-Alanine metabolism
METABOLISM OF COFACTORS AND VITAMINS				
1	PSPA7_0436	thya	thymidylate synthase	One carbon pool by folate
2	PSPA7_5802	ubiE	ubiquinone/menaquinone biosynthesis methyltransferase	Ubiquinone biosynthesis
3	PSPA7_5804	ubiB	2-polyprenylphenol 6-hydroxylase	Ubiquinone biosynthesis
XENOBIOTICS BIODEGRADATION AND METABOLISM				
1	PSPA7_1347	guaA	Bifunctional gmp synthase	Drug metabolism - other enzymes
TRANSLATION				
1	PSPA7_2507	rplT	50S ribosomal protein L20	Aminoacyl-tRNA biosynthesis
2	PSPA7_3509	cysS	cysteinyl-tRNA synthetase	Aminoacyl-tRNA biosynthesis
REPLICATION AND REPAIR				
1	PSPA7_0010	tag	Dna-3-methyladenine glycosylase1	Base excision repair
2	PSPA7_2922		maltose/mannitol ABC transporter ATP-binding protein	ABC transporters
SIGNAL TRANSDUCTION				
1	PSPA7_1236		nitrate reductase1,delta chain	Two-component system

TABLE 4.3: RESULTS OF BLAST OF *Pseudomonas aeruginosa* PA7 WITH HOST(homo sapiens)

	GENE ID	GENE NAME	DEG ID	SCORE	PATHWAY
	CARBOHYDRATE METABOLISM				
1	PSPA7_1323	leuA2	DEG10100584	78	Pyruvate metabolism
2	PSPA7_1500	accA	DEG10040052	68	Pyruvate metabolism
3	PSPA7_1500	accA	DEG10040052	68	Propanoate metabolism
4	PSPA7_2117	leuA1	DEG10100584	72	Pyruvate metabolism
	ENERGY METABOLISM				
1	PSPA7_1235	narH	DEG10110103	377	Nitrogen metabolism
2	PSPA7_3847	ccmF	DEG10050396	62	Nitrogen metabolism
	LIPID METABOLISM				

1	PSPA7_1311	ispG	DEG10040387	84	Biosynthesis of steroids
2	PSPA7_1500	accA	DEG10040052	68	Fatty acid biosynthesis
3	PSPA7_3664	fabA	DEG10040157	94	Fatty acid biosynthesis
	NUCLEOTIDE METABOLISM				
1	PSPA7_0862	rpoA	DEG10110177	90	Purine metabolism
2	PSPA7_0862	rpoA	DEG10110177	90	Pyrimidine metabolism
3	PSPA7_1485	pyrH	DEG10110017	68	Pyrimidine metabolism
4	PSPA7_1499	dnaE	DEG10040051	192	Purine metabolism
5	PSPA7_1499	dnaE	DEG10040051	192	Pyrimidine metabolism
6	PSPA7_1966	cmk	DEG10100298	188	Pyrimidine metabolism
7	PSPA7_5589	ureC	DEG10100307	76	Purine metabolism
	AMINO ACID METABOLISM				
1	PSPA7_0009	glyQ	DEG10040536	254	Glycine, serine and threonine metabolism
2	PSPA7_1323	leuA2	DEG10100584	78	Valine, leucine and isoleucine biosynthesis
3	PSPA7_1473	dapD	DEG10100177	64	Lysine biosynthesis
4	PSPA7_1556	metE	DEG10100167	72	Methionine metabolism
5	PSPA7_2016	asd	DEG10040525	62	Glycine, serine and threonine metabolism
6	PSPA7_2016	asd	DEG10040525	62	Lysine biosynthesis
7	PSPA7_2117	leuA1	DEG10100584	72	Valine, leucine and isoleucine biosynthesis
8	PSPA7_5408	ilvC	DEG10100483	113	Valine, leucine and isoleucine biosynthesis
9	PSPA7_5589	ureC	DEG10100307	76	Urea cycle and metabolism of amino groups
	METABOLISM OF OTHER AMINO ACIDS				
1	PSPA7_4983	murC	DEG10110010	76	D-Glutamine and D-glutamate metabolism
	GLYCAN BIOSYNTHESIS AND METABOLISM				
1	PSPA7_4983	murC	DEG10110010	76	Peptidoglycan biosynthesis
	METABOLISM OF COFACTORS AND VITAMINS				
1	PSPA7_1054	ribA	DEG10040224	78	Riboflavin metabolism
2	PSPA7_5408	ilvC	DEG10100483	113	Pantothenate and CoA biosynthesis
3	PSPA7_5702	thiC	DEG10100055	234	Thiamine metabolism
	XENOBIOTICS BIODEGRADATION AND METABOLISM				
1	PSPA7_5589	ureC	DEG10100307	76	Atrazine degradation
	TRANSCRIPTION				
1	PSPA7_0471	rpoH	DEG10040527	165	RNA polymerase
2	PSPA7_0718	rpoD	DEG10110169	377	RNA polymerase
3	PSPA7_0862	rpoA	DEG10110177	90	RNA polymerase
	TRANSLATION				
1	PSPA7_0849	rplE	DEG10110185	60	Ribosome
2	PSPA7_0852	rplF	DEG10110184	78	Ribosome
3	PSPA7_0854	rpsE	DEG10040500	100	Ribosome
4	PSPA7_0859	rpsM	DEG10040497	168	Ribosome
5	PSPA7_5661	rpsR	DEG10050172	58	Ribosome
6	PSPA7_5662		DEG10040597	113	Ribosome
	FOLDING , SORTING AND DEGRADATION				
1	PSPA7_0857	secY	DEG10110180	88	Protein export
2	PSPA7_4751	lepB1	DEG10040403	60	Protein export
3	PSPA7_4974	secA	DEG10100514	94	Protein export
	REPLICATION AND REPAIR				

1	PSPA7_1499	dnaE	DEG10040051	192	DNA replication
2	PSPA7_1499	dnaE	DEG10040051	192	Mismatch repair
3	PSPA7_1499	dnaE	DEG10040051	192	Homologous recombination
4	PSPA7_5658	dnaB1	DEG10040586	78	DNA replication
	MEMBRANE TRANSPORT				
1	PSPA7_0526	mexB	DEG10110035	339	
	SIGNAL TRANSDUCTION				
1	PSPA7_1235	narH	DEG10110103	377	Two-component system

Table: results of blast of *P.aeruginosa* with the human proteome

CARBOHYDRATE METABOLISM					
	Aminosugars metabolism				
1	PSPA7_1357	K02804			putative phosphotransferase system protein
	Fructose and mannose metabolism				
1	PSPA7_1585	K02770		fruA	fructose-specific IIBC component
	Nucleotide sugars metabolism				
1	PSPA7_5903	K01790	EC:5.1.3.13	rfbC	dTDP-4-dehydrorhamnose 3,5-epimerase
2	PSPA7_1032	K00067	EC:1.1.1.133		hypothetical protein ;
3	PSPA7_1593	K07806	EC:2.6.1.-		UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase
ENERGY METABOLISM					
	Nitrogen metabolism				
1	PSPA7_2726	K05602	EC:1.14.12.3	antB	anthranilate dioxygenase small subunit
2	PSPA7_2726	K05602	EC:1.14.12.3	antB	anthranilate dioxygenase small subunit
3	PSPA7_1234	K00370	EC:1.7.99.4		nitrate reductase 1, alpha subunit
4	PSPA7_1235	K00371	EC:1.7.99.4		nitrate reductase 1, beta subunit
5	PSPA7_1236	K00373	EC:1.7.99.4		nitrate reductase 1, delta subunit
6	PSPA7_1237	K00374	EC:1.7.99.4	narI	nitrate reductase 1, gamma subunit
AMINO ACID METABOLISM					
	Alanine and aspartate metabolism				
1	PSPA7_5657	K01775	EC:5.1.1.1	alr2	alanine racemase
	Phenylalanine, tyrosine and tryptophan biosynthesis				
1	PSPA7_0037	K01695	EC:4.2.1.20	trpA	tryptophan synthase subunit alpha
2	PSPA7_0038	K01696	EC:4.2.1.20	trpB	tryptophan synthase subunit beta
3	PSPA7_0792	K01609	EC:4.1.1.48	trpC	indole-3-glycerol-phosphate synthase
4	PSPA7_0791	K00766	EC:2.4.2.18	trpD	anthranilate phosphoribosyltransferase
5	PSPA7_0753	K01657	EC:4.1.3.27	trpE	anthranilate synthetase component I
METABOLISM OF OTHER AMINO ACIDS					
	D-Alanine metabolism				
1	PSPA7_5658	K01776	EC:5.1.1.2	alr3	alanine racemase
2	PSPA7_0897	K01921	EC:6.3.2.4	ddlA	D-alanine-D-alanine ligase A ;
GLYCAN BIOSYNTHESIS AND METABOLISM					
	Peptidoglycan biosynthesis				
1	PSPA7_0898	K01922	EC:6.3.2.5	ddlA	D-alanine-D-alanine ligase A ;
BIOSYNTHESIS OF POLYKETIDES AND NON-RIBOSOMAL PEPTIDES					
	Biosynthesis of siderophore group nonribosomal peptides				
1	PSPA7_0870	K01852	EC:5.4.4.3	pchA	salicylate biosynthesis isochorismate synthase
	Polyketide sugar unit biosynthesis				
1	PSPA7_5905	K01792	EC:5.1.3.15	rfbC	dTDP-4-dehydrorhamnose 3,5-epimerase
2	PSPA7_1034	K00069	EC:1.1.135		hypothetical protein ;

METABOLISM OF COFACTORS AND VITAMINS					
	Porphyrin and chlorophyll metabolism				
1	PSPA7_0187	K02259			hypothetical protein
	Ubiquinone biosynthesis				
1	PSPA7_0869	K01851	EC:5.4.4.2	pchA	salicylate biosynthesis isochorismate synthase
BIOSYNTHESIS OF SECONDARY METABOLITES					
	Streptomycin biosynthesis				
1	PSPA7_5904	K01791	EC:5.1.3.14	rfbC	dTDP-4-dehydrorhamnose 3,5-epimerase
2	PSPA7_1033	K00068	EC:1.1.1.134		hypothetical protein ;
XENOBIOTICS BIODEGRADATION AND METABOLISM					
	2,4-Dichlorobenzoate degradation				
1	PSPA7_0232	K00450	EC:1.13.11.4	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
2	PSPA7_0232	K00450	EC:1.13.11.4	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
3	PSPA7_0333	K00482	EC:1.14.13.3	pobA	4-hydroxybenzoate 3-monooxygenase
4	PSPA7_0232	K00450	EC:1.13.11.4	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
	Benzoate degradation via hydroxylation				
1	PSPA7_2731	K03464	EC:5.3.3.4	catC	muconolactone delta-isomerase
	Fluorobenzoate degradation				
1	PSPA7_2734	K03383	EC:1.13.11.3	catA	catechol 1,2-dioxygenase
2	PSPA7_2732	K01858	EC:5.5.1.3	catB	muconate cycloisomerase I
	1,4-Dichlorobenzene degradation				
1	PSPA7_2736	K03385	EC:1.13.11.5	catA	catechol 1,2-dioxygenase
2	PSPA7_2736	K03385	EC:1.13.11.5	catA	catechol 1,2-dioxygenase ();
3	PSPA7_2736	K03385	EC:1.13.11.5	catA	catechol 1,2-dioxygenase
4	PSPA7_2734	K01860	EC:5.5.1.5	catB	muconate cycloisomerase I
	2,4-Dichlorobenzoate degradation				
1	PSPA7_0333	K00482	EC:1.14.13.3	pobA	4-hydroxybenzoate 3-monooxygenase
2	PSPA7_5628	K03862		vanA	vanillate O-demethylase oxygenase subunit
3	PSPA7_0232	K00450	EC:1.13.11.4	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
	Benzoate degradation via CoA ligation				
1	PSPA7_2721	K05551	EC:1.14.12.12	benA	benzoate 1,2-dioxygenase, large subunit
2	PSPA7_5624	K01576	EC:4.1.1.7	mdlC	benzoylformate decarboxylase ;
3	PSPA7_2733	K03382	EC:1.13.11.2	catA	catechol 1,2-dioxygenase
4	PSPA7_0332	K00481	EC:1.14.13.2	pobA	4-hydroxybenzoate 3-monooxygenase
5	PSPA7_0231	K00449	EC:1.13.11.3	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
6	PSPA7_5624	K01576	EC:4.1.1.7	mdlC	benzoylformate decarboxylase ;
7	PSPA7_2733	K03382	EC:1.13.11.2	catA	catechol 1,2-dioxygenase ();
8	PSPA7_2724	K05600	EC:1.14.12.1	antB	anthranilate dioxygenase small subunit
9	PSPA7_3224	K01821	EC:5.3.2.-		hypothetical protein ();
10	PSPA7_2724	K05600	EC:1.14.12.1	antB	anthranilate dioxygenase small subunit
11	PSPA7_2719	K05549	EC:1.14.12.10	benA	benzoate 1,2-dioxygenase, large subunit
12	PSPA7_2733	K03382	EC:1.13.11.2	catA	catechol 1,2-dioxygenase
13	PSPA7_2731	K01857	EC:5.5.1.2	catB	muconate cycloisomerase I
14	PSPA7_0231	K00449	EC:1.13.11.3	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
15	PSPA7_0231	K00449	EC:1.13.11.3	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
16	PSPA7_0332	K00481	EC:1.14.13.2	pobA	4-hydroxybenzoate 3-monooxygenase
17	PSPA7_0231	K00449	EC:1.13.11.3	pcaH	protocatechuate 3,4-dioxygenase, beta subunit
18	PSPA7_5624	K01576	EC:4.1.1.7	mdlC	benzoylformate decarboxylase
	Carbazole degradation				
1	PSPA7_2737	K03386	EC:1.13.11.6	catA	catechol 1,2-dioxygenase
2	PSPA7_2737	K03386	EC:1.13.11.6	catA	catechol 1,2-dioxygenase ();
3	PSPA7_2725	K05601	EC:1.14.12.2	antB	anthranilate dioxygenase small subunit
4	PSPA7_2725	K05601	EC:1.14.12.2	antB	anthranilate dioxygenase small subunit
5	PSPA7_2737	K03386	EC:1.13.11.6	catA	catechol 1,2-dioxygenase
6	PSPA7_2735	K01861	EC:5.5.1.6	catB	muconate cycloisomerase I

	Fluorobenzoate degradation				
1	PSPA7_2734	K03383	EC:1.13.11.3	catA	catechol 1,2-dioxygenase
2	PSPA7_2734	K03383	EC:1.13.11.3	catA	catechol 1,2-dioxygenase ();
3	PSPA7_2720	K05550	EC:1.14.12.11	benA	benzoate 1,2-dioxygenase, large subunit
	gamma-Hexachlorocyclohexane degradation				
1	PSPA7_2732	K03381	EC:1.13.11.1	catA	catechol 1,2-dioxygenase
2	PSPA7_2732	K03381	EC:1.13.11.1	catA	catechol 1,2-dioxygenase ();
3	PSPA7_2732	K03381	EC:1.13.11.1	catA	catechol 1,2-dioxygenase
4	PSPA7_2730	K01856	EC:5.5.1.1	catB	muconate cycloisomerase I
	Toluene and xylene degradation				
1	PSPA7_5625	K01577	EC:4.1.1.8	mdlC	benzoylformate decarboxylase ;
2	PSPA7_2735	K03384	EC:1.13.11.4	catA	catechol 1,2-dioxygenase
3	PSPA7_5625	K01577	EC:4.1.1.8	mdlC	benzoylformate decarboxylase ;
4	PSPA7_2735	K03384	EC:1.13.11.4	catA	catechol 1,2-dioxygenase ();
5	PSPA7_2735	K03384	EC:1.13.11.4	catA	catechol 1,2-dioxygenase
6	PSPA7_2733	K01859	EC:5.5.1.4	catB	muconate cycloisomerase I
7	PSPA7_5625	K01577	EC:4.1.1.8	mdlC	benzoylformate decarboxylase
FOLDING ,SORTING AND DEGRADATION					
	Type II secretion system				
1	PSPA7_1416	K02453		gspD4	general secretion pathway protein D
2	PSPA7_2031	K02455		gspF2	general secretion pathway protein F
3	PSPA7_2029	K02452		xcpP	secretion protein XcpP ;
4	PSPA7_2033	K02457		gspH2	general secretion pathway protein H ;
5	PSPA7_2035	K02459		gspJ1	general secretion pathway protein J ;
6	PSPA7_2036	K02460		xcpX	general secretion pathway protein K
7	PSPA7_2037	K02461		gspL	general secretion pathway protein L
8	PSPA7_2038	K02462		xcpZ	general secretion pathway protein M
9	PSPA7_1414	K02505		gspF3	general secretion pathway protein F
10	PSPA7_2177	K02238		ComE	DNA internalization-related competence protein
11	PSPA7_4870	K02282			hypothetical protein
13	PSPA7_4873	K02651			hypothetical protein pilus assembly
14	PSPA7_4869	K02283		CpaF	putative type II secretion system protein
15	PSPA7_4871	K02280		CpaC	putative type II secretion system protein
16	PSPA7_5777	K02666		pilQ	type 4 fimbrial biogenesis protein
17	PSPA7_5162	K02652		pilB	type 4 fimbrial biogenesis protein
18	PSPA7_5163	K02653		PilC	pilin biogenesis protein ;type IV pilus
19	PSPA7_5162	K02651		PilA	type IV pilin structural subunit type
20	PSPA7_5164	K02654	EC:3.4.23.43 2.1.1.-	PilD	type 4 prepilin peptidase
21	PSPA7_5781	K02662		pilM	type 4 fimbrial biogenesis protein
22	PSPA7_5779	K02664		PilO	type 4 fimbrial biogenesis protein
23	PSPA7_5778	K02665		pilP	type 4 fimbrial biogenesis protein
24	PSPA7_0514	K02488			type IV pili sensor histidine kinase/response regulator
25	PSPA7_0512	K02661		PilK	methyltransferase
26	PSPA7_0511	K02660		pilI	twitching motility protein
27	PSPA7_0512	K02661		pilJ	twitching motility protein
28	PSPA7_0509	K02658		pilG	twitching motility protein
29	PSPA7_0509	K02658		pilH	twitching motility protein PilH
30	PSPA7_5195	K02675		pilY2	type 4 fimbrial biogenesis protein
31	PSPA7_5194	K02674		PilY1	type 4 fimbrial biogenesis protein
32	PSPA7_5193	K02673		pilX	type 4 fimbrial biogenesis protein PilX
33	PSPA7_5191	K02671		pilV	type IV pilus modification protein PilV
34	PSPA7_5192	K02672		pilW	type 4 fimbrial biogenesis protein
35	PSPA7_5188	K02668		pilR	two-component response regulator
36	PSPA7_5187	K02669	EC:2.7.13.4	pilS	two-component sensor PilS
37	PSPA7_2201	K02676		pilZ	type 4 fimbrial biogenesis protein
38	PSPA7_0496	K02670		pilU	twitching motility protein
39	PSPA7_0495	K02669		pilT	twitching motility protein

40	PSPA7_3876	K02401		flhA	flagellar biosynthesis protein A
41	PSPA7_3879	K02402		flhB	flagellar biosynthesis protein
42	PSPA7_3880	K02422		FliR ;	flagellar biosynthesis protein
43	PSPA7_3881	K02421		fliQ	flagellar biosynthesis protein
44	PSPA7_3882	K02420		fliP	flagellar biosynthesis protein
45	PSPA7_5187	K02669	EC:2.7.13.4	PilS	two-component system sensor kinase
46	PSPA7_0514	K02488			type IV pili sensor histidine kinase/response regulator
47	PSPA7_0510	K02659		pilH	twitching motility two-component system response regulator
48	:PSPA7_0509	K02658		pilG	twitching motility two-component system response regulator PilG
49	PSPA7_5162	K02651		PilA	type IV pilus assembly protein
50	PSPA7_0511	K02660		pilI	twitching motility protein
51	PSPA7_0512	K02661		pilJ	twitching motility protein
52	PSPA7_5188	K02668		pilR	two-component system response regulator
	Type III secretion system				
1	PSPA7_3885	K02419		fliN	flagellar motor switch protein
2	PSPA7_4268	K02412		FliH	flagellar assembly protein H
3	PSPA7_4271	K02410		fliF	flagellar MS-ring protein
4	PSPA7_3885	K02419		fliN	flagellar motor switch protein
5	PSPA7_3868	K02558		MotA	flagellar motor protein
6	PSPA7_4271	K02410		FliF	flagellar MS-ring protein
7	PSPA7_3885	K02419		fliN	flagellar motor switch protein
8	PSPA7_3876	K02401		flhA	flagellar biosynthesis protein
9	PSPA7_3878	K02401		FlhB	flagellar biosynthesis protein
10	PSPA7_4269	K02412		FliH	flagellar assembly protein
11	PSPA7_3881	K02421		fliQ	flagellar biosynthesis protein
12	PSPA7_3882	K02420		FliP	flagellar biosynthesis protein
13	PSPA7_3880	K02422		FliR	flagellar biosynthesis protein
	Type IV secretion system				
1	PSPA7_3708	K03205			hypothetical protein
2	PSPA7_3706	K03196		trbB	P-type conjugative transfer ATPase
3	PSPA7_3697	K03195		VirB10	conjugation TrbI family protein
4	PSPA7_3698	K03204		VirB9	conjugal transfer protein
5	PSPA7_3699	K03200		VirB5	conjugal transfer protein
6	PSPA7_3703	K03199		VirB4 (A)	ATPase
7	PSPA7_3704	K03198		VirB3	conjugal transfer protein
8	PSPA7_3705	K03197		TrbC	conjugal transfer protein
MEMBRANE TRANSPORT					
	Phosphotransferase system (PTS)				
1	PSPA7_1583	K02768	EC:2.7.1.69		putative phosphotransferase system enzyme I
2	PSPA7_5037	K02806	EC:2.7.1.69		nitrogen regulatory IIA protein
3	PSPA7_5039	K02784		HPr	phosphocarrier protein
4	PSPA7_0430	K08484	EC:2.7.3.9		phosphoenolpyruvate-protein phosphotransferase
5	PSPA7_1586	K02771		fruA	fructose-specific IIBC component
6	PSPA7_1358	K02805			putative phosphotransferase system protein
SIGNAL TRANSDUCTION					
	Two-component system				
1	PSPA7_5943	K07638	EC:2.7.13.3	envZ	osmolarity sensor histidine kinase
2	PSPA7_2347	K07644	EC:2.7.13.3		heavy metal sensor histidine kinase
3	PSPA7_4763	K07649	EC:2.7.13.3	TctE	sensor histidine kinase
4	PSPA7_2708	K07799		MdtA	membrane fusion protein precursor
5	PSPA7_2709	K07788		MdtB	RND efflux transporte
6	PSPA7_2710	K07789		MdtC	RND efflux transporter
7	PSPA7_3872	K02405		fliA	flagellar biosynthesis sigma factor
8	PSPA7_3640	K01546	EC:3.6.3.12	kdpA2	potassium-transporting ATPase subunit

9	PSPA7_4582	K07678	EC:2.7.13.3	BarA	sensor/response regulator hybrid
10	PSPA7_5161	K02650		PilA	type IV pilin structural subunit type
11	PSPA7_0513	K02487			type IV pili sensor histidine kinase/response regulator
12	PSPA7_0510	K02659		pilI	twitching motility protein
13	PSPA7_0511	K02660		pilJ	twitching motility protein
14	PSPA7_0508	K02657		pilG	twitching motility protein
15	PSPA7_0508	K02657		pilH	twitching motility protein PilH
16	PSPA7_5187	K02667		pilR	two-component response regulator
17	PSPA7_5186	K02668	EC:2.7.13.3	pilS	two-component sensor PilS
18	PSPA7_3770	K03776		aer	aerotaxis receptor
19	PSPA7_0253	K03412	EC:3.1.1.61		chemotaxis-specific methylesterase
20	PSPA7_0258	K03407	EC:2.7.13.3		putative two-component sensor
21	PSPA7_0257	K03408		CheW	putative purine-binding chemotaxis protein
22	PSPA7_1780	K03414		CheV	putative chemotaxis protein
23	PSPA7_0259	K03413		CheY	utative two-component response regulator
24	PSPA7_3866	K02556		MotA	flagellar motor protein
25	PSPA7_4279	K02406			A-type flagellin
26	PSPA7_3866	K02556		MotA	flagellar motor protein
27	PSPA7_6139	K07636	EC:2.7.3.-	PhoR	two-component sensor
28	PSPA7_4199	K07637	EC:2.7.13.3	phoQ	two-component sensor
29	PSPA7_3499	K07639	EC:2.7.13.3	RpeA	histidine kinase
30	PSPA7_0569	K07641	EC:2.7.13.3	creC	sensory histidine kinase
31	PSPA7_5498	K07645	EC:2.7.13.3	QseC	two-component regulator system signal sensor kinase
32	PSPA7_3637	K07646	EC:2.7.13.3	kdpD	sensor histidine kinase
33	PSPA7_6138	K07657		PhoB	two-component response regulato
34	PSPA7_4200	K07660		PhoP	two-component response regulator
35	PSPA7_5944	K07659		ompR	osmolarity response regulator
36	PSPA7_0568	K07663		creB	DNA-binding response regulator
37	PSPA7_2348	K07665		CusR	putative two-component response regulator
38	PSPA7_3636	K07667		KdpE	operon response regulator
39	PSPA7_4764	K07774		TctD	two-component response regulator
40	PSPA7_6138	K07657		phoB	two-component response regulator
41	PSPA7_1594	K07807	EC:2.6.1.-		UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase
42	PSPA7_4279	K02406			A-type flagellin
43	PSPA7_3866	K02556		MotA	flagellar motor protein
44	PSPA7_0188	K02260			hypothetical protein
45	PSPA7_3638	K01548	EC:3.6.3.12	kdpC2	potassium-transporting ATPase subunit C
46	PSPA7_4766	K07795			putative tricarboxylic transport membrane protein
47	PSPA7_0038	K01696	EC:4.2.1.21	trpA	tryptophan synthase subunit alpha
48	PSPA7_0039	K01697	EC:4.2.1.21	trpB	tryptophan synthase subunit beta
49	PSPA7_0793	K01610	EC:4.1.1.49	trpC	indole-3-glycerol-phosphate synthase
50	PSPA7_0792	K00767	EC:2.4.2.19	trpD	anthranilate phosphoribosyltransferase
51	PSPA7_0754	K01658	EC:4.1.3.28	trpE3	anthranilate synthetase component I
52	PSPA7_0293	K05966	EC:2.7.8.25	mdcB	triphosphoribosyl-dephospho-CoA synthase
53	PSPA7_1230	K07673	EC:2.7.13.3	narX	wo-component sensor
54	PSPA7_3015	K07677	EC:2.7.13.3	NarL	kinase sensor protein
55	PSPA7_5186	K02668	EC:2.7.13.3	PilS	two-component system sensor kinase
56	PSPA7_0257	K03408		CheW	putative purine-binding chemotaxis protein
57	PSPA7_3770	K03776		aer	aerotaxis receptor
58	PSPA7_0258	K03407	EC:2.7.13.3	CheA	putative two-component sensor
59	PSPA7_0510	K02659		pilI	twitching motility protein
60	PSPA7_0511	K02660		pilJ	twitching motility protein
61	PSPA7_0513	K02487			type IV pili sensor histidine kinase/response regulator
62	PSPA7_0509	K02658		pilH	twitching motility two-component system response regulator
63	PSPA7_0508	K02657		pilG	twitching motility two-component system response regulator PilG
64	PSPA7_1781	K03415		CheV	two-component system, chemotaxis family

65	PSPA7_0253	K03412	EC:3.1.1.61	CheB	chemotaxis-specific methylesterase
66	PSPA7_0259	K03413		CheY	two-component system, chemotaxis family, response regulator
67	PSPA7_5187	K02667		pilR	two-component system response regulator
68	PSPA7_5859	K07712		ntrC	nitrogen regulation response regulator
69	PSPA7_1160	K07690		EvgA	putative two-component response regulato
70	PSPA7_1229	K07684		NarL	nitrate/nitrite response regulator
71	PSPA7_5161	K02650		PilA	type IV pilus assembly protein
72	PSPA7_3291	K02106			hypothetical
73	PSPA7_1481	K00990	EC:2.7.7.59	glnD	PII uridylyl-transferase
74	PSPA7_1235	K00371	EC:1.7.99.5		nitrate reductase 1, alpha subunit
75	PSPA7_1236	K00372	EC:1.7.99.5		nitrate reductase 1, beta subunit
76	PSPA7_1237	K00374	EC:1.7.99.5		nitrate reductase 1, delta subunit
77	PSPA7_1238	K00375	EC:1.7.99.5	narI	nitrate reductase 1, gamma subunit
78	PSPA7_1922	K07662			two-component response regulator
CELL MOTILITY					
	Bacterial chemotaxis				
1	PSPA7_3883	K02417		fliN	flagellar motor switch protein
2	PSPA7_3771	K03777		aer	aerotaxis receptor
3	PSPA7_0254	K03413	EC:3.1.1.62		chemotaxis-specific methylesterase
4	PSPA7_0259	K03408	EC:2.7.13.4		putative two-component sensor
5	PSPA7_0258	K03409		CheW	putative purine-binding chemotaxis protein
6	PSPA7_1781	K03415		CheV	putative chemotaxis protein
7	PSPA7_0255	K00575	EC:2.1.1.80	CheR	putative chemotaxis protein methyltransferase
8	PSPA7_0260	K03414		CheY	utative two-component response regulator
9	PSPA7_3870	K03414		cheZ	chemotaxis protein
10	PSPA7_0254	K03411	EC:3.5.1.44	CheD	hypothetical protein
11	PSPA7_3883	K02417		fliN	flagellar motor switch protein
12	PSPA7_3884	K02416		fliM	lagellar motor switch protein
13	PSPA7_4269	K02410		fliG	flagellar motor switch protein G
14	PSPA7_3867	K02557		MotA	flagellar motor protein
15	PSPA7_3865	K02557		MotB	flagellar motor protein
16	PSPA7_3865	K02557		MotB	flagellar motor protein
17	PSPA7_4269	K02410		FliG	flagellar motor switch protein
18	PSPA7_3884	K02416		FliM	flagellar motor switch protein
19	PSPA7_3883	K02417		fliN	flagellar motor switch protein
20	PSPA7_3867	K02557		MotA	flagellar motor protein
21	PSPA7_0258	K03409		CheW	putative purine-binding chemotaxis protein
22	PSPA7_3771	K03777		aer	aerotaxis receptor
23	PSPA7_0259	K03408	EC:2.7.13.4	CheA	putative two-component sensor
24	PSPA7_1782	K03416		CheV	two-component system, chemotaxis family
25	PSPA7_0254	K03413	EC:3.1.1.62	CheB	chemotaxis-specific methylesterase
26	PSPA7_0260	K03414		CheY	two-component system, chemotaxis family, response regulator
	Flagellar assembly				
1	PSPA7_3875	K02400		flhA	flagellar biosynthesis protein A
2	PSPA7_3878	K02401		flhB	flagellar biosynthesis protein
3	PSPA7_3879	K02421		FliR	flagellar biosynthesis protein
4	PSPA7_3880	K02420		fliQ	flagellar biosynthesis protein
5	PSPA7_3881	K02419		fliP	flagellar biosynthesis protein
6	PSPA7_3884	K02418		fliN	flagellar motor switch protein
7	PSPA7_4268	K02411		FliH	flagellar assembly protein H
8	PSPA7_4270	K02409		fliF	flagellar MS-ring protein
9	PSPA7_3884	K02418		fliN	flagellar motor switch protein
10	PSPA7_3885	K02417		fliM	lagellar motor switch protein
11	PSPA7_4270	K02411		fliG	flagellar motor switch protein G
12	PSPA7_3868	K02558		MotA	flagellar motor protein
13	PSPA7_3866	K02558		MotB	flagellar motor protein

14	PSPA7_4277	K02407		FliD	flagellar cap protein
15	PSPA7_4280	K02407			A-type flagellin
16	PSPA7_4290	K02397		FlgL	flagellar hook-associated protein
17	PSPA7_4291	K02396		flgK	flagellar hook-associated protein
18	PSPA7_4297	K02390		FlgE	flagellar hook protein
19	PSPA7_3888	K02414		FliK	hypothetical protein
20	PSPA7_4298	K02389		flgD	flagellar basal-body rod modification protein
21	PSPA7_4296	K02391		flgF	flagellar basal-body rod protein
22	PSPA7_4295	K02392		flgG	flagellar basal body rod protein
23	PSPA7_4294	K02393		FlgH	flagellar basal body L-ring protein
24	PSPA7_4293	K02394		FlgI	flagellar basal body P-ring protein
25	PSPA7_3866	K02558		MotB	flagellar motor protein
26	PSPA7_3867	K02557		MotA	flagellar motor protein
27	PSPA7_4271	K02408		FliE	flagellar hook-basal body protein
28	PSPA7_4299	K02388		flgC	flagellar basal body rod protein
29	PSPA7_4270	K02409		FliF	flagellar MS-ring protein
30	PSPA7_4270	K02411		FliG	flagellar motor switch protein
31	PSPA7_3885	K02417		FliM	flagellar motor switch protein
32	PSPA7_3884	K02418		fliN	flagellar motor switch protein
33	PSPA7_3875	K02400		flhA	flagellar biosynthesis protein
34	PSPA7_3877	K02400		FlhB	flagellar biosynthesis protein
35	PSPA7_4268	K02411		FliH	flagellar assembly protein
36	PSPA7_3882	K02418		FliO	flagellar protein
37	PSPA7_3880	K02420		fliQ	flagellar biosynthesis protein
38	PSPA7_3881	K02419		FliP	flagellar biosynthesis protein
39	PSPA7_3879	K02421		FliR	flagellar biosynthesis protein
40	PSPA7_1778	K02399		FlgN	hypothetical protein
41	PSPA7_4266	K02413		fliJ	flagellar biosynthesis chaperone
42	PSPA7_1780	K02386		FlgA	flagellar basal body P-ring biosynthesis protein
43	PSPA7_1779	K02398		FlgM	hypothetical protein
44	PSPA7_4300	K02387		FlgB	flagellar basal body rod protein
45	PSPA7_4280	K02407			A-type flagellin
46	PSPA7_3868	K02558		MotA	flagellar motor protein

TABLE 4.4 : RESULTS OF BLAST OF *Pseudomonas syringea* B728a WITH DEG.

	GENE ID	NAME	DEG ID	SCORE	PATHWAY
	CARBOHYDRATE METABOLISM				
1	Psyr_0517		DEG10040028	66	Glycolysis / Gluconeogenesis
2	Psyr_0517		DEG10040028	66	Pyruvate metabolism
3	Psyr_0924		DEG10100033	64	Nucleotide sugars metabolism
4	Psyr_0944		DEG10040205	74	Pentose phosphate pathway
5	Psyr_1120		DEG10040316	62	Pentose phosphate pathway
6	Psyr_1257		DEG10100584	64	Pyruvate metabolism
7	Psyr_1363	eno	DEG10110158	157	Glycolysis / Gluconeogenesis
8	Psyr_2009	kgd	DEG10040125	137	Citrate cycle (TCA cycle)
9	Psyr_2010		DEG10040126	113	Citrate cycle (TCA cycle)
10	Psyr_2176		DEG10110051	70	Propanoate metabolism
11	Psyr_2988		DEG10100164	68	Pentose phosphate pathway
12	Psyr_3404	acnA	DEG10100247	111	Citrate cycle (TCA cycle)
13	Psyr_3404	acnA	DEG10100247	111	Glyoxylate & dicarboxylate metabolism
14	Psyr_4792		DEG10040444	206	Pentose phosphate pathway
	ENERGY METABOLISM				
1	Psyr_0624		DEG10040601	66	Oxidative phosphorylation
2	Psyr_1238		DEG10040392	70	Nitrogen metabolism
3	Psyr_3404	acnA	DEG10100247	111	Reductive carboxylate cycle
4	Psyr_3483		DEG10040067	70	Nitrogen metabolism
5	Psyr_4126		DEG10100190	113	Sulfur metabolism

6	Psyr_4270	glyA	DEG10040397	137	Methane metabolism
7	Psyr_4712		DEG10040397	78	Methane metabolism
9	Psyr_4817		DEG10100352	125	Nitrogen metabolism
10	Psyr_4833	hutH	DEG10110048	105	Nitrogen metabolism
	LIPID METABOLISM				
1	Psyr_0500		DEG10100527	74	Fatty acid biosynthesis
2	Psyr_0713	ispH	DEG10040008	62	Biosynthesis of steroids
3	Psyr_1349		DEG10040043	62	Biosynthesis of steroids
4	Psyr_1647	fabG	DEG10040181	72	Fatty acid biosynthesis
5	Psyr_1647	fabG	DEG10040181	72	Biosynthesis of unsaturated fatty acids
6	Psyr_1754		DEG10110097	62	Fatty acid biosynthesis
7	Psyr_2019		DEG10110135	133	Fatty acid biosynthesis
8	Psyr_2020		DEG10040157	107	Fatty acid biosynthesis
9	Psyr_2176		DEG10110051	70	Fatty acid metabolism
	NUCLEOTIDE METABOLISM				
1	Psyr_0002		DEG10110201	68	Purine metabolism
2	Psyr_0002		DEG10110201	68	Pyrimidine metabolism
3	Psyr_0944		DEG10040205	74	Purine metabolism
4	Psyr_1269		DEG10130197	72	Purine metabolism
5	Psyr_1345	pyrH	DEG10110017	70	Pyrimidine metabolism
6	Psyr_1358	dnaE	DEG10040051	159	Purine metabolism
7	Psyr_1358	dnaE	DEG10040051	159	Pyrimidine metabolism
8	Psyr_1361		DEG10110159	107	Pyrimidine metabolism
9	Psyr_1825		DEG10040089	86	Purine metabolism
10	Psyr_1825		DEG10040089	86	Pyrimidine metabolism
11	Psyr_3717		DEG10130065	100	Purine metabolism
12	Psyr_3717		DEG10130065	100	Pyrimidine metabolism
13	Psyr_3721		DEG10130064	113	Purine metabolism
14	Psyr_3721		DEG10130064	113	Pyrimidine metabolism
15	Psyr_4126		DEG10100190	113	Purine metabolism
16	Psyr_4176		DEG10040475	113	Purine metabolism
17	Psyr_4176		DEG10040475	113	Pyrimidine metabolism
18	Psyr_4191	carB	DEG10130216	133	Pyrimidine metabolism
19	Psyr_4524		DEG10110177	82	Purine metabolism
20	Psyr_4524		DEG10110177	82	Pyrimidine metabolism
21	Psyr_4554		DEG10110213	521	Purine metabolism
22	Psyr_4554		DEG10110213	521	Pyrimidine metabolism
23	Psyr_4555	rpoB	DEG10040577	412	Purine metabolism
24	Psyr_4555	rpoB	DEG10040577	412	Pyrimidine metabolism
	AMINO ACID METABOLISM				
1	Psyr_0012	glyQ	DEG10040536	230	Glycine, serine and threonine metabolism
2	Psyr_0183	dapF	DEG10040559	64	Lysine biosynthesis
3	Psyr_0460		DEG10100518	82	Methionine metabolism
4	Psyr_0487 -		DEG10040446	66	Glutamate metabolism
5	Psyr_0517		DEG10040028	66	Alanine and aspartate metabolism
6	Psyr_0710	ileS	DEG10040006	66	Valine,leucine,isoleucine biosynthesis
7	Psyr_0848		DEG10100483	74	Valine,leucine,isoleucine biosynthesis
8	Psyr_1086	valS	DEG10040605	123	Valine,leucine,isoleucine biosynthesis
9	Psyr_1096		DEG10100306	76	Glycine,serine,threonine metabolism
10	Psyr_1249	hisS	DEG10040386	100	Histidine metabolism
11	Psyr_1257		DEG10100584	64	Valine,leucine,isoleucine biosynthesis
12	Psyr_1311	lysS	DEG10110164	98	Lysine biosynthesis
13	Psyr_1399		DEG10040054	100	Arginine and proline metabolism
14	Psyr_1406	aspS	DEG10130053	101	Alanine and aspartate metabolism
15	Psyr_1734		DEG10110046	64	Glutamate metabolism

16	Psyr_1735	cysS	DEG10040100	74	Cysteine metabolism
17	Psyr_1985		DEG10100480	66	Valine,leucine,isoleucine biosynthesis
18	Psyr_1986		DEG10040525	86	Glycine,serine,threonine metabolism
19	Psyr_1986		DEG10040525	86	Lysine biosynthesis
20	Psyr_2009	kgd	DEG10040125	137	Lysine degradation
21	Psyr_2009	kgd	DEG10040125	137	Tryptophan metabolism
22	Psyr_2010		DEG10040126	113	Lysine degradation
23	Psyr_2162	thrS	DEG10040294	98	Glycine, serine,threonine metabolism
24	Psyr_2162	thrS	DEG10040294	98	Glycine, serine,threonine metabolism
25	Psyr_2176		DEG10110051	70	Valine, leucine,isoleucine degradation
26	Psyr_3556	alaS	DEG10040423	147	Alanine and aspartate metabolism
27	Psyr_3886	metG	DEG10130067	84	Methionine metabolism
28	Psyr_4107	murE	DEG10110006	64	Lysine biosynthesis
29	Psyr_4191	carB	DEG10130216	133	Glutamate metabolism
30	Psyr_4270	glyA	DEG10040397	137	Glycine, serine and threonine metabolism
31	Psyr_4352	leuS	DEG10040114	84	Valine, leucine and isoleucine biosynthesis
32	Psyr_4712		DEG10040397	78	Glycine, serine and threonine metabolism
33	Psyr_4794		DEG10100226	149	Methionine metabolism
34	Psyr_4817		DEG10100352	125	Glutamate metabolism
35	Psyr_4833	hutH	DEG10110048	105	Histidine metabolism
36	Psyr_4893		DEG10100263	66	Histidine metabolism
	METABOLISM OF OTHER AMINO ACID METABOLISM				
1	Psyr_0696		DEG10040306	70	Selenoamino acid metabolism
2	Psyr_1120		DEG10040316	62	Glutathione metabolism
3	Psyr_2176		DEG10110051	70	beta-Alanine metabolism
4	Psyr_2988		DEG10100164	68	Glutathione metabolism
5	Psyr_3886	metG	DEG10130067	84	Selenoamino acid metabolism
6	Psyr_4101	murC	DEG10110010	60	D-Glutamine and D-glutamate metabolism
7	Psyr_4126		DEG10100190	113	Selenoamino acid metabolism
8	Psyr_4270	glyA	DEG10040397	137	Cyanoamino acid metabolism
9	Psyr_4712		DEG10040397	78	Cyanoamino acid metabolism
10	Psyr_4794		DEG10100226	149	Selenoamino acid metabolism
	GLYCAN BIOSYNTHESIS AND METABOLISM				
1	Psyr_4101	murC	DEG10110010	60	Peptidoglycan biosynthesis
2	Psyr_4107	murE	DEG10110006	64	Peptidoglycan biosynthesis
3	Psyr_4817		DEG10100352	125	Peptidoglycan biosynthesis
	METABOLISM OF COFACTORS AND VITAMINS				
1	Psyr_0302		DEG10110206	165	Ubiquinone biosynthesis
2	Psyr_0387		DEG10040563	96	Ubiquinone biosynthesis
3	Psyr_0414	hemE	DEG10110214	88	Porphyrin and chlorophyll metabolism
4	Psyr_0544		DEG10100055	178	Thiamine metabolism
5	Psyr_0848		DEG10100483	74	Pantothenate and CoA biosynthesis
6	Psyr_1237	iscS	DEG10040393	84	Thiamine metabolism
7	Psyr_4342	hemL	DEG10040034	123	Porphyrin and chlorophyll metabolism
	BIOSYNTHESIS OF SECONDARY METABOLITES				
1	Psyr_0924		DEG10100033	64	Streptomycin biosynthesis
	TRANSCRIPTION				
1	Psyr_4524		DEG10110177	82	RNA polymerase
2	Psyr_4554		DEG10110213	521	RNA polymerase
3	Psyr_4555	rpoB	DEG10040577	412	RNA polymerase
	TRANSLATION				
1	Psyr_0012	glyQ	DEG10040536	230	Aminoacyl-tRNA biosynthesis

2	Psyr_0581	rpsF	DEG10040597	72	Ribosome
3	Psyr_0582	rpsR	DEG10130171	76	Ribosome
4	Psyr_0710	ileS	DEG10040006	66	Aminoacyl-tRNA biosynthesis
5	Psyr_1086	valS	DEG10040605	123	Aminoacyl-tRNA biosynthesis
6	Psyr_1249	hisS	DEG10040386	100	Aminoacyl-tRNA biosynthesis
7	Psyr_1311	lysS	DEG10110164	98	Aminoacyl-tRNA biosynthesis
8	Psyr_1343	rpsB	DEG10040039	172	Ribosome
9	Psyr_1399		DEG10040054	100	Aminoacyl-tRNA biosynthesis
10	Psyr_1406	aspS	DEG10130053	101	Aminoacyl-tRNA biosynthesis
11	Psyr_1734		DEG10110046	64	Aminoacyl-tRNA biosynthesis
12	Psyr_1735	cysS	DEG10040100	74	Aminoacyl-tRNA biosynthesis
13	Psyr_2165	rplT	DEG10010205	64	Ribosome
14	Psyr_3556	alaS	DEG10040423	147	Aminoacyl-tRNA biosynthesis
15	Psyr_3886	metG	DEG10130067	84	Aminoacyl-tRNA biosynthesis
16	Psyr_4119	rpsI	DEG10040486	190	Ribosome
17	Psyr_4120	rplM	DEG10040487	86	Ribosome
18	Psyr_4352	leuS	DEG10040114	84	Aminoacyl-tRNA biosynthesis
19	Psyr_4525	rpsD	DEG10130246	62	Ribosome
20	Psyr_4526		DEG10130247	103	Ribosome
21	Psyr_4527	rpsM	DEG10040497	139	Ribosome
22	Psyr_4531	rpsE	DEG10110183	88	Ribosome
23	Psyr_4533	rplF	DEG10110184	76	Ribosome
24	Psyr_4536	rplE	DEG10110185	76	Ribosome
25	Psyr_4538	rplN	DEG10050275	78	Ribosome
26	Psyr_4540	rpmC	DEG10110186	145	Ribosome
27	Psyr_4544	rpsS	DEG10040508	58	Ribosome
28	Psyr_4545	rplB	DEG10110187	109	Ribosome
29	Psyr_4549	rpsJ	DEG10040513	157	Ribosome
30	Psyr_4552		DEG10110191	84	Ribosome
31	Psyr_4553	rpsL	DEG10040516	180	Ribosome
32	Psyr_4558	rplA	DEG10050166	78	Ribosome
33	Psyr_4639	rpsU	DEG10130100	60	Ribosome
	FOLDING,SORTING AND DEGRADING				
	Psyr_1281		DEG10040412	121	Protein export
	Psyr_4094	secA	DEG10040026	121	Protein export
	Psyr_4528	secY	DEG10110180	64	Protein export
	REPLICATION AND REPAIR				
1	Psyr_0002		DEG10110201	68	DNA replication
2	Psyr_0002		DEG10110201	68	Mismatch repair
3	Psyr_0002		DEG10110201	68	Homologous recombination
4	Psyr_0398		DEG10040569	60	Homologous recombination
5	Psyr_0585		DEG10040586	68	DNA replication
6	Psyr_1358	dnaE	DEG10040051	159	DNA replication
7	Psyr_1358	dnaE	DEG10040051	159	DNA replication
8	Psyr_1358	dnaE	DEG10040051	159	Homologous recombination
9	Psyr_1376		DEG10110156	74	Mismatch repair
10	Psyr_1825		DEG10040089	86	DNA replication
11	Psyr_1825		DEG10040089	86	Mismatch repair
12	Psyr_1825		DEG10040089	86	Homologous recombination
	MEMBRANE TRANSPORT				
1	Psyr_0081		DEG10100381	82	ABC transporters
2	Psyr_0759		DEG10100446	60	ABC transporters
3	Psyr_1075		DEG10040115	101	ABC transporters
4	Psyr_2263		DEG10040186	60	ABC transporters
5	Psyr_3076		DEG10110057	62	ABC transporters
6	Psyr_3911		DEG10040115	86	ABC transporters

	SIGNAL TRANSDUCTION				
1	Psyr_4817		DEG10100352	125	Two-component system

TABLE 4.5 : RESULTS OF BLAST OF *Pseudomonas syringe* b728a WITH *Pseudomonas syringe* KT2440

	GENE ID	**GENE NAME**	**PROTEIN NAME**
carbohydrate metabolism			
	Citrate cycle (TCA cycle)		
1	Psyr_0518		dihydrolipoamide acetyltransferase
	Glycolysis / Gluconeogenesis		
1	Psyr_0517		dihydrolipoamide acetyltransferase
	Pentose phosphate pathway		
1	Psyr_1120		glucose-6-phosphate 1-dehydrogenase
2	Psyr_2988		glucose-6-phosphate 1-dehydrogenase
	Propanoate metabolism		
1	Psyr_2179		Acyl-CoA dehydrogenase,
	Pyruvate metabolism		
1	Psyr_0519		dihydrolipoamide acetyltransferase
Energy metabolism			
	Nitrogen metabolism		
1	Psyr_3483		cyanate hydratase
Lipid metabolism			
	Biosynthesis of steroids		
1	Psyr_1349		1-deoxy-D-xylulose 5-phosphate reductoisomerase
2	Psyr_1648	fabG	3-ketoacyl-(acyl-carrier-protein) reductase
	Fatty acid biosynthesis		
1	Psyr_0500		carbamoyl-phosphate synthase subunit L
2	Psyr_1647	fabG	3-ketoacyl-(acyl-carrier-protein) reductase
3	Psyr_1754		Short-chain dehydrogenase/reductase SDR
4	Psyr_2176		Acyl-CoA dehydrogenase
Nucleotide metabolism			
	Purine metabolism		
1	Psyr_1269		phosphoribosylformylglycinamidine synthase
Aminoacid metabolism			
	Cysteine metabolism		
1	Psyr_1735	cysS	cysteinyl-tRNA synthetase
	Glutamate metabolism		
1	Psyr_0487		glutathione synthetase
	Glycine, serine and threonine metabolism		
1	Psyr_0012	glyQ	glycyl-tRNA synthetase subunit alpha
	Histidine metabolism		
1	Psyr_4893		imidazole glycerol phosphate synthase subunit HisF
	Lysine biosynthesis		

1	Psyr_0183	dapF	diaminopimelate epimerase
2	Psyr_4107	murE	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase
	Methionine metabolism		
1	Psyr_0460		S-adenosyl-L-homocysteine hydrolase
	Valine, leucine and isoleucine degradation		
1	Psyr_2177		Acyl-CoA dehydrogenase,
Metabolism of other amino acids			
	Selenoamino acid metabolism		
1	Psyr_0461		S-adenosyl-L-homocysteine hydrolase
	beta-Alanine metabolism		
1	Psyr_2178		Acyl-CoA dehydrogenase
	Glutathione metabolism		
1	Psyr_0488		glutathione synthetase
2	Psyr_1121		glucose-6-phosphate 1-dehydrogenase
3	Psyr_2989		glucose-6-phosphate 1-dehydrogenase
	D-Glutamine and D-glutamate metabolism		
1	Psyr_4104	murD	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
	Peptidoglycan biosynthesis		
1	Psyr_4108	murE	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase
2	Psyr_4104	murD	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
Metabolism of cofactors and vitamins			
	Ubiquinone biosynthesis		
1	Psyr_0387		2-polyprenylphenol 6-hydroxylase
Translation			
	Aminoacyl-tRNA biosynthesis		
1	Psyr_0013	glyQ	glycyl-tRNA synthetase subunit alpha
2	Psyr_1736	cysS	cysteinyl-tRNA synthetase
	Ribosome		
1	Psyr_2165	rplT	50S ribosomal protein L20
2	Psyr_4120	rplM	50S ribosomal protein L13
3	Psyr_4544	rpsS	30S ribosomal protein S19
4	Psyr_4639	rpsU	30S ribosomal protein S21
Membrane transport			
	ABC transporters		
1	Psyr_0759		ABC transporter:TOBE
2	Psyr_3076	HlyB	Type I secretion system ATPase,

TABLE 4.6 : RESULTS OF BLAST OF *Pseudomonas syringe* b728a *WITH EST OF PHASEOLUS VULGARIS*

	GENE ID	GENE NAME	DEG ID	SCORE	PATHWAY
	CARBOHYDRATE METABOLISM				
1	Psyr_0517		DEG10040028	66	Glycolysis / Gluconeogenesis
2	Psyr_0517		DEG10040028	66	Pyruvate metabolism
3	Psyr_1120		DEG10040316	62	Pentose phosphate pathway
4	Psyr_2176		DEG10110051	70	Propanoate metabolism
5	Psyr_2988		DEG10100164	68	Pentose phosphate pathway

	ENERGY METABOLISM				
1	Psyr_1238		DEG10040392	70	Nitrogen metabolism
2	Psyr_3483		DEG10040067	70	Nitrogen metabolism
	LIPID METABOLISM				
1	Psyr_0500		DEG10100527	74	Fatty acid biosynthesis
2	Psyr_1349		DEG10040043	62	Biosynthesis of steroids
3	Psyr_1647	fabG	DEG10040181	72	Fatty acid biosynthesis
4	Psyr_1647	fabG	DEG10040181	72	Biosynthesis of unsaturated fatty acids
5	Psyr_1754		DEG10110097	62	Fatty acid biosynthesis
6	Psyr_2176		DEG10110051	70	Fatty acid metabolism
	NUCLEOTIDE METABOLISM				
1	Psyr_1269		DEG10130197	72	Purine metabolism
	AMINO ACID METABOLISM				
1	Psyr_0012	glyQ	DEG10040536	230	Glycine, serine and threonine metabolism
2	Psyr_0183	dapF	DEG10040559	64	Lysine biosynthesis
3	Psyr_0487		DEG10040446	66	Glutamate metabolism
4	Psyr_0517		DEG10040028	66	Alanine and aspartate metabolism
5	Psyr_1735	cysS	DEG10040100	74	Cysteine metabolism
6	Psyr_1985		DEG10100480	66	Valine, leucine and isoleucine biosynthesis
7	Psyr_2176		DEG10110051	70	Valine, leucine and isoleucine degradation
8	Psyr_4107	murE	DEG10110006	64	Lysine biosynthesis
9	Psyr_4893		DEG10100263	66	Histidine metabolism
10	Psyr_4104	murD	DEG10110006	52	D-Glutamine and D-glutamate metabolism
	METABOLISM OF OTHER AMINO ACID METABOLISM				
1	Psyr_0487		DEG10040446	66	Glutathione metabolism
2	Psyr_1120		DEG10040316	62	Glutathione metabolism
3	Psyr_2176		DEG10110051	70	beta-Alanine metabolism
4	Psyr_2988		DEG10100164	68	Glutathione metabolism
	GLYCAN BIOSYNTHESIS AND METABOLISM				
1	Psyr_4107	murE	DEG10110006	64	Peptidoglycan biosynthesis
2	Psyr_4104	murD	DEG10110006	52	Peptidoglycan biosynthesis
	METABOLISM OF COFACTORS AND VITAMINS				
1	Psyr_0387		DEG10040563	96	Ubiquinone biosynthesis
	TRANSLATION				
1	Psyr_0012	glyQ	DEG10040536	230	Aminoacyl-tRNA biosynthesis
2	Psyr_1735	cysS	DEG10040100	74	Aminoacyl-tRNA biosynthesis
3	Psyr_2165	rplT	DEG10010205	64	Ribosome
4	Psyr_4120	rplM	DEG10040487	86	Ribosome
5	Psyr_4544	rpsS	DEG10040508	58	Ribosome
6	Psyr_4639	rpsU	DEG10130100	60	Ribosome
	MEMBRANE TRANSPORT				
1	Psyr_0759		DEG10100446	60	ABC transporters
2	Psyr_3076		DEG10110057	62	ABC transporters

4.1 Perl Program used to download essential bacterial genes in Html files from DEG:

```
use LWP::Simple;
$id='DEG';
$counter=10010001;
foreach(1..243)
{
push(@array,$counter);
$counter++;
}
$count=0;
foreach(@array)
{
$_=$id.$_;
open(GG,">>Seq.html");
$count++;
print $count."). Fetching ".$_."....\n";
my $esearch_result = get("http://tubic.tju.edu.cn/deg/information.php?ac=".$_);
print GG $esearch_result."\n";
print "Fetched....\n";
close GG;
}
print "Finished...";
<stdin>;
}
```

4.2 Perl Program used to concatenate the spilt blast result :

```
#$file='result.txt';
#open(F1,">>$file);
$count=1;
while($count<642)
{
open(F1,"./".$count."\.txt");
@lines=<F1>;
$file='result.txt';
open(F2,">>$file");
print F2 @lines;
close (F1);
$count=$count+1;
}
```

4.3 Perl Program used to spilt the blast result generated and assigning unique ids

```
open(F2,">result.txt");
for($i=1;$i<6286;$i++)
{
open(F1,"./1/".$i."\.txt");
while(<F1>)
{
if(/No/)
```

```
{
print F2 $i." ".$_;
}
}
}
```

4.4 Perl Program used to make excel file by extracting the information in the folder :

```
open(F2,">result.xls");
for($i=1;$i<637;$i++)
{
open(F1,$i."\.txt");
$ref_q=0;
$ref_l=0;

$ref=0;
while(<F1>)
{
if(/No hits/)
{
$ref=1;
}
}
if($ref eq 1 )
{
open(F1,$i."\.txt");
print F2 $i."\t";
while(<F1>)
{

if(/^Query/)
{
chomp($_);
@w=split(/[|]/,$_);
print F2 $w[1]."|".$w[2]."\t";
}
if(/letters/)
{
@w=split(/[(]/,$_);
@w1=split(/[ ]/,$w[1]);
print F2 $w1[0]."\n";
}
 }
}
}
```

4.5 Perl Program used to get the NCBI id by giving the sequence number :

```
 open(F1,"egenes.txt");
$i=0;
$r=<stdin>;
chomp($r);
while(<F1>)
```

```
{
if(/>/)
{
$i=$i+1;
if($i eq $r )
{
print $_;
}

}
}
$r=<stdin>;
```

4.6 Perl Program used to find the sequence with no homology with the datase sequence :

```
open(F2,">result.xls");
for($i=1;$i<637;$i++)
{
open(F1,$i."\.txt");
$ref_q=0;
$ref_l=0;

$ref=0;
while(<F1>)
{
if(/No hits/)
{
$ref=1;
}
}
if($ref eq 1 )
{
open(F1,$i."\.txt");
print F2 $i."\t";
while(<F1>)
{

if(/^Query/)
{
chomp($_);
@w=split(/[|]/,$_);
print F2 $w[1]."|".$w[2]."\t";
}
if(/letters/)
{
@w=split(/[(]/,$_);
@w1=split(/[ ]/,$w[1]);
print F2 $w1[0]."\n";
}

}
}
}
```

4.7 Perl Program used to concatenate the spilt blastn result :

```
print "enetr file name";
$file=<stdin>;
chomp($file);
 open(F1,$file."\.txt");
mkdir "./1";
 while(<F1>)
{
if(/BLASTN/)
{
$count=$count+1;
open(F2,">./1/".$count."\.txt");
}
print F2 $_;
}
```

4.7 Perl Program used to concatenate the spilt blastx result :

```
print "enetr file name";
$file=<stdin>;
chomp($file);
 open(F1,$file."\.txt");
mkdir "./1";
while(<F1>)
{
if(/BLASTX/)
{
$count=$count+1;
open(F2,">./1/".$count."\.txt");
}
print F2 $_;
 }
```