Computer aided target Identification and drug design for pathogen *Chlamydophila psittaci*

A Major Project II dissertation submitted

In partial fulfillment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Surya Kant Singh

(2K13/BIO/17)

Delhi Technological University, Delhi, India

Under the supervision of

Dr.Yasha Hasija

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
ShahbadDaulatpur, Main Bawana Road,
Delhi-110042, INDIA

CERTIFICATE

This is to certify that the dissertation entitled **Computer aided target identification and drug design for pathogen *Chlamydophila psittaci* (Surya Kant Singh, 2k13/bio/17)** in the partial fulfillment of the requirements for the award the degree of Master of Technology. (Bioinformatics), Delhi Technological University (Formerly Delhi College of Engineering), is a *bona fide* record of the candidate's own work carried out by him under my guidance. The information and data enclosed in this thesis is original and has not been submitted elsewhere for honoring of any other degree.

Date:

Dr. Yasha Hasija

(Mentor)

Assistant professor

Department of Bio-Technology

Delhi Technological University

New Delhi

Prof. D Kumar

(Head of Department)

Department of Bio-Technology

Delhi Technological University

New Delhi

DECLARATION

I hereby declare that the thesis entitled "Computer aided target identification and drug design for pathogen *Chlamydophila psittaci*" being submitted by me to Delhi Technological University for the partial fulfillment of the Degree of M.Tech in Bioinformatics, is an original and authentic research work conducted by myself under the guidance of Dr. Yasha Hasija.

To reiterate, no part of this research work or in any full form has been submitted elsewhere for a similar degree. Due acknowledgement have been made whenever facilities and suggestions are availed of.

Date: Name:

Place: Signature:

ACKNOWLEDGEMENT

The project in this report is an outcome of continual work over and intellectual support from various sources. It is therefore almost impossible to express adequately the debts owed to many persons who have been instrumental in imparting this work, a successful status. It is however a matter of pleasure to express my gratitude and appreciation to those who have been contributing to bring about this project.

I am extremely glad to make use of this opportunity to express our deep sense of gratitude to Dr. Yasha Hasija, Department of Biotechnology, DTU, for lending me stimulating suggestions, innovative quality guidance and creative thinking. Her practicality, constructive criticism, constant encouragement and advice helped me in all stages of the project. Her scientific views and scientific approach will always be the source of motivation for me. I am grateful to her for the support she provided in doing things at my pace and for being patient with my mistakes.

Words are inadequate to express my deep sense of gratitude to my classmates and friends for their constant support and keeping my spirits up throughout the project and their companionship during my stay in DTU.

Last but not the least; I am grateful to my family for their unflinching support and guidance during the course of the study. I am lucky and proud of them.

Date:
Place:

Surya Kant Singh
2K13/Bio/17
M.Tech (Bioinformatics)
Delhi Technological University
New Delhi

CONTENTS

LIST OF FIGURES AND TABLES

Computer aided target Identification and drug design for pathogen *Chlamydophila psittaci*

Surya Kant Singh
Delhi Technological University, Delhi, India

ABSTRACT

Whole genome sequences of three strains of the human pathogen *Chlamydophila psittaci* were analyzed to identify common drug targets. Total number of 2926 protein sequences were studied from three strains; in which 2720 proteins were having more than 100 amino acids were selected; Further, 3 sequences were identified as non-human homologs which are common in all the three strains. Bifunctional 3, 4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein has been found to have a structural hit in Protein Data Bank with the ID 4I14 which can be used as the target protein. Ligands were identified based on the active sites and docked subsequently to find out the best ligand, N,N'-bis[(1-benzyl-4-piperidylidene)amino]butanediamide. This ligand has better binding energy than the natural ligand as well as the available drug molecules. Further, Lipinski's filters, various other physicochemical properties and toxicity studies were also done to check the bioavailability and toxicity of the top ligands.

Keywords: *Chlamydophila psittaci*, DEG, BLAST, Subtractive genomic approach, AutoDock.

INTRODUCTION

Chlamydophila psittaci

Chlamydophila psittaci is a gram negative bacterium which is a parasite. Its cells are of length 0.2 – 1.5 m. *Chlamydophila psittaci* causes an infective disease, psittacosis, within avian species as well as vertebrate species. The pathogen is generally detected in the infected birds' feathers, faecal matter, etc. This can also be transferred to human species through various ways like inhalation of dirt from infected bird. In the year 1930, a large epidemic of the infectious disease, psittacosis affected approximately 800 people that cause the closing off of *Chlamydophila psittaci* in US and Europe. A total number of 923 cases of the infectious disease, psittacosis are bruited to UN Centers for the prevention and management between the year 1988 and 2003.

The pathogen was detected through organism isolation by cell culture that involved the pathogenic cells to be scratched from the infection site. Improvement in PCR as well as ligase chain reaction amends this sample detection. These improved techniques include enzyme linked immunoassays and microscopy as well.

Chlamydophila psittaci is a harmful microorganism which is responsible for chlamydiosis in birds; and in humans, the psittacosis. Wild birds, humans, domesticated poultry and the cattle, etc are the potential hosts. *Chlamydophila psittaci* is transferred by contact, inhalation among the avian species and then shifted to mammalians. The disease, psittacosis within humans as well as avian species begins with an indication of flu but lead to pneumonia which may be sometimes life-threatening. The pathogenic strains remain quiet in avian species unless activated during the stress condition. The avian species are generally good vectors because of their mobility in spreading the infectious disease like psittacosis as they kill; get in contact with any type of diseased animals frequently.
At first, *Chlamydophila psittaci* was termed as *Chlamydia psittaci*. The beta sheets aere generally found in the outer membrane proteins of the pathogen, *Chlamydophila psittaci*, which function similar to the prion protein. These ducts are generally porous for the adenosine triphosphate and are the path through which the pathogen gets the benefit of nucleoside triphosphates. The pathogen is unable to metabolise the glucose. So, it is well understood that *Chlamydophila psittaci* gain adenosine triphosphate and other necessary residues generally from the host cell (Kaye *et al.*, 1996; Stead, 1997 and Sreevatsan *et al.*, 1997).
The pathogen is unable to perform the purines or pyrimidines synthesis, therefore the bacterium is dependent upon its host to collect pyrimidines or purines.

REVIEW OF LITERATURE

Chlamydophila psittaci is a harmful microorganism which is responsible for chlamydiosis in birds; and in humans, the psittacosis. Wild birds, humans, domesticated poultry and the cattle, etc are the potential hosts. *Chlamydophila psittaci* is transferred by contact, inhalation among the avian species and then shifted to mammalians. The disease, psittacosis within humans as well as avian species begins with an indication of flu but lead to pneumonia which may be sometimes life-threatening. The pathogenic strains remain quiet in avian species unless activated during the stress condition. The avian species are generally good vectors because of their mobility in spreading the infectious disease like psittacosis as they kill; get in contact with any type of diseased animals frequently.

Life cycle

In its life cycle, *Chlamydophila psittaci* meet with diverse conversion. It occurs in between the hosts as Elementary body. Elementary body is inactive in terms of life, still it can resist against the stresses and also capable enough to live without any host. It is transferred from diseased one to normal person or bird's lungs in the form of a little bead and becomes the main cause of disease. After that it is transferred by cells by the process of phagocytosis in the form of a pouch, termed as endosome. But still it cannot be ruined by the lysosome fusion. It translates into the reticulate body and starts to duplicate in the endosome. To complete its replication, reticulate bodies must use host's body. Now, the conversion of reticulate body to elementary body takes place and then the elementary body reaches the lung generally after cell death. Now, elementary bodies are capable to infect new cells of same as well as new host. Therefore, *Chlamydophila psittaci* life cycle lies between elementary body and reticulate body. The elementary body is incapable of replication, however capable enough to infect the new hosts, and the reticulate body, which can replicate, but cannot cause new infection.

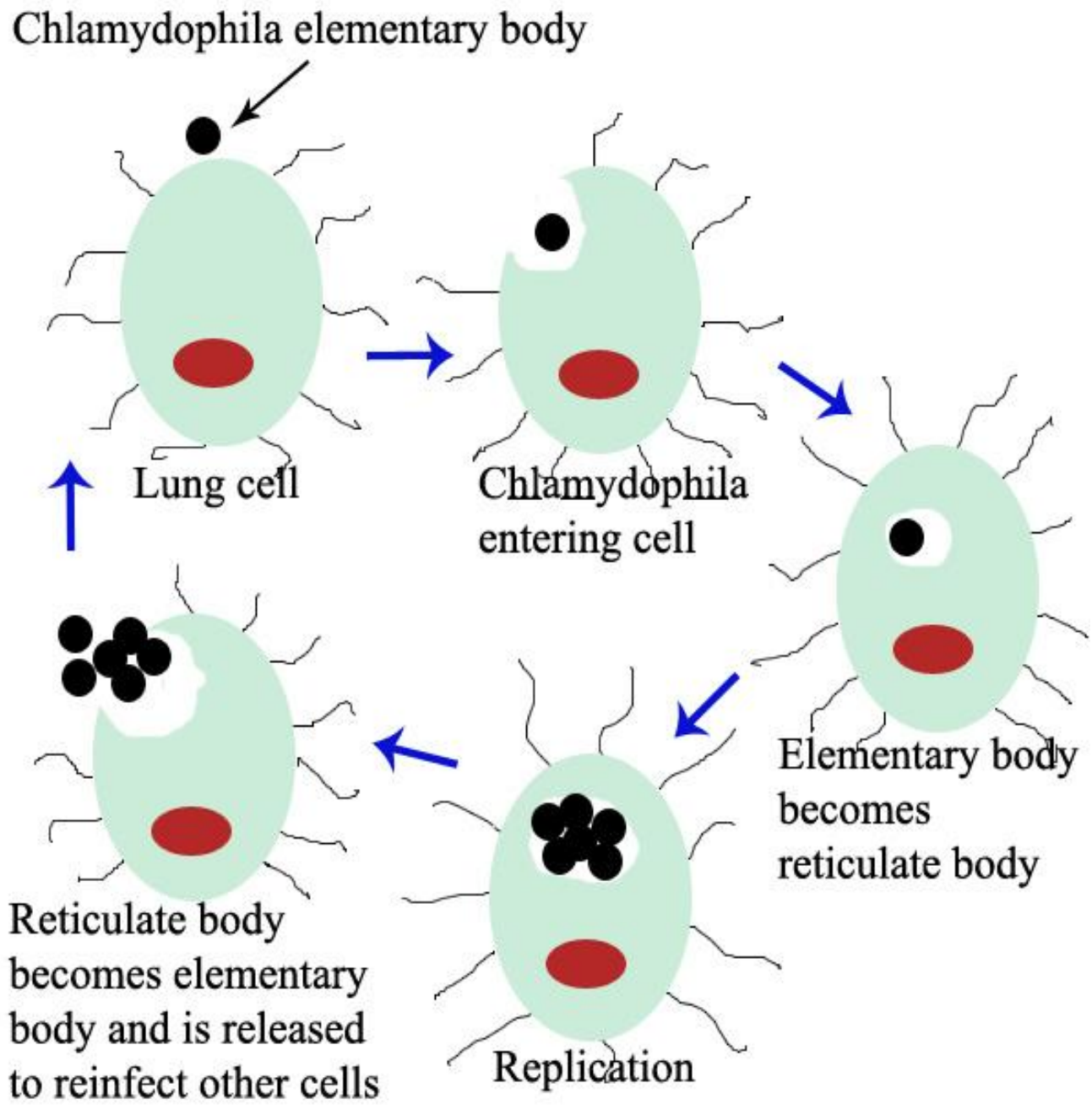

Fig1. Life cycle of *Chlamydophila psittaci*

Treatment

Drug of Choice

Tetracyclines are widely used as drugs of choice for the psittacosis treatment purpose, either the tetracycline hydrochloride or doxycycline (Harding *et al.,* 1962; Schaffner *et al.,* 1967; Verweij *et al.,* 1995; Jawetz, 1979; MacFarlane and Macrae, 1983). Although they can suppress pathogen growth and replication but the agent cannot be removed from host (Jawetz, 1979). The improvement can be seen in patient after twenty four to forty eight hours; some patients respond slowly or not the least bit (Schaffner *et al.,* 1967). Sensitivity to tetracycline is universal. And resistant mutants identified till date are very less (Jawetz, 1979; Yung *et al.,* 1988; MacFarlane *et al.,* 1983).
Children below the age of eight years ought to be treated with erythromycin unless they're severely sick or don't respond to erythromycin, during which case tetracycline or doxycycline ought to be used.

Researchers have done the comparative analysis of different metabolic pathways of human and *Chlamydophila psittaci*. Enzymes participated in *Chlamydophila psittaci* biochemical pathways taken from KEGG database were examined against proteins of humans, by protein BLAST search i.e., BLASTP against database of *Homo sapiens*. The threshold for e-value was 0.005. Enzymes having similarity below this cut-off value were separated out as potential targets.

The research was very helpful in identifying drug targets against the pathogen. A study of host and pathogen metabolic pathway by different bioinformatic methods can be helpful in studying several pathogens which are of medical concern. (Altschul *et al*., 1997). The fully genome sequence and inventions in structural biology furnishes activity for novel drug targets discovery. An international pool has different labs from all around the world. The main aims of this consortium is the identification of more than four hundred targets from the pathogen genome and then study them using their functional data. Possible and novel targets can be identified with the use of bioinformatics ways.

Subtractive genome approach

Genomics are often used for the evaluation of possible targets quality. It includes two measures:

1. Essentiality,

2. Selectivity.

The encoded proteins from the essential gene are of utmost important for any organism to survive (Kobayashi *et al.,* 2003; Mushegian *et al.,* 1996; Itaya, 1995). Galperin proposed that previous drug targets finding among qualified proteins are essential as well as specific for any pathogen (Galperin *et al.,* 1999). Database of Essential Genes, developed by Zhang, contains essential genes of different bacteria (Zhang *et al.,* 2004).

There is increase in number of target identification employing genomics application. A remarkable method known as Subtractive genomics (also known as Differential genome display) is suggested for identification of possible targets (Huynen *et al.*, 1997). It is depending upon the concept that parasite has usually smaller genome and a lower number of proteins are encoded in comparison with the nonparasites. Genes available only in the parasite and absent within the genome of a associated nonparasite are essential for the pathogen. And in addition to this, target shouldn't show any homolog within *Homo sapiens*. Therefore, those can be treated as the possible targets (Dutta *et al.*, 2006; Sakharkar *et al.*, 2004). This approach is used by various researchers to identify new drug targets in *Mycobacterum tuberculosis* (Anishetty *et al.*, 2005), *Clostridium perfringes* (Chhabra *et al.*, 2010), etc. These works had been perfectly done by using DEG (Zhang *et al.*, 2004).

Docking

Docking is a process by which the best configuration of binding molecules is determined. In this process, a complex structure is obtained having stable structure (Lengauer and Rarey, 1996). Knowledge of favored orientation in turn can be used in predicting strength of association between the two molecules and binding energy can be measured in terms of scoring function.

Docking is often used for predicting the binding of drug candidates to their target protein to predict the activity and affinity of the drug candidates. Thus, docking perform a very important role in the rational drug design (Kitchen *et al.*, 2004).

Docking approaches

There are 2 approaches popular in molecular docking community. First approach employs a method of matching where protein and ligand molecular surfaces are reported as complementary to each other (Goldman and Wipke, 2000; Meng *et al.*, 2004; Morris *et al.*, 1998). In the next approach, process of docking takes place and the interaction energy of protein-ligand complex is determined (Feig *et al.*, 2004).

METHODOLOGY

Online Tools and Database:

1. **NCBI**

NCBI is one of the part of National Institutes of Health branch (United States National Library of Medicine). There are a number of databases available in NCBI which are useful for biomedicine and biotechnology. Major databases are:

-GenBank: For DNA sequences,

-PubMed: For biomedical literature,

-Protein: For protein sequences, etc.

2. BLAST

Basic Local Alignment Search Tool detects the local regions of similarity between protein or nucleotide sequences. The BLAST program compares the query peptide or nucleotide sequences to its own sequence databases and then computes the matching statistical significance. BLAST is useful in inferring the functional as well as evolutionary relationships between the sequences. It is also useful in the identification of gene family members.

Five different types of BLAST program available:

-BLASTP: Comparison of protein sequence against the database of protein sequence.

-BLASTN: Comparison of nucleotide sequence against the database of nucleotide sequence.

-BLASTX: Comparison of six framed translational of nucleotide sequence against the database of protein sequence.

-TBLASTX: Comparison of six framed translation product of nucleotide sequence against the database of six framed translational product of a nucleotide sequence.

-TBLASTN: Comparison of protein sequence against the database of nucleotide sequence dynamically translated product in all six framed.

3. DEG

Essential genes can be defined as those genes required for any organism to live and so regarded as a base of life. DEG hosts records of presently accessible essential genomic elements, like non-coding RNAs and protein-coding genes, among eukaryotes, bacteria, and archaea. In the bacteria, essential genes comprise a minimal genome, forming functional modules set that play key roles in the multiple fields.

Users have four options to perform BLAST against DEG:

-A single gene: It gives raw BLAST output

-Multiple genes: It sum-ups homologous genes with the database of essential genes

-Annotated genome: Complete genome sequence is accepted with the gene annotation.

-Unannotated genome: It will find out protein-coding genes.

4. PDB

The Protein Data Bank is a repository for 3D biological molecules (nucleotides and proteins) structural data. The structural data found experimentally by NMR spectroscopy or X-ray crystallography and put in by the biochemists or biologists and are accessible freely on Internet.

Users can search in this database by PDB ID,macromolecule, author, sequence or ligands and download the required files in pdb format.

5. ZINC Database

The ZINC database comprises of commercially available chemical compounds. ZINC database is mainly used for virtual screening. ZINC is used by various scholars in research field as well as by the investigators in pharmaceutical companies or biotech companies.

Users can search in this database by IDs, SMILES, etc. The ZINC database finds the compounds based on similarity to the query compound. The output result of query can be downloaded in the mol2, sdf, SMILES, ddb (flexibase) format.

Other uses of the ZINC database:

-obtaining a compound for purchasing,

-obtaining compounds which can be used as a drug molecule, etc.

6. RASPD

RASPD is used to excluding the ligand molecules in the beginning based on physicochemical properties of ligands and active site of the target protein molecule. This tool searches based on various physicochemical properties like chemical formula, H-bond donors as well as acceptors, number of rings, etc. for every molecules.
Four methods are available for users in this tool:
-Method A: If protein-ligand complex information is available,
-Method B: If protein 3D structure is available but no information related to ligand is available,
-Method C: Customized Dataset,
-Method D: Customized Molecule.

7. Toxicity checker

As the name suggests, Toxicity Checker aims to identify whether any toxic substructure of the query compound is available or not and it also calculates the different properties of the

compounds. It is available freely to the users. It helps the scholars, companies and research institutes by allowing them to use the available tools online.
In this tool, users have two options to check for the toxic substructure in the compound. They can check it either by drawing the molecule or by providing molecule ID, SMILES, InChI, InChIKey.

Software:

1. **AUTODOCK 4.2.5**

Auto Dock 4.2.5 is a software used for the purpose of molecular docking of ligand to macromolecules like DNA, proteins, etc. There are two main programs in Auto Dock: (a) Auto Grid program for the identification of pre-computing grids, and (b) Auto Dock program for docking ligand molecule to a number of grids of the target protein.
Binding energy calculated is the combination of intermolecular and torsional energies.

2. **OSIRIS DATA WARRIOR**

OSIRIS Data Warrior is data analysis and visualization software. OSIRIS data warrior is helpful in predicting various physico-chemical properties and toxicity risk indication that must be optimized while designing pharmaceutically active compounds.

STEPS:

1. **Novel Drug Target Identification:**

Data Base Search Sequence

Whole genome sequences of three strains of *Chlamydophila psittaci* (*C.psitattaci 082DC60, C.psitattaci 02DC15, C.psitattaci 01DC11*) were downloaded from the National Center for Biotechnology Information (NCBI). These three strains are having 975, 978 and 973 respectively protein sequences. From the available complete genome sequences data, the encoded peptide sequences whose sequence length is greater than 100 amino acids were chosen. The selected protein sequences were then subjected to BLASTP against the DEG database to screen out essential genes. A random threshold of E-value, i.e., 10^-100and a score of 100 is applied as the threshold value. Then the screened out essential genes of *Chlamydophila psittaci* were submitted to BLASTP against the human genome to identify the non-human homologous proteins in the bacteria. The protein sequences showing homology were removed and the list of non-homologous was piled up for all the three strains of *Chlamydophila psittaci*. Further, non-human homologous proteins common to all the three strains were identified.
This protein is considered as common drug target for all the three *Chlamydophila psittaci* strains. The flow chart of the process is shown in the Figure-2.

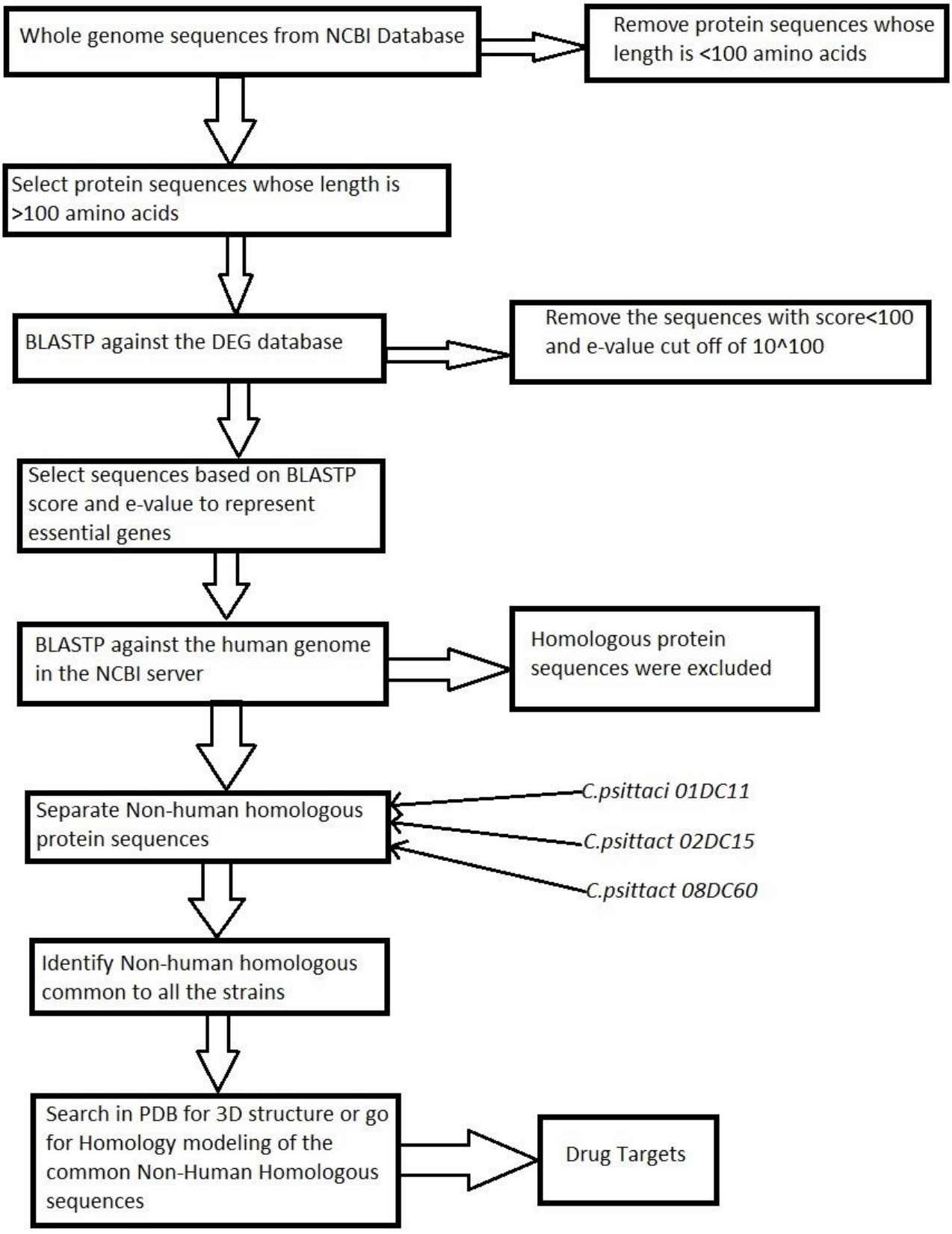

Fig2. *In silico* genomic approach for prediction of drug targets for *Chlamydophila psittaci*

Finally, those common non-human homologous protein sequences were searched in PDB (Protein Data Bank) to find the 3D structure of these proteins so as to find the drug target.

2. Docking of natural ligand with the target protein molecule

From Protein Data Bank, the natural ligand of our target protein molecule was found. And the natural ligand molecule was downloaded in mol2 format from ZINC database.

Then, we had used the docking software, AutoDock 4.2.5 for docking the natural ligand with the target protein molecule and binding energy had been recorded. Now we have to identify the ligands having lesser binding energy.

3. Finding ligands based on the target protein active sites

Method B of the online available tool RASPD had been used to identify various ligands. This method is useful when only target protein molecule is available. RASPD-Method B determines ligands depending upon the active grooves present on the target molecule. So, we obtained the library of ligands with their ZINC ID, IUPAC name and the 3D coordinates of the atoms involved.

Then, all the ligands were downloaded from the ZINC database and thus we had generated the virtual library of ligands. And in addition to ligands, the above mentioned drug molecules (doxycycline and tetracycline hydrochloride) were also downloaded from the ZINC database.

4. Docking of different ligands with the target protein molecule

Each of the ligands as well as drug molecules was docked one by one with the target molecule using AutoDock 4.2.5 software.

And obtained binding energies were noted down and the data is shown at Table 2. First 10 ligands having batter binding energy were also shown in fig. 15.

5. Checking Lipinski's filters

Lipinski's rule of five had been checked for top 10 ligand molecules, according to binding energy. These filters are obtained from one of the drug design tools at scfbio-IIT Delhi, "Lipinski Rule of five". The results are summarized in Table 3.

6. Cheking Toxicity of ligand molecules

a) Using Toxicity checker

Using SMILES sequence of the ligand, molecule structures was drawn and was checked for the toxic substructure. It is done for top 10 ligands and is shown in the Result section.

b) Using OSIRIS data warrior software

Using OSIRIS data warrior software, toxicity as well as physicochemical properties was obtained for top 10 ligand molecules and compared with the available drug molecules and summarized in Table 4.

RESULTS

1. Novel Drug Target Identification:

Whole genome sequences of three strains of *Chlamydophila psittaci* (*C.psitattaci 082DC60, C.psitattaci 02DC15, C.psitattaci 01DC11*) were analyzed to find out the drug target. Total 2926 protein sequences were found out from the three strains of *Chlamydophila psittaci* (*C.psitattaci 082DC60, C.psitattaci 02DC15, C.psitattaci 01DC11*) by an *in silico* genomic approach. These three strains are having 975, 978 and 973 respectively protein sequences. And out of these sequences, 2720 were having more than 100 residues in their coding sequence. This approach is applied based on the assumption that protein sequences having less than 100 residues can affect proteins catalytic activity and they can also form protein complex that affect their enzymatic activity [Yang *et al.*, 1991].
The BLASTP against DEG database results to equal number of essential genes across the strains. Each strain shows 526 essential gene hits against DEG database.

BLASTP RESULTS against DEG database for each strain are shown up in Fig. 3, 4, 5.

Organism:	Acinetobacter baylyi ADP1; Bacillus subtilis 168; Bacteroides fragilis 638R; Bacteroides thetaiotaomicron VPI-548 ; Burkholderia pseudomallei K96243; Burkholderia thailandensis E264; Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819; Caulobacter crescentus; Escherichia coli MG1655 I; Escherichia coli MG1655 II; Francisella novicida U112; Haemophilus influenzae Rd KW20; Helicobacter pylori 26695; Mycobacterium tuberculosis H37Rv; Mycobacterium tuberculosis H37Rv II; Mycobacterium tuberculosis H37Rv III; Mycoplasma genitalium G37; Mycoplasma pulmonis UAB CTIP; Porphyromonas gingivalis ATCC 33277; Pseudomonas aeruginosa PAO1; Pseudomonas aeruginosa UCBPP-PA14; Salmonella enterica serovar Typhi; Salmonella enterica serovar Typhi Ty2; Salmonella enterica serovar Typhimurium SL1344; Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S; Salmonella typhimurium LT2; Shewanella oneidensis MR-1; Sphingomonas wittichii RW1; Staphylococcus aureus N315; Staphylococcus aureus NCTC 8325; Streptococcus pneumoniae; Streptococcus sanguinis; Vibrio cholerae N16961;
Parameters:	deg.py -i /var/www/tubic/cgi-bin/blast/temp_seq/gosRO4jy6Y/seq.txt -db /var/www/tubic/cgi-bin/blast/temp_seq/gosRO4jy6Y/db -type seq -score 100 -email suryakantsingh123@gmail.com -job gosRO4jy6Y -F T -e 0.00001 -M BLOSUM62 -g T -v 100 -b 100 -blastprogram blastp
Total protein-coding genes in your sequence:	906 genes
In your sequence, the No. of genes having homologs with DEG:	526 genes.
In DEG, the No. of genes having homologs with your sequence:	9311 genes.

Your Query Protein	No. of homologs in DEG	DEG AC Number
gi|384451107|ref|YP_005663705.1| delta-aminolevulinic acid dehydratase [Chlamydia psittaci 01DC11]	12	DEG10130153; DEG10250102; DEG10190046; DEG10320050; DEG10280124; DEG10270093; DEG10200167; DEG10180069; DEG10020187; DEG10330177; DEG10120143; DEG10160174;
gi|384451110|ref|YP_005663708.1| fused transcript	8	DEG10170220; DEG10250200; DEG10050182; DEG10070073; DEG10140210; DEG10310029; DEG10020176; DEG10060232;

Fig3. *C.psittaci* 01DC11

Organism:	Acinetobacter baylyi ADP1; Bacillus subtilis 168; Bacteroides fragilis 638R; Bacteroides thetaiotaomicron VPI-548 ; Burkholderia pseudomallei K96243; Burkholderia thailandensis E264; Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819; Caulobacter crescentus; Escherichia coli MG1655 I; Escherichia coli MG1655 II; Francisella novicida U112; Haemophilus influenzae Rd KW20; Helicobacter pylori 26695; Mycobacterium tuberculosis H37Rv; Mycobacterium tuberculosis H37Rv II; Mycobacterium tuberculosis H37Rv III; Mycoplasma genitalium G37; Mycoplasma pulmonis UAB CTIP; Porphyromonas gingivalis ATCC 33277; Pseudomonas aeruginosa PAO1; Pseudomonas aeruginosa UCBPP-PA14; Salmonella enterica serovar Typhi; Salmonella enterica serovar Typhi Ty2; Salmonella enterica serovar Typhimurium SL1344; Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S; Salmonella typhimurium LT2; Shewanella oneidensis MR-1; Sphingomonas wittichii RW1; Staphylococcus aureus N315; Staphylococcus aureus NCTC 8325; Streptococcus pneumoniae; Streptococcus sanguinis; Vibrio cholerae N16961;
Parameters:	deg.py -i /var/www/tubic/cgi-bin/blast/temp_seq/QMHrYsXTOk/seq.txt -db /var/www/tubic/cgi-bin/blast/temp_seq/QMHrYsXTOk/db -type seq -score 100 -email suryakantsingh123@gmal.com -job QMHrYsXTOk -F T -e 0.00001 -M BLOSUM62 -g T -v 100 -b 100 -blastprogram blastp
Total protein-coding genes in your sequence:	909 genes
In your sequence, the No. of genes having homologs with DEG:	526 genes.
In DEG, the No. of genes having homologs with your sequence:	9313 genes.

Your Query Protein	No. of homologs in DEG	DEG AC Number
gi|384454036|ref|YP_005666631.1| delta-aminolevulinic acid dehydratase [Chlamydia psittaci 02DC15]	12	DEG10130153; DEG10250102; DEG10190046; DEG10320050; DEG10280124; DEG10270093; DEG10200167; DEG10180069; DEG10020187; DEG10330177; DEG10120143; DEG10160174;
gi|384454039|ref|YP_005666634.1| transcription elongation factor GreA domain protein [Chlamydia	9	DEG10170220; DEG10250200; DEG10050182; DEG10070073; DEG10140210; DEG10310029; DEG10020176; DEG10060232; DEG10080151;

Fig4. *C.psittaci* 02DC15

Organism:	Acinetobacter baylyi ADP1; Bacillus subtilis 168; Bacteroides fragilis 638R; Bacteroides thetaiotaomicron VPI-548 ; Burkholderia pseudomallei K96243; Burkholderia thailandensis E264; Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819; Caulobacter crescentus; Escherichia coli MG1655 I; Escherichia coli MG1655 II; Francisella novicida U112; Haemophilus influenzae Rd KW20; Helicobacter pylori 26695; Mycobacterium tuberculosis H37Rv; Mycobacterium tuberculosis H37Rv II; Mycobacterium tuberculosis H37Rv III; Mycoplasma genitalium G37; Mycoplasma pulmonis UAB CTIP; Porphyromonas gingivalis ATCC 33277; Pseudomonas aeruginosa PAO1; Pseudomonas aeruginosa UCBPP-PA14; Salmonella enterica serovar Typhi; Salmonella enterica serovar Typhi Ty2; Salmonella enterica serovar Typhimurium SL1344; Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S; Salmonella typhimurium LT2; Shewanella oneidensis MR-1; Sphingomonas wittichii RW1; Staphylococcus aureus N315; Staphylococcus aureus NCTC 8325; Streptococcus pneumoniae; Streptococcus sanguinis; Vibrio cholerae N16961;
Parameters:	deg.py -i /var/www/tubic/cgi-bin/blast/temp_seq/YxpjC8KMR1/seq.txt -db /var/www/tubic/cgi-bin/blast/temp_seq/YxpjC8KMR1/db -type seq -score 100 -email suryakantsingh123@gmal.com -job YxpjC8KMR1 -F T -e 0.00001 -M BLOSUM62 -g T -v 100 -b 100 -blastprogram blastp
Total protein-coding genes in your sequence:	905 genes
In your sequence, the No. of genes having homologs with DEG:	526 genes.
In DEG, the No. of genes having homologs with your sequence:	9313 genes.

Your Query Protein	No. of homologs in DEG	DEG AC Number
gi|384452083|ref|YP_005664680.1| delta-aminolevulinic acid dehydratase [Chlamydia psittaci 08DC60]	12	DEG10130153; DEG10250102; DEG10190046; DEG10320050; DEG10280124; DEG10270093; DEG10200167; DEG10180069; DEG10020187; DEG10330177; DEG10120143; DEG10160174;
gi|384452086|ref|YP_005664683.1| transcription elongation factor GreA domain protein [Chlamydia psittaci 08DC60]	9	DEG10170220; DEG10250200; DEG10050182; DEG10070073; DEG10140210; DEG10310029; DEG10020176; DEG10060232; DEG10080151;

Fig5. *C.psittaci* 08DC60

Out of these 526 protein sequences, the sequences having score > 100 and e-value less than 10^100 were taken as essential genes in each of the strains. 60 essential genes were identified for each strain of *Chlamydophila psittaci.*

Screenshots of some of the essential genes are shown in Fig.6, 7 and 8 for each strain:

>gi|384451133|ref|YP_005663731.1| DNA topoisomerase IV subunit B [Chlamydia psittaci 01DC11]
MATYTEASVVSLASLEHIRLRAGMYIGRLGDGSQVEDGIYTLFKEVVDNAIDEFIMGYGKTISIFTDDSTITVRDSGRGIPLGKMIDCVSKINTGA
KYTQDVFHFSVGLNGVGLKAVNALSEKFTVRSVRKKKYHYATFYKGVLQDSRQGSTKDPDGTEITFSPDPTIFTNFAFNDEFLRKKIRRYTYLHPG
LEIICNNEVFISQQGLLDLFKEEIPEETLYPPIAFQNSELSFLFSHLETHSERYFSFVNGQETLDGGSHLAAFKEAVVKGINEYFGKNFTSNDIRE
GIVGCIAIKIASPIFESQTKNKLGNTQIRSGIIKEVKSAIIQELKKNKSRSDLLLEKIKLNEKTRKNIQFIKQDLKDKQKKLHYKIPKLRDCKFHY
NERSLYGEASSIFVTEGESASASILSSRNPLTQAVFSLRGKPMNVFSLEEEKMYKNDELFYLATALGITKNSTQHLRYNKIILATDADVDGMHIRN
LLITFFLKTFLSVVENQHLFILETPLFKVRYKDTTLYCYSDQEKTQAIQKLGKKEAHIEVTRFKGLGEISPKEFKTTIGADMRLTPVTISSLESLD
SLLQFYMGKNTKERKQFIMDNLITNL
>gi|384451158|ref|YP_005663756.1| FeS assembly protein SufB [Chlamydia psittaci 01DC11]
MSQSIEDFLHNHEDYPYGFVTPIESEGLTRGLSEETIIKISQLRNEPSFILDFRLKAYQHWKKLQEPAWARLSYPTIDYDSIVYFSAPKQKNPLGR
LEEADPEILETFKKLGIPLDEQKRLLNVQNVAVDLVFDSVSIGTTFKEALDKAGVIFCSMNEAIREYPELVKKYLGSVVSHRDNYFAALNAAVFSD
GSFVYVPKGVRCPMEISTYFRINDKESGQFERTLIIAEDDSFVSYLEGCTAPSYSSNQLHAAVVELVAHERAVVRYSTVQNWFSGDKKTGKGGIYN
FVTKRGLCAGYKSKISWSQVEVGAAITWKYPSCILRGKESVGEFYSIALTSGKMQADTGTKMIHVGEKTTSTIVSKGISSEDSHNTFRSLVSISEG
AVGSRNHTQCDSMLIGRACGAYTDPKISVENSRSSVEHEATTSKLRADQLMYLRSRGLSAEEAVSLVVHGFCREIIEQLPLEFAREASKLLFVKLE
NSVG
>gi|384451184|ref|YP_005663782.1| aspartyl-tRNA synthetase [Chlamydia psittaci 01DC11]
MKYRTHRCNELSLSNVGERVRVSGWVHRYRNHGGVVFIDLRDRFGITQIVCREDEKPELHQLVDSVRSEWVLSIEGTVCRRLEGMENANLATGEIE
VDIEKVEILSKAKNLPFSISDDHIHVNEELRLEYRYLDMRRGQILDRLVCRHKVMMACRQYMDKQGFTEVVTPILGKSTPEGARDYLVPSRIYPGS
FYALPQSPQLFKQILMVGGLDRYFQIATCFRDEDLRADRQPEFAQIDIEMSFATPNDLFPIIEQLVVEMFAVQGIKIDLPLPRMTYQEAKDLYGTD
KPDLRFGLQLHDCRENAKEFSFSIFLDQLAQGGTIKGFCVPGGADISRKQLDVYTEFVKRYGAMGLVWIKKQENGIASNVAKFASEAVFQAMFADF
GAEDNDILLLIAAPEDVANQSLDHLRRLIAKERNLYNESQYNCVWITDFPLFAKEDGKICSEHHPFTSPLDEDIPLLDKDPLSVRSSSYDLVLNGY
EIASGSQRIHNADLQNKIFSILELSPESIKEKFGFFIDALSFGTPPHLGIALGLDRIMMVLTGAEGIREVIAFPKTQKAADLMMDAPSKIMTSQLK
ELNIKVTS
>gi|384451222|ref|YP_005663820.1| glyceraldehyde-3-phosphate dehydrogenase [Chlamydia psittaci 01DC11]
MKVVINGFGRIGRLVLRQLLKRNSSIEVVAVNDLVPGEALTYLFKHDSTHGRFPAEVSHENGCLVVDGRKIQLLAQSDVQKLPWKDLGVDVVIEST
GLFTKKEDAEKHLASGAKRVLITAPAKGDVPTFVMGVNEHKFDPEKDLVISNASCTTNCLAPLAKVLLDSFGIEEGLMTTVHAATATQSVVDGPSK

Fig6. *C.psittaci* 01DC11

>gi|384454062|ref|YP_005666657.1| DNA topoisomerase IV subunit B [Chlamydia psittaci 02DC15]
MATYTEASVVSLASLEHIRLRAGMYIGRLGDGSQVEDGIYTLFKEVVDNAIDEFIMGYGKTISIFTDDSTITVRDSGRGIPLGKMIDCVSKINTGA
KYTQDVFHFSVGLNGVGLKAVNALSEKFTVRSVRKKKYHYATFYKGVLQDSRQGSTKDPDGTEITFSPDPTIFTNFAFNDEFLRKKIRRYTYLHPG
LEIICNNEVFISQQGLLDLFKEEIPEETLYPPIAFQNSELSFLFSHLETHSERYFSFVNGQETLDGGSHLAAFKEAVVKGINEYFGKNFTSNDIRE
GIVGCIAIKIASPIFESQTKNKLGNTQIRSGIIKEVKSAIIQELKKNKSRSDLLLEKIKLNEKTRKNIQFIKQDLKDKQKKLHYKIPKLRDCKFHY
NERSLYGEASSIFVTEGESASASILSSRNPLTQAVFSLRGKPMNVFSLEEEKMYKNDELFYLATALGITKNSTQHLRYNKIILATDADVDGMHIRN
LLITFFLKTFLSVVENQHLFILETPLFKVRYKDTTLYCYSDQEKTQAIQKLGKKEAHIEVTRFKGLGEISPKEFKTFIGADMRLTPVTISSLESLD
SLLQFYMGKNTKERKQFIMDNLITNL
>gi|384454087|ref|YP_005666682.1| FeS assembly protein SufB [Chlamydia psittaci 02DC15]
MSQSIEDFLHNHEDYPYGFVTPIESEGLTRGLSEETIIKISQLRNEPSFILDFRLKAYQHWKKLQEPAWARLSYPTIDYDSIVYFSAPKQKNPLGR
LEEADPEILETFKKLGIPLDEQKRLLNVQNVAVDLVFDSVSIGTTFKEALDKAGVIFCSMNEAIREYPELVKKYLGSVVSHRDNYFAALNAAVFSD
GSFVYVPKGVRCPMEISTYFRINDKESGQFERTLIIAEDDSFVSYLEGCTAPSYSSNQLHAAVVELVAHERAVVRYSTVQNWFSGDKKTGKGGIYN
FVTKRGLCAGYKSKISWSQVEVGAAITWKYPSCILRGKESVGEFYSIALTSGKMQADTGTKMIHVGEKTTSTIVSKGISSEDSHNTFRSLVSISEG
AVGSRNHTQCDSMLIGRACGAYTDPKISVENSRSSVEHEATTSKLRADQLMYLRSRGLSAEEAVSLVVHGFCREIIEQLPLEFAREASKLLFVKLE
NSVG
>gi|384454112|ref|YP_005666707.1| aspartyl-tRNA synthetase [Chlamydia psittaci 02DC15]
MKYRTHRCNELSLSNVGERVRVSGWVHRYRNHGGVVFIDLRDRFGITQIVCREDEKPELHQLVDSVRSEWVLSIEGTVCRRLEGMENANLATGEIE
VDIEKVEILSKAKNLPFSISDDHIHVNEELRLEYRYLDMRRGQILDRLVCRHKVMMACRQYMDKQGFTEVVTPILGKSTPEGARDYLVPSRIYPGS
FYALPQSPQLFKQILMVGGLDRYFQIATCFRDEDLRADRQPEFAQIDIEMSFATPNDLFPIIEQLVVEMFAVQGIKIDLPLPRMTYQEAKDLYGTD
KPDLRFGLQLHDCREHAKEFSFSIFLDQLAQGGTIKGFCVPGGADISRKQLDVYTEFVKRYGAMGLVWIKKQENGIASNVAKFASEAVFQAMFADF
GAEDNDILLLIAAPEDVANQSLDHLRRLIAKERNLYNESQYNCVWITDFPLFAKEDGKICSEHHPFTSPLDEDIPLLDKDPLSVRSSSYDLVLNGY
EIASGSQRIHNADLQNKIFSILELSPESIKEKFGFFIDALSFGTPPHLGIALGLDRIMMVLTGAEGIREVIAFPKTQKAADLMMDAPSKIMTSQLK
ELNIKVTS
>gi|384454150|ref|YP_005666745.1| glyceraldehyde-3-phosphate dehydrogenase [Chlamydia psittaci 02DC15]
MKVVINGFGRIGRLVLRQLLKRNSSIEVVAVNDLVPGEALTYLFKHDSTHGRFPAEVSHENGCLVVDGRKIQLLAQSDVQKLPWKDLGVDVVIEST
GLFTKKEDAEKHLASGAKRVLITAPAKGDVPTFVMGVNEHKFDPEKDLVISNASCTTNCLAPLAKVLLDSFGIEEGLMTTVHAATATQSVVDGPSK
KDWRGGRGAFQNIIPASTGAAKAVALCLPELKNKLTGMAFRVPVADVSVVDLTVRLQKSTTYEEICKVVKEASETHLSGILGYTDQEVVSSDFIGC
EYSSIFDAGAGIALTDRFFKLVAWYDNEIGYATRIVDLLEYVAKNSK

Fig7. *C.psittaci* 02DC15

>gi|384452109|ref|YP_005664706.1| DNA topoisomerase IV subunit B [Chlamydia psittaci 08DC60]
MATYTEASVVSLASLEHIRLRAGMYIGRLGDGSQVEDGIYTLFKEVVDNAIDEFIMGYGKTISIFTDDSTITVRDSGRGIPLGKMIDCVSKINTGA
KYTQDVFHFSVGLNGVGLKAVNALSEKFTVRSVRKKKYHYATFYKGVLQDSRQGSTKDPDGTEITFSPDPTIFTNFAFNDEFLRKKIRRYTYLHPG
LEIICNNEVFISQQGLLDLFKEEIPEETLYPPIAFQNSELSFLFSHLETHSERYFSFVNGQETLDGGSHLAAFKEAVVKGINEYFGKNFTSNDIRE
GIVGCIAIKIASPIFESQTKNKLGNTQIRSGIIKEVKSAIIQELKKNKSRSDLLLEKIKLNEKTRKNIQFIKQDLKDKQKKLHYKIPKLRDCKFHY
NERSLYGEASSIFVTEGESASASILSSRNPLTQAVFSLRGKPMNVFSLEEEKMYKNDELFYLATALGITKNSTQHLRYNKIILATDADVDGMHIRN
LLITFFLKTFLSVVENQHLFILETPLFKVRYKDTTLYCYSDQEKTQAIQKLGKKEAHIEVTRFKGLGEISPKEFKTTIGADMRLTPVTISSLESLD
SLLQFYMGKNTKERKQFIMDNLITNL
>gi|384452134|ref|YP_005664731.1| FeS assembly protein SufB [Chlamydia psittaci 08DC60]
MSQSIEDFLHNHEDYPYGFVTPIESEGLTRGLSEETIIKISQLRNEPSFILDFRLKAYQHWKKLQEPAWARLSYPTIDYDSIVYFSAPKQKNPLGR
LEEADPEILETFKKLGIPLDEQKRLLNVQNVAVDLVFDSVSIGTTFKEALDKAGVIFCSMNEAIREYPELVKKYLGSVVSHRDNYFAALNAAVFSD
GSFVYVPKGVRCPMEISTYFRINDKESGQFERTLIIAEDDSFVSYLEGCTAPSYSSNQLHAAVVELVAHERAVVRYSTVQNWFSGDKKTGKGGIYN
FVTKRGLCAGYKSKISWSQVEVGAAITWKYPSCILRGKESVGEFYSIALTSGKMQADTGTKMIHVGEKTTSTIVSKGISSEDSHNTFRSLVSISEG
AVGSRNHTQCDSMLIGRACGAYTDPKISVENSRSSVEHEATTSKLRADQLMYLRSRGLSAEEAVSLVVHGFCREIIEQLPLEFAREASKLLFVKLE
NSVG
>gi|384452160|ref|YP_005664757.1| aspartyl-tRNA synthetase [Chlamydia psittaci 08DC60]
MKYRTHRCNELSLSNVGERVRVSGWVHRYRNHGGVVFIDLRDRFGITQIVCREDEKPELHQLVDSVRSEWVLSIEGTVCRRLEGMENANLATGEIE
VDIEKVEILSKAKNLPFSISDDHIHVNEELRLEYRYLDMRRGQILDRLVCRHKVMMACRQYMDKQGFTEVVTPILGKSTPEGARDYLVPSRIYPGS
FYALPQSPQLFKQILMVGGLDRYFQIATCFRDEDLRADRQPEFAQIDIEMSFATPNDLFPIIEQLVVEMFAVQGIKIDLPLPRMTYQEAKDLYGTD
KPDLRFGLQLHDCREHAKEFSFSIFLDQLAQGGTIKGFCVPGGADISRKQLDVYTEFVKRYGAMGLVWIKKQENGIASNVAKFASEAVFQAMFADF
GAEDNDILLLIAAPEDVANQSLDHLRRLIAKERNLYNESQYNCVWITDFPLFAKEDGKICSEHHPFTSPLDEDIPLLDKDPLSVRSSSYDLVLNGY
EIASGSQRIHNADLQNKIFSILELSPESIKEKFGFFIDALSFGTPPHLGIALGLDRIMMVLTGAEGIREVIAFPKTQKAADLMMDAPSKIMTSQLK
ELNIKVTS
>gi|384452198|ref|YP_005664795.1| glyceraldehyde-3-phosphate dehydrogenase [Chlamydia psittaci 08DC60]
MKVVINGFGRIGRLVLRQLLKRNSSIEVVAVNDLVPGEALTYLFKHDSTHGRFPAEVSHENGCLVVDGRKIQLLAQSDVQKLPWKDLGVDVVIEST
GLFTKKEDAEKHLASGAKRVLITAPAKGDVPTFVMGVNEHKFDPEKDLVISNASCTTNCLAPLAKVLLDSFGIEEGLMTTVHAATATQSVVDGPSK
KDWRGGRGAFQNIIPASTGAAKAVALCLPELKNKLTGMAFRVPVADVSVVDLTVRLQKSTTYEEICKVVKEASETHLSGILGYTDQEVVSSDFIGC
EYSSIFDAGAGIALTDRFFKLVAWYDNEIGYATRIVDLLEYVAKNSK

Fig8. *C.psittaci* 08DC60

Human non homologous gene:

Comparison of the essential genes of the bacterial strain with human genome illustrates ----- number of human non homolog essential genes in the genome of three strains of the pathogen. All essential genes of each strain can be put in the protein blast of NCBI which were aligned against human to find non human homologous sequences. Sequences showing homology with any of the human proteins are filtered out from total sequences determined. This time *Chlamydophila psittaci 02DC15* and *Chlamydophila psittaci 08DC60* shows 5 non human homologous essential protein sequences but the *Chlamydophila psittaci 01DC11* shows only 4 non human homologous protein sequences. These non human homologous protein sequences which are essential for the pathogen can be putative targets. And these proteins inhibition can inhibit the pathogen without affecting the host metabolism.

Fig9. NCBI BLASTP result for "bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein".

Screenshot of **NCBI BLASTP** result is shown in Fig. 9 for "**bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein"** which is one of the non-human homologous among all the three strains.

CHARACTER	*C. psittaci 01DC11*	*C. psittaci 02DC15*	*C. psittaci 08DC60*
TOTAL SEQUENCES	975	978	973
SEQUENCES HAVE >100 RESIDUES	906	909	905
DEG BLASTP RESULTS	526	526	526
ESSENTIAL GENE	60	60	60
HUMAN HOMOLOG	56	55	55
NON HUMAN HOMOLOG	4	5	5

Table1. Variation of the various genes across different strains of *C.psittaci* .

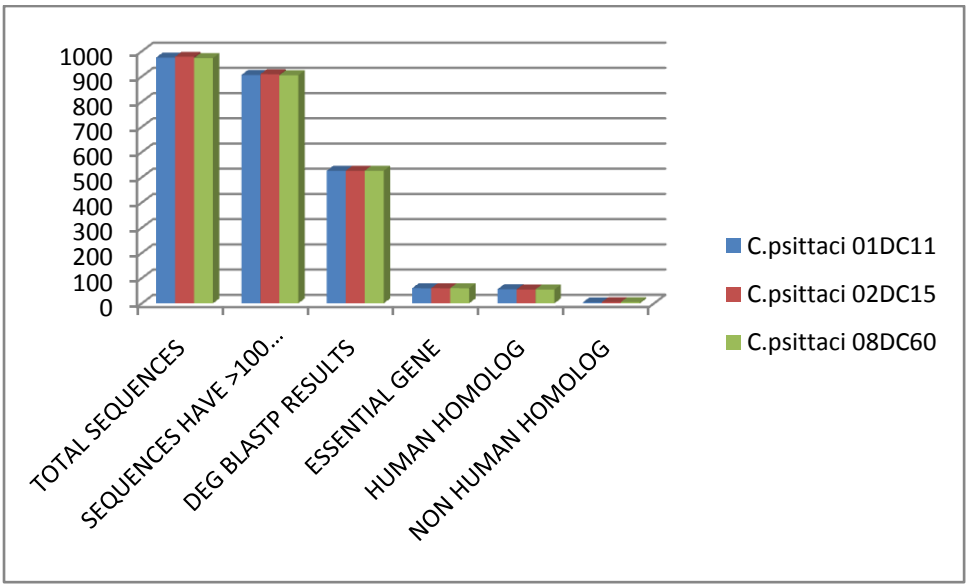

Fig10. Shows that blue, red and green bar represent the analysis of sequences found in this experiment for all the three strains.

The **common** non-human homologous protein sequences in all the three strains are:

- FeS assembly protein SufB
- bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein
- RNA polymerase sigma factor

These common non-human homologous protein sequences were searched in **Protein Data Bank (PDB)** to find the 3D structure of these proteins. Out of these three proteins I got one structure hit for "**bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein**" with the **PDB ID 4I14.**

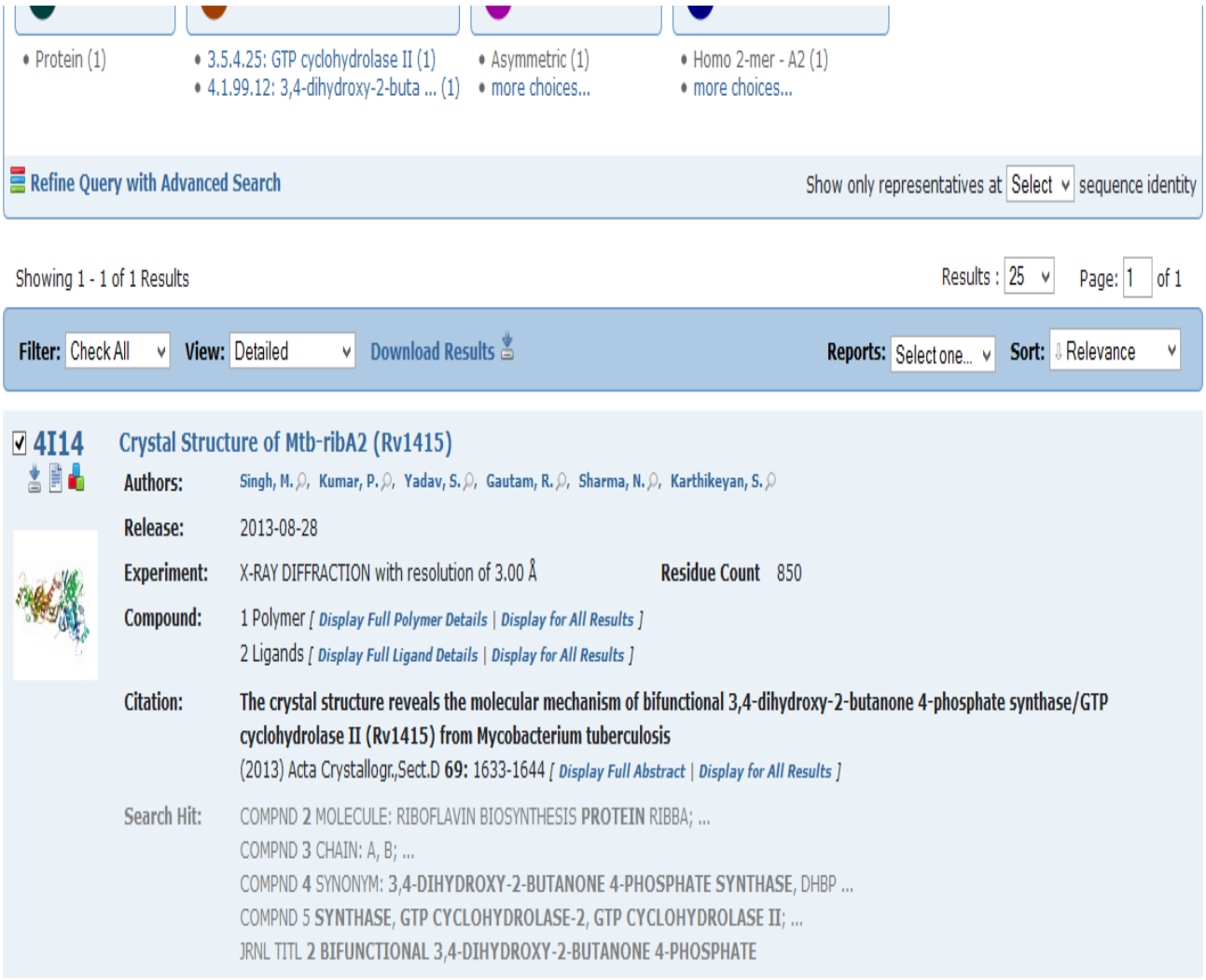

Fig11. Structure hit for "bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II" protein in PDB.

For the other two common proteins, no 3D structure is available in PDB. So, the protein sequence with PDB ID 4I14 can be used as the drug target molecule against the pathogen *Chlamydophila psittaci.*

3D structure view of the hit molecule for "**bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein" (PDB ID 4I14)** is shown in fig.12.

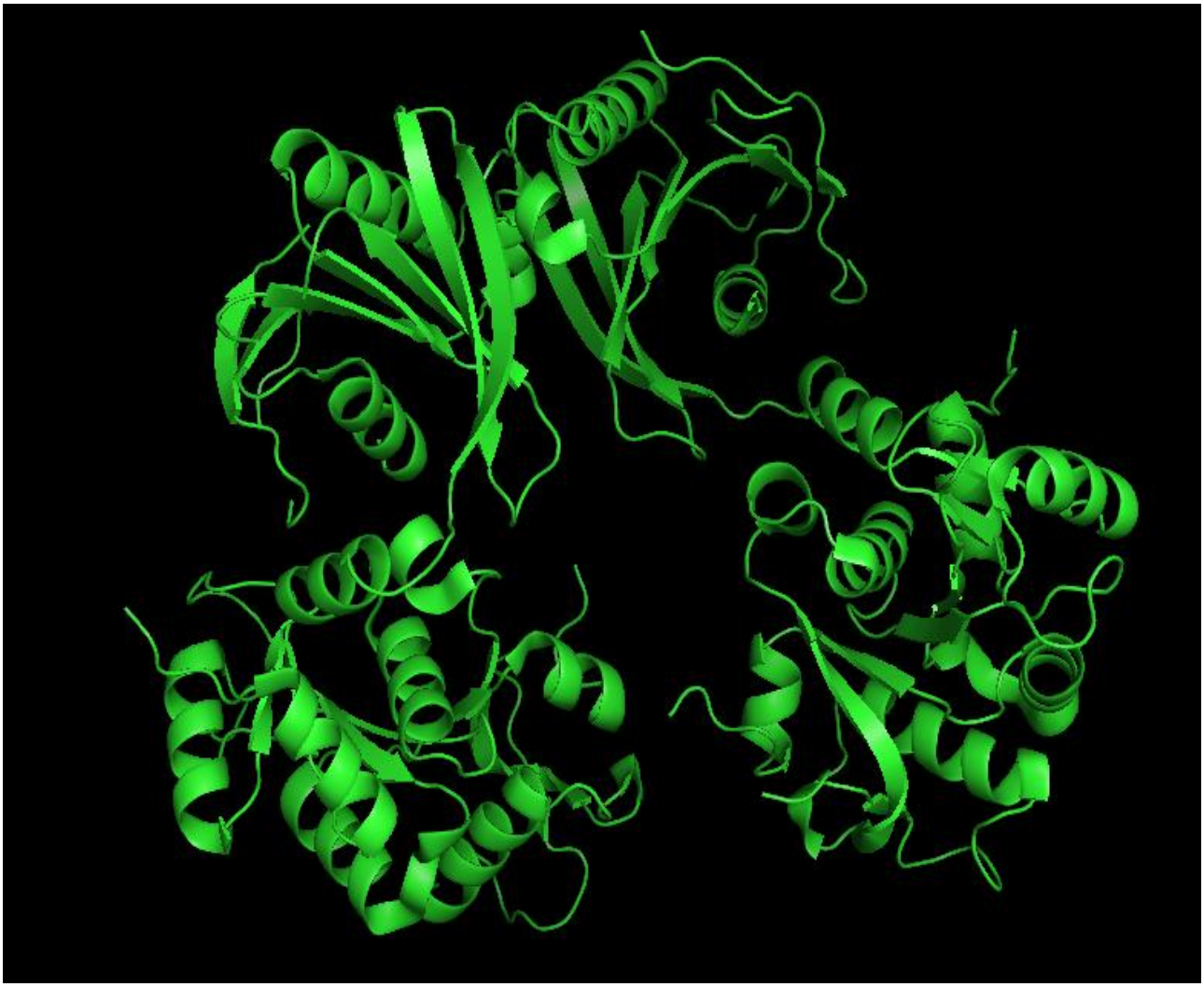

Fig12. 3D structure view of the protein with PDB ID 4I14.

2. Docking of natural ligand with the target protein molecule

The natural ligand identified for protein with PDB ID 4I14 is Sulfate ion. And ZINC ID of Sulfate ion is ZINC06827621. After docking of natural ligand with the target protein results into the complex with binding energy -2.79 kcal/mol. Now we have to identify the ligands having lesser binding energy.

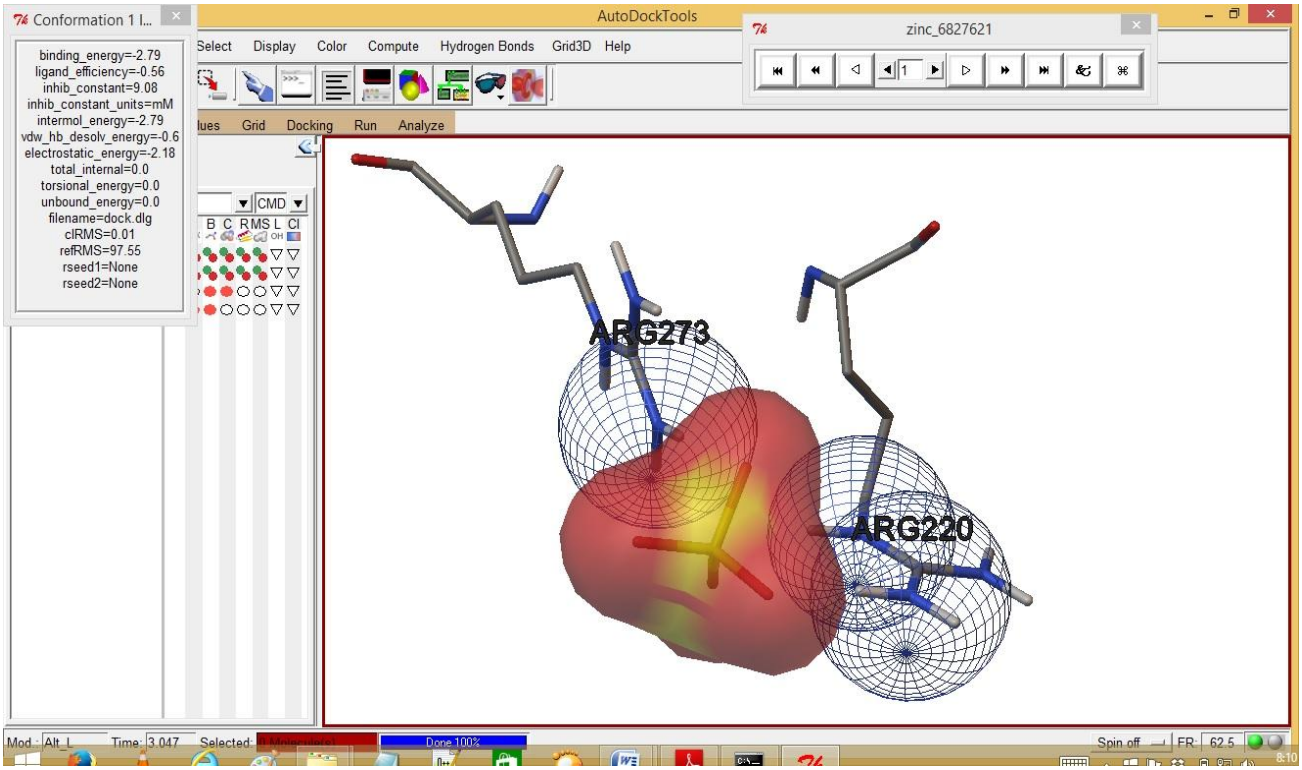

Fig13. Docking result of natural ligand with target protein.

3. Finding virtual library of ligands and docking results

Submitting of our target protein to the Method B of RASPD results into a virtual library of 100 ligands with their ZINC ID, IUPAC name and involved atomic 3D co-ordinates. The docking results of each ligands with the target protein molecule are shown in Table 2 and screenshots for top 10 ligands are also shown below. Then, all the ligands were downloaded from the ZINC database and thus we had generated the virtual library of ligands.

Docking of doxycycline and tetracycline hydrochloride results into the binding energies of -3.33 and -3.87 respectively in terms of kcal/mol shown in figures.

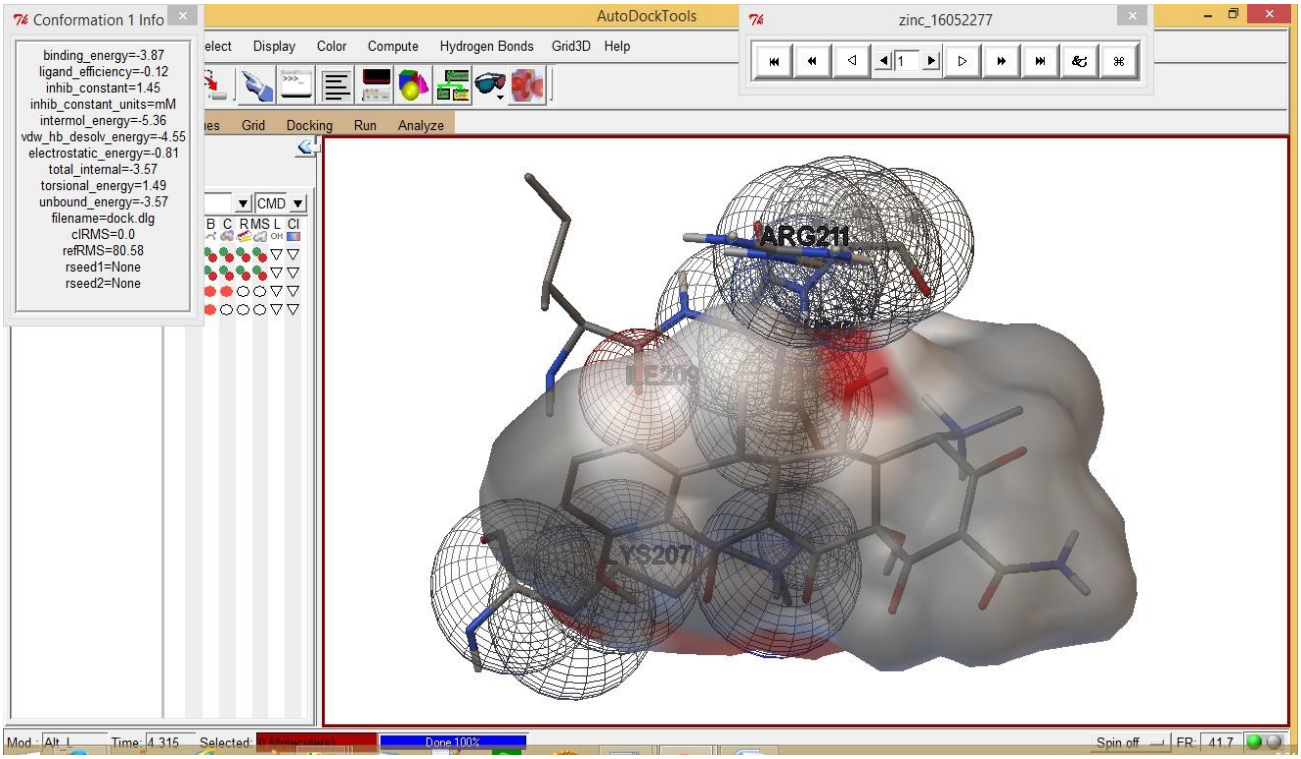

Fig14a. Docking result of Doxycycline with the target protein.

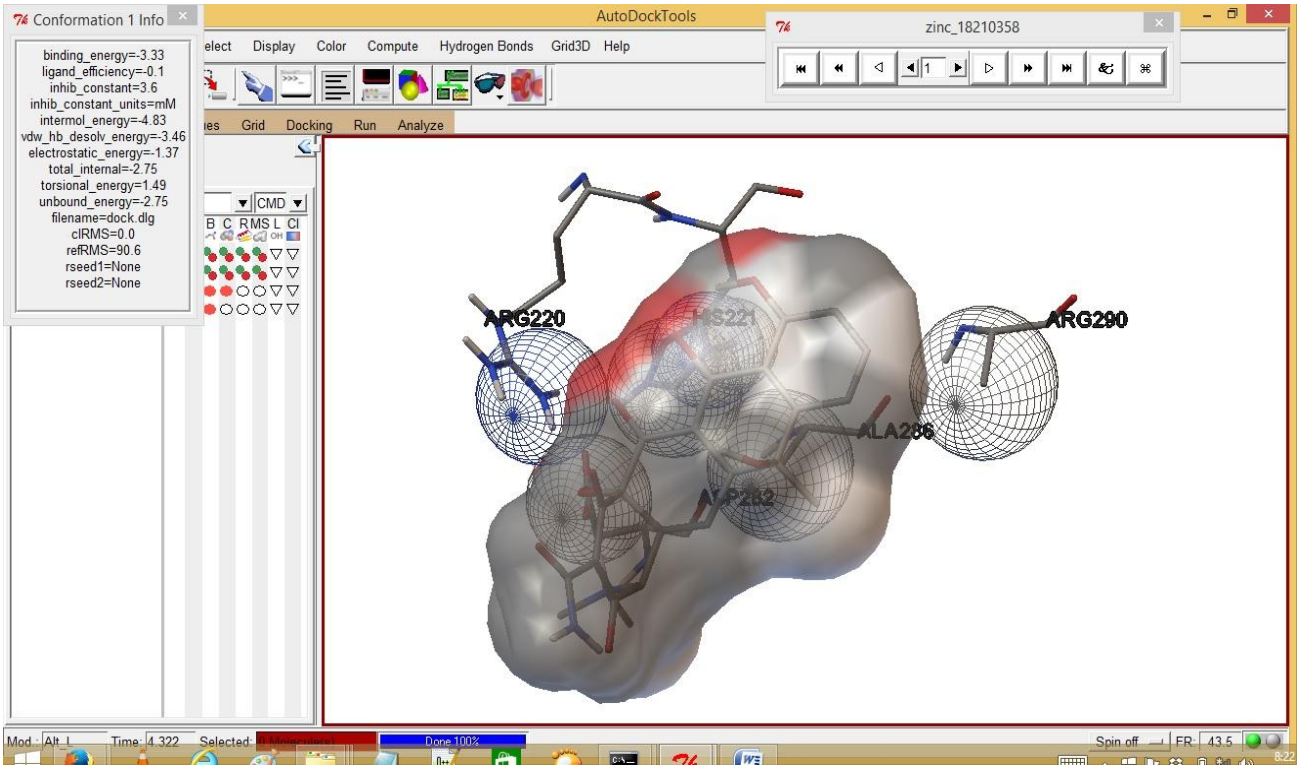

Fig14b. Docking result of Tetracycline hydrochloride with the target protein.

Rank	Ligand (ZINC ID)	Binding energy
1	zinc_4113772	-8.86
2	zinc_11963881	-8.09
3	zinc_12243261	-7.8
4	zinc_12191590	-7.75
5	zinc_12453654	-7.25
5	zinc_19813096	-7.25
7	zinc_15001976	-6.86
8	zinc_11881196	-6.84
9	zinc_11786333	-6.74
10	zinc_2811191	-6.72
11	zinc_11790367	-6.71
12	zinc_20919473	-6.67
12	zinc_2895958	-6.67
14	zinc_19842927	-6.63
15	zinc_3664754	-6.6
16	zinc_19811543	-6.47
17	zinc_11783578	-6.42
18	zinc_11974393	-6.34
19	zinc_12333572	-6.17
20	zinc_11882026	-6.08
21	zinc_1807072	-6.02
22	zinc_11913293	-6
23	zinc_19835705	-5.97
24	zinc_20451377	-5.95
25	zinc_19758600	-5.94
26	zinc_20601870	-5.91
27	zinc_12455413	-5.89
28	zinc_12242812	-5.77
29	zinc_32603782	-5.76
29	zinc_11841107	-5.76
31	zinc_4122191	-5.75
32	zinc_15005335	-5.72
33	zinc_11784200	-5.66
34	zinc_15823058	-5.65
35	zinc_1158015	-5.61
36	zinc_11784161	-5.59
36	zinc_12242780	-5.59
38	zinc_14992522	-5.47
38	zinc_20909250	-5.47
40	zinc_2760064	-5.38
40	zinc_12190465	-5.38

42	zinc_12247658	-5.33
43	zinc_14990472	-5.17
44	zinc_12245742	-5.16
45	zinc_19797529	-5.02
46	zinc_15003908	-4.99
47	zinc_12576410	-4.98
47	zinc_20562945	-4.98
49	zinc_2745710	-4.96
50	zinc_13081002	-4.95
51	zinc_11840986	-4.84
52	zinc_12050585	-4.76
53	zinc_14885566	-4.74
54	zinc_2952200	-4.72
55	zinc_2985323	-4.61
56	zinc_12247323	-4.56
57	zinc_3143011	-4.5
57	zinc_14981325	-4.5
59	zinc_808592	-4.49
60	zinc_19774479	-4.48
61	zinc_20999348	-4.47
62	zinc_3656658	-4.45
63	zinc_2822264	-4.28
64	zinc_19830686	-4.26
65	zinc_12217375	-4.12
66	zinc_16480347	-4.09
67	zinc_20508667	-4.04
68	zinc_15856729	-3.97
69	zinc_11980933	-3.94
70	zinc_11783262	-3.88
71	zinc_9041389	-3.87
72	zinc_8769779	-3.86
73	zinc_8892130	-3.85
74	zinc_9226916	-3.78
75	zinc_2836173	-3.71
76	zinc_19114279	-3.68
77	zinc_12524375	-3.53
78	zinc_2831975	-3.51
79	zinc_9191993	-3.48
80	zinc_15441607	-3.39
81	zinc_675736	-3.34
82	zinc_1139950	-3.3
83	zinc_2109070	-3.27
83	zinc_8935093	-3.27

85	zinc_6239462	-3.25
86	zinc_8442648	-3.21
87	zinc_12419770	-3.2
88	zinc_16248054	-3.18
89	zinc_19805326	-3.16
90	zinc_16275934	-3.14
91	zinc_3877717	-3.04
92	zinc_19857639	-2.95
93	zinc_9261062	-2.93
94	zinc_22147048	-2.91
95	zinc_2851420	-2.87
96	zinc_1794178	-2.69
97	zinc_20494835	-2.55
98	zinc_22064237	-2.5
99	zinc_17195094	-1.62
100	zinc_16667348	-1.08

Table2. Binding energy of different ligands with the target protein

AutoDock result of first 10 ligands:

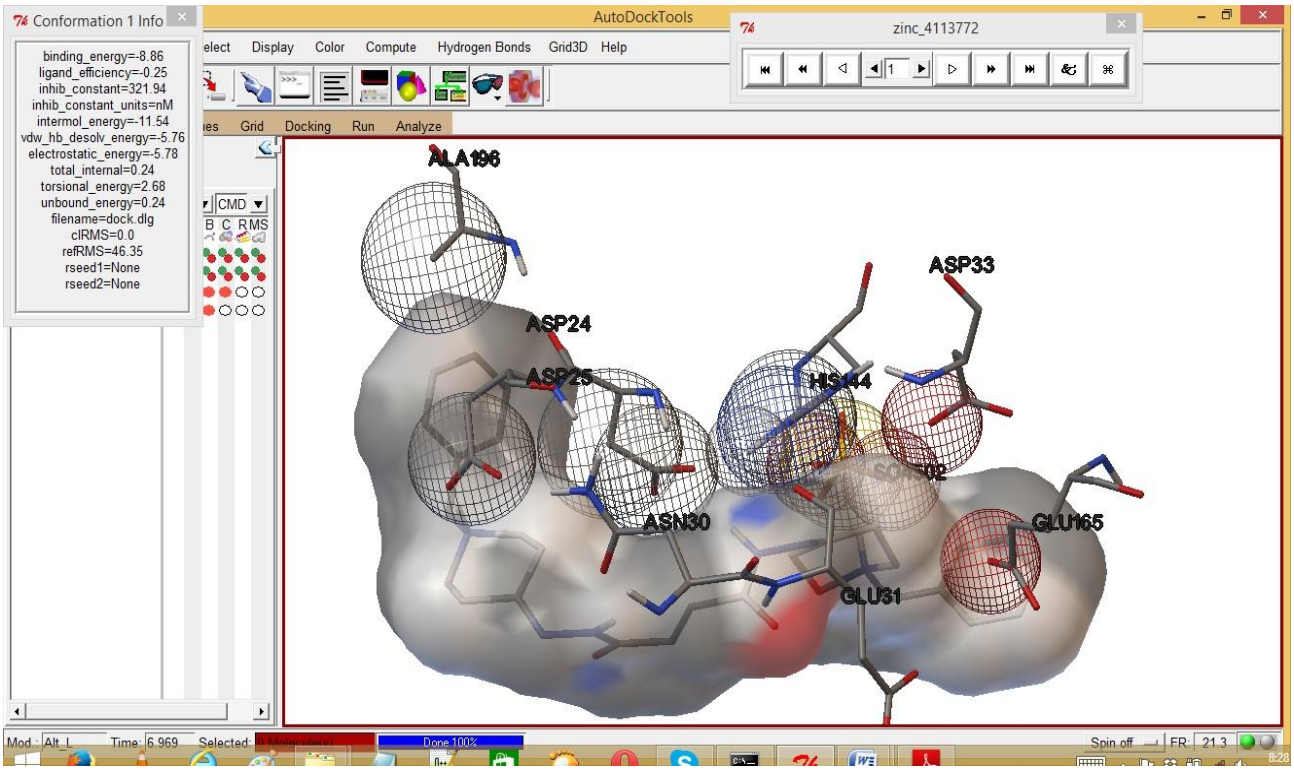

Fig15a. Docking with ligand 1 (ZINC04113772).

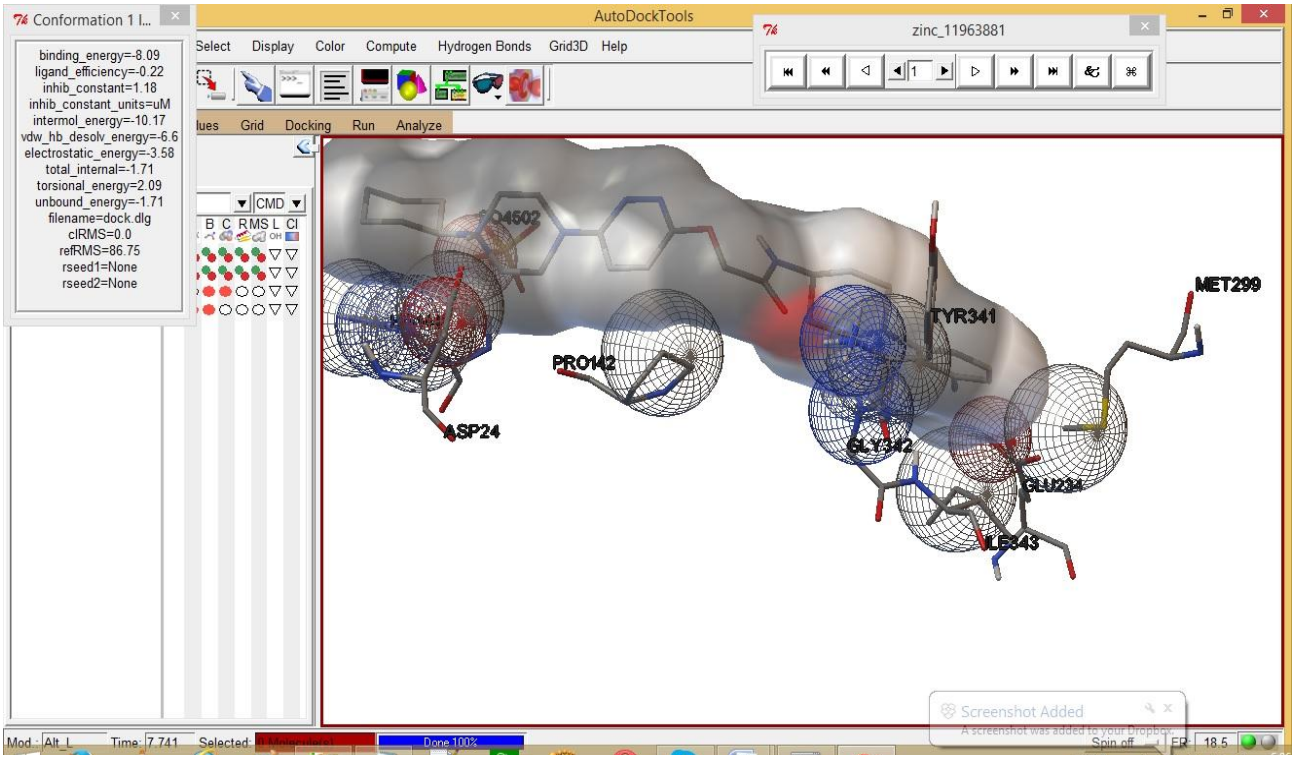

Fig15b. Docking with ligand 2 (ZINC11963881)

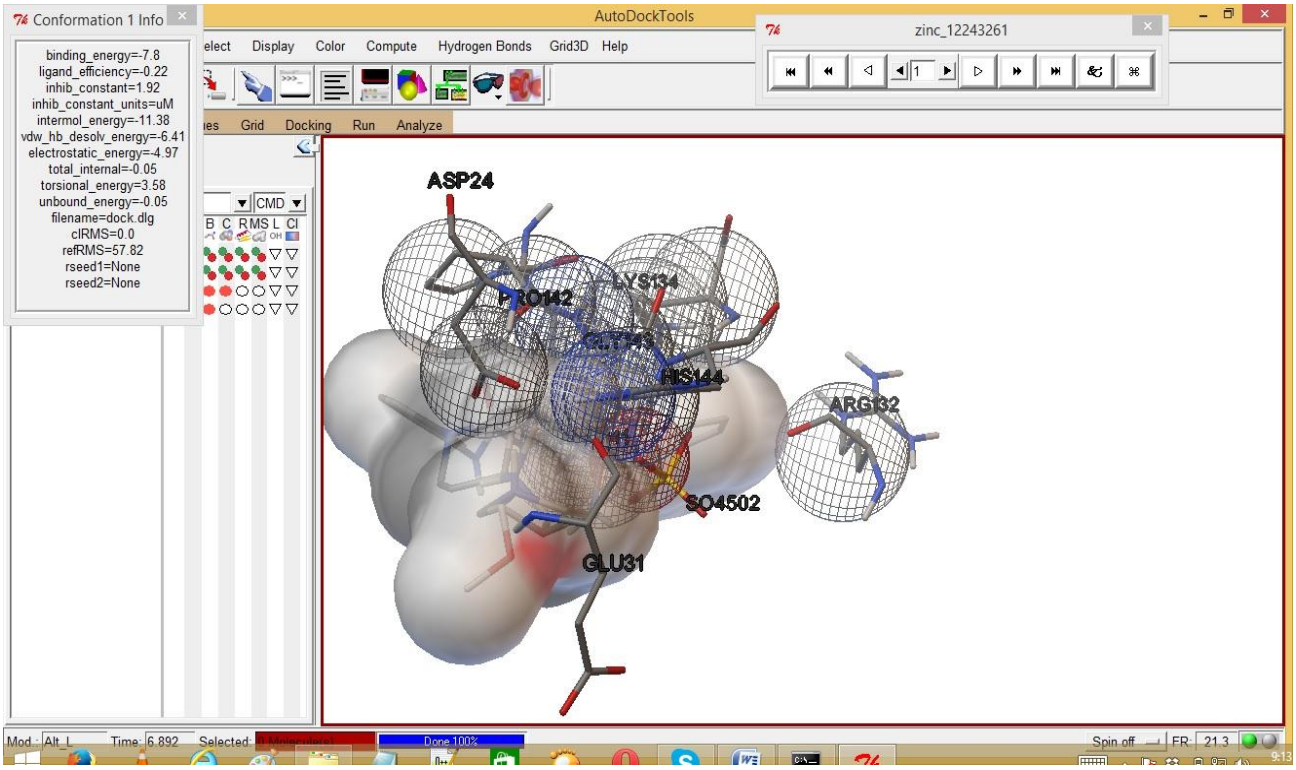

Fig15c. Docking with ligand 3 (ZINC12243261)

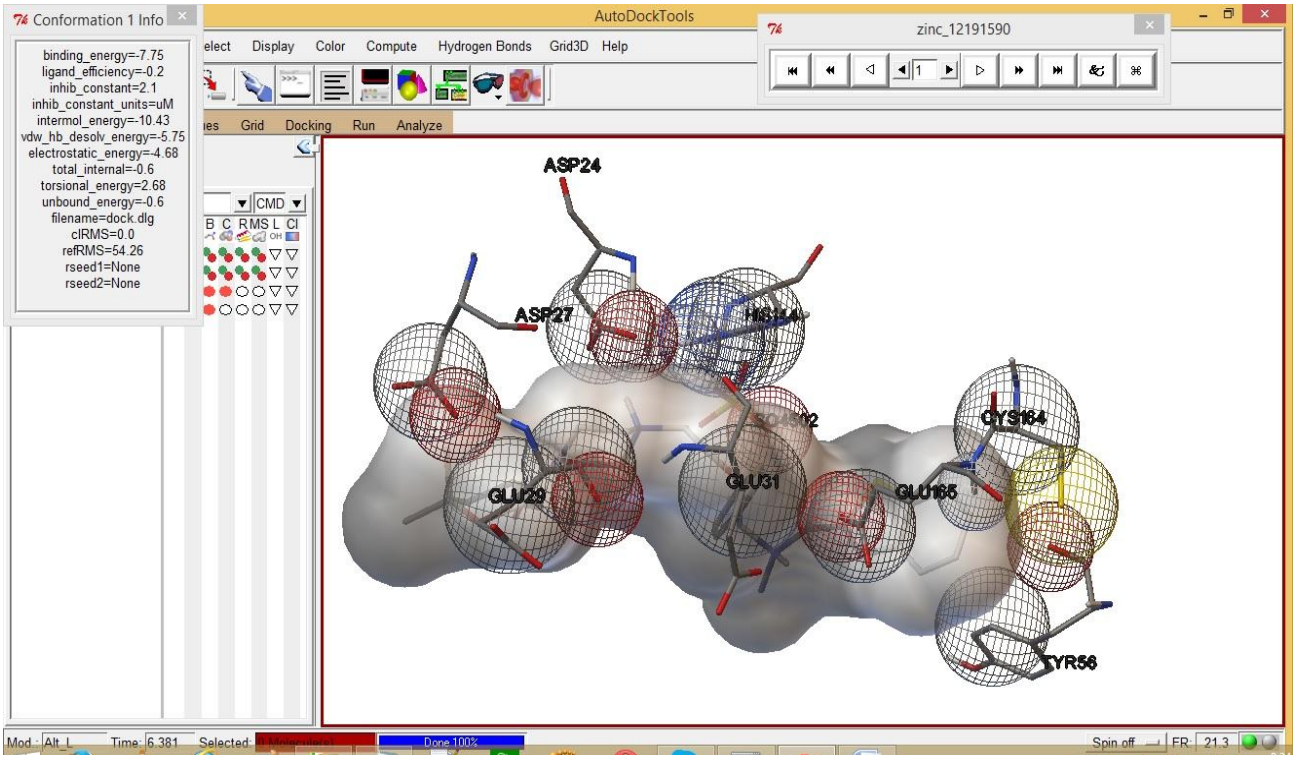

Fig15d. Docking with ligand 4 (ZINC12191590).

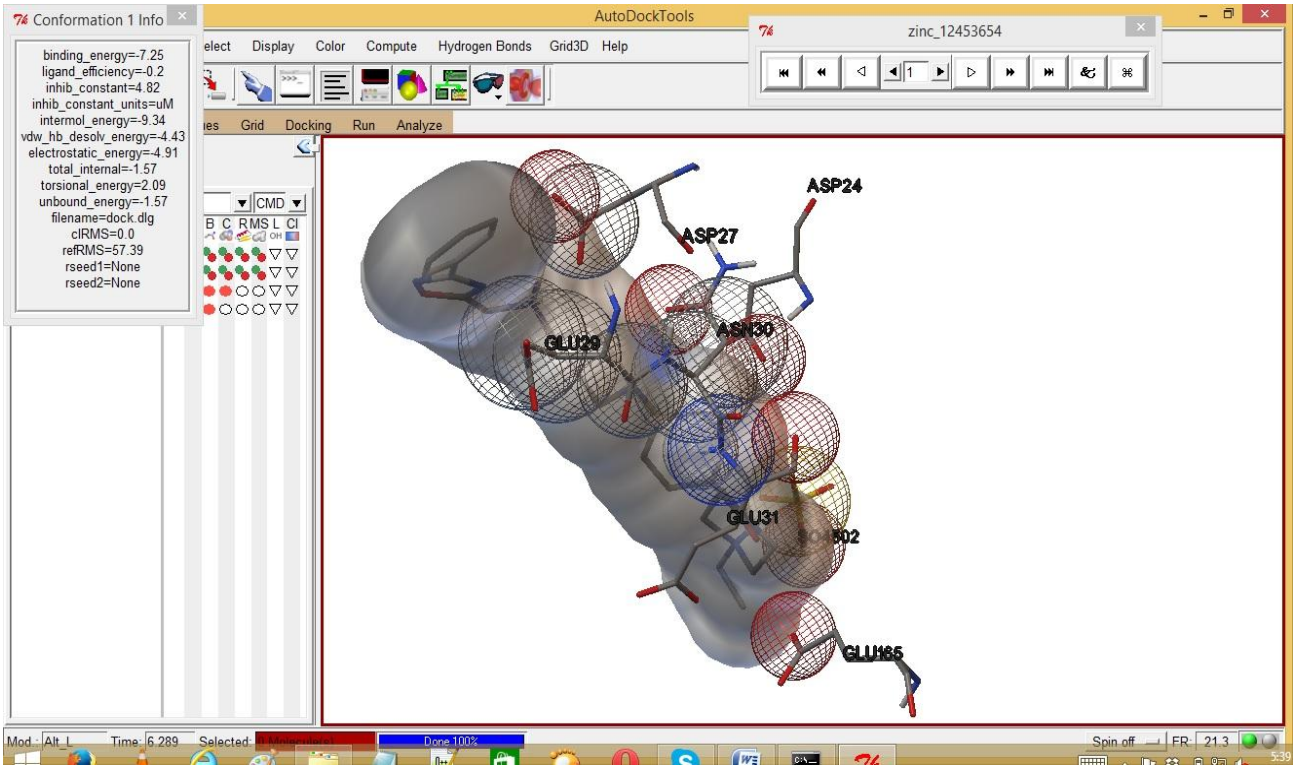

Fig15e. Docking with ligand 5 (ZINC12453654).

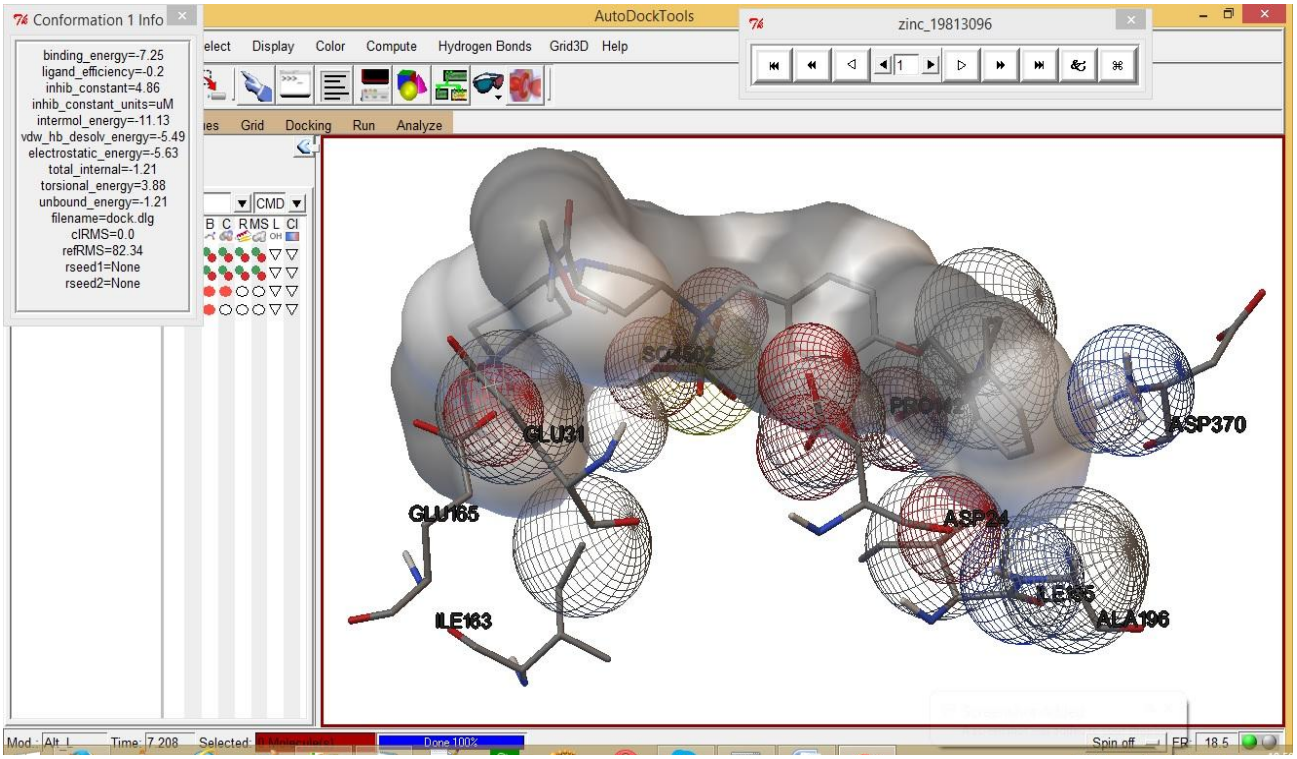

Fig15f. Docking with ligand 6 (ZINC19813096).

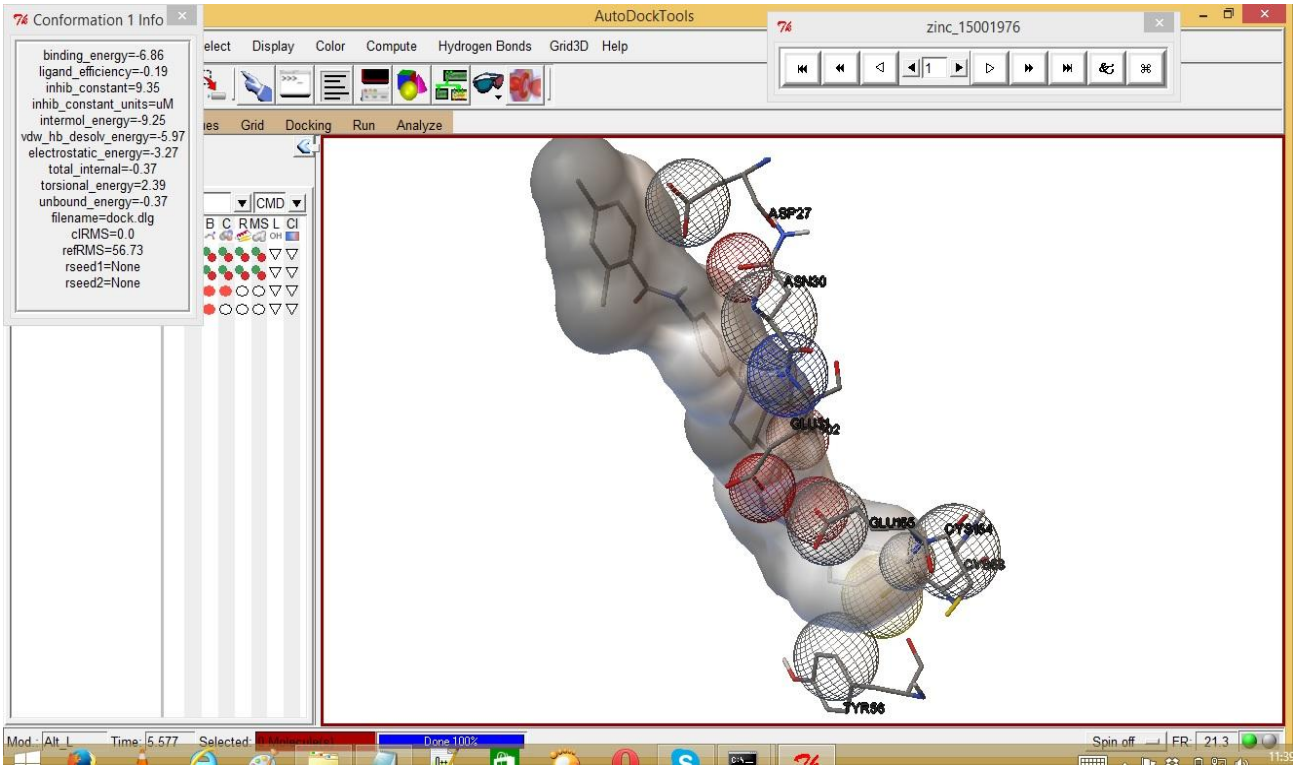

Fig15g. Docking with ligand 7 (ZINC15001976).

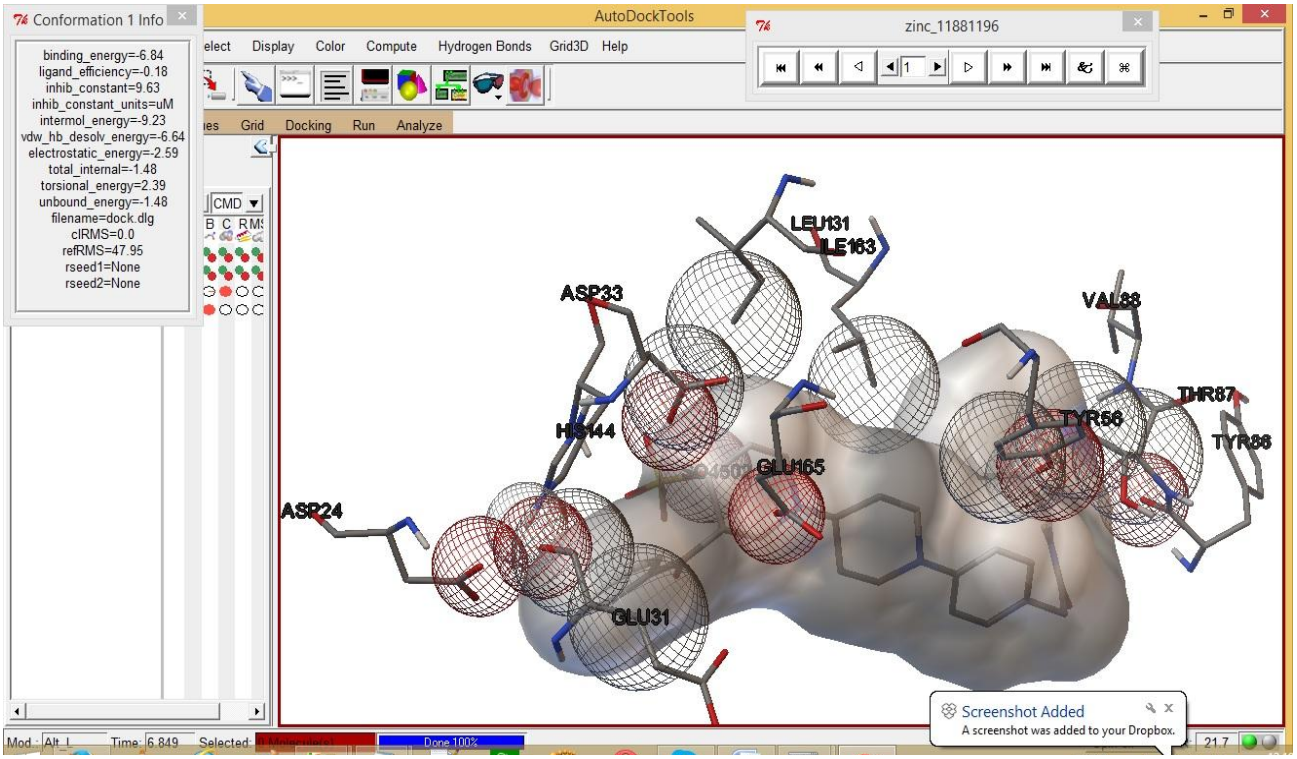

Fig15h. Docking with ligand 8 (ZINC11881196).

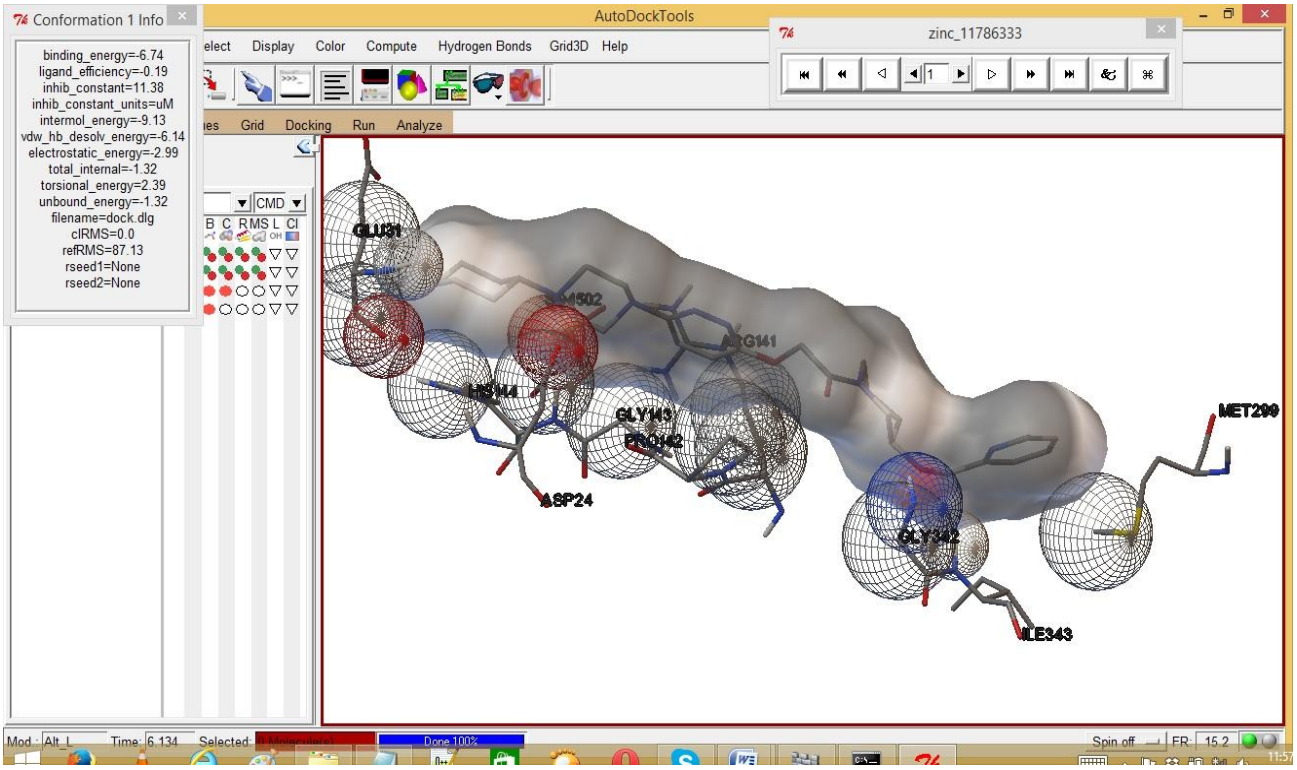

Fig15i. Docking with ligand 9 (ZINC11786333).

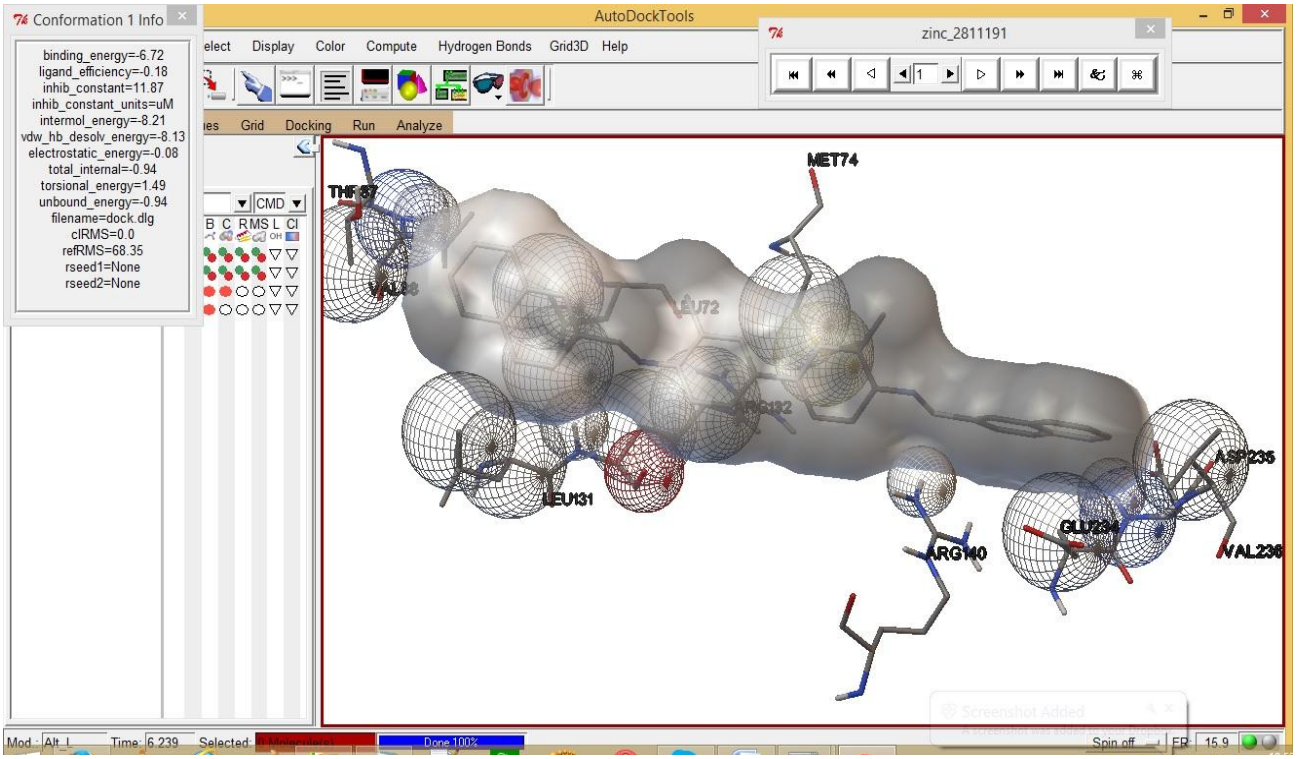

Fig15j. Docking with ligand 10 (ZINC02811191)

4. Checking Lipinski's filters

Lipinski's rule of five had been checked for top 10 ligand molecules selected based on the binding energy. The results are shown below in Table 2.

S.N	Ligands (ZINC ID)	Molecular Mass	LogP	Hydrogen Bond Donors	Hydrogen Bond Acceptors
1	zinc_4113772	490	0.47	4	6
2	zinc_11963881	493	1.99	2	7
3	zinc_12243261	494	1.68	3	4
4	zinc_12191590	537	2.43	2	4
5	zinc_12453654	490	0.59	2	6
6	zinc_19813096	511	-0.42	5	4
7	zinc_15001976	494	4.53	3	4
8	zinc_11881196	497	3.09	2	5
9	zinc_11786333	495	1.49	1	8
10	zinc_2811191	488	9.78	0	2

Table3. Lipinski's filters for top 10 ligands

5. Cheking Toxicity of ligand molecules

a) Using Toxicity checker

Presence of any toxic substructure was checked by the online tool Toxicity checker by providing the SMILES sequence of the ligand. It is done for the top 10 ligands and the result is shown below.

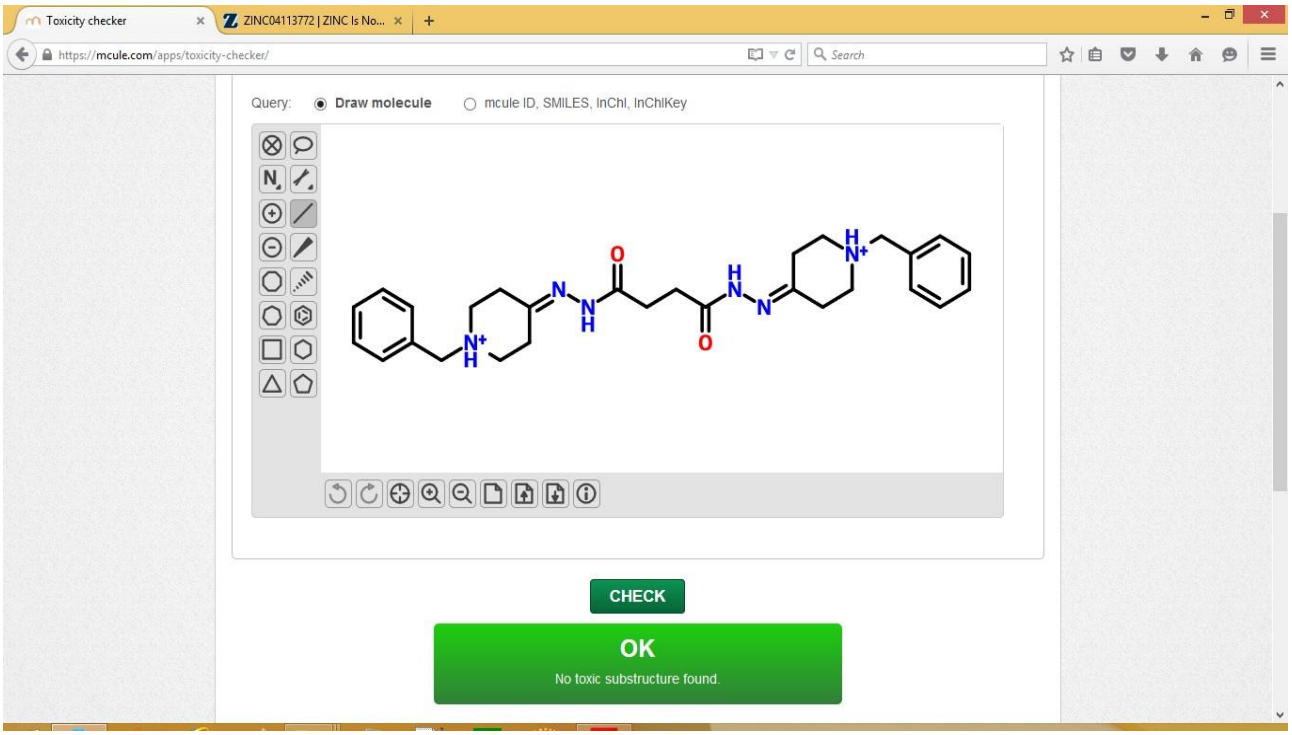

Fig16a. Toxicity checker result of ligand 1.

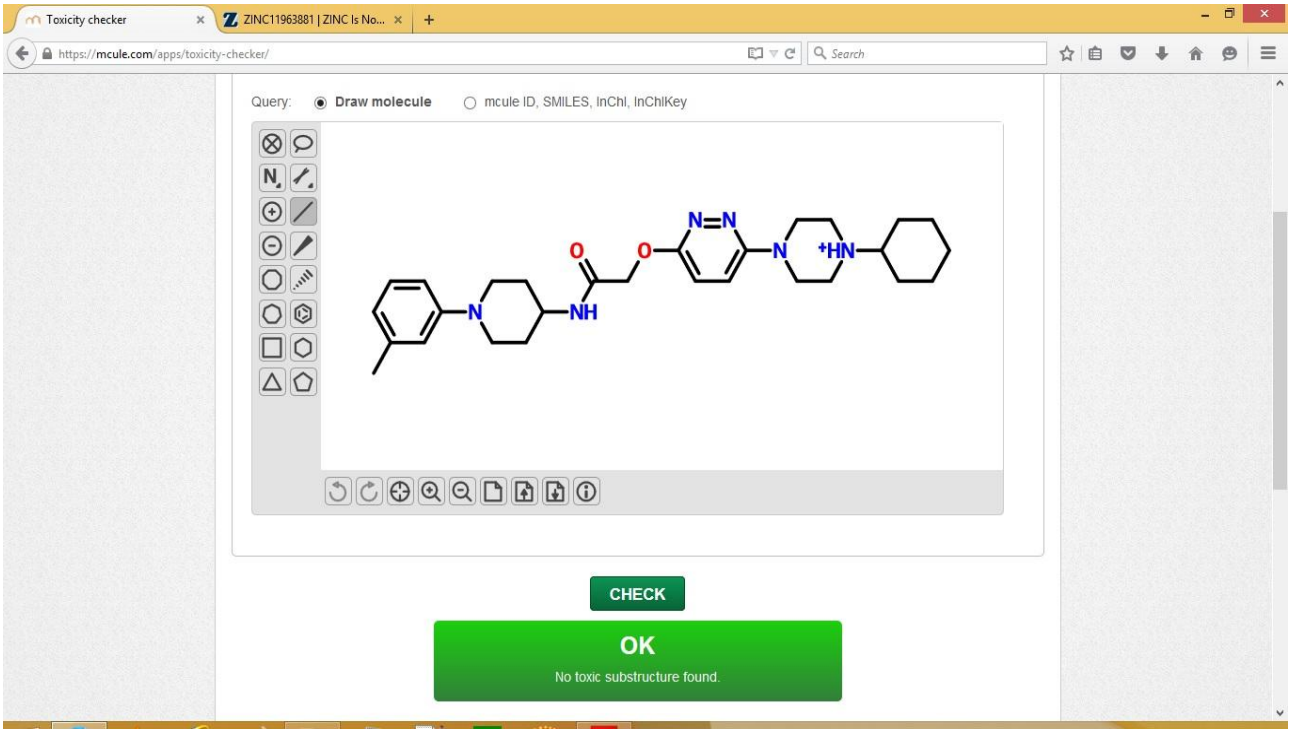

Fig16b. Toxicity checker result of ligand 2.

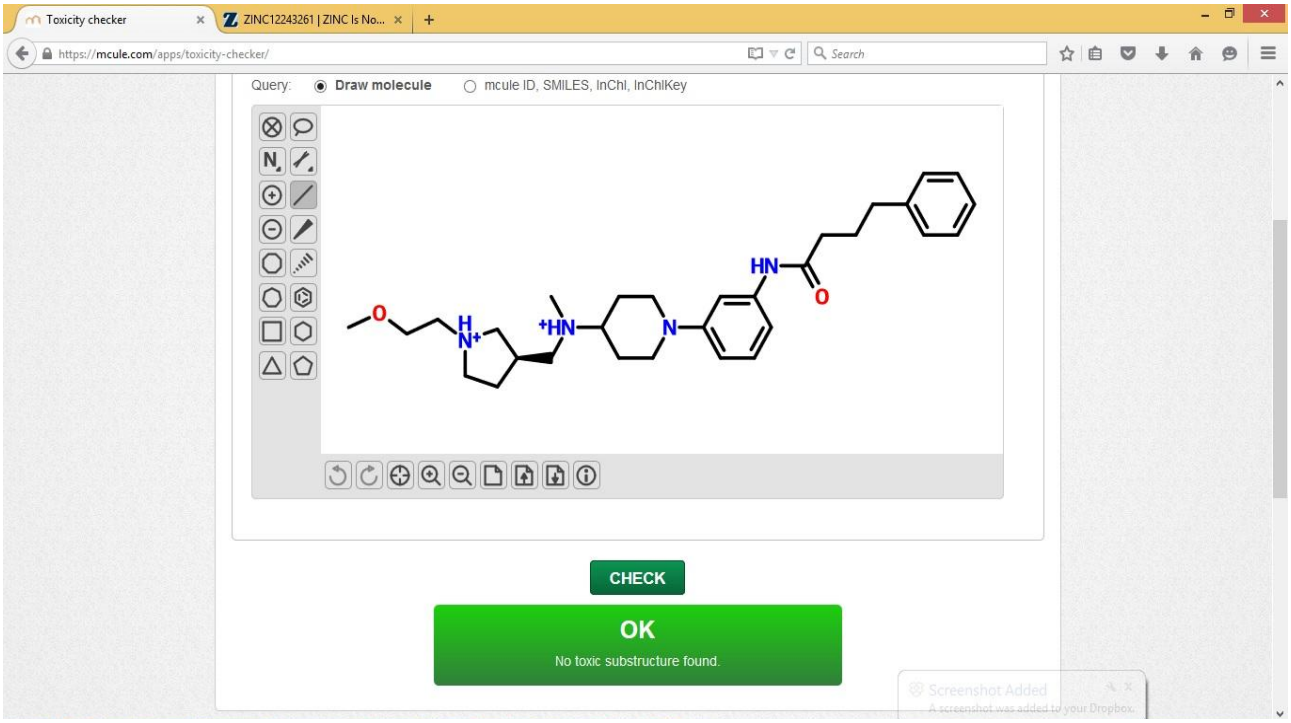

Fig16c. Toxicity checker result of ligand 3.

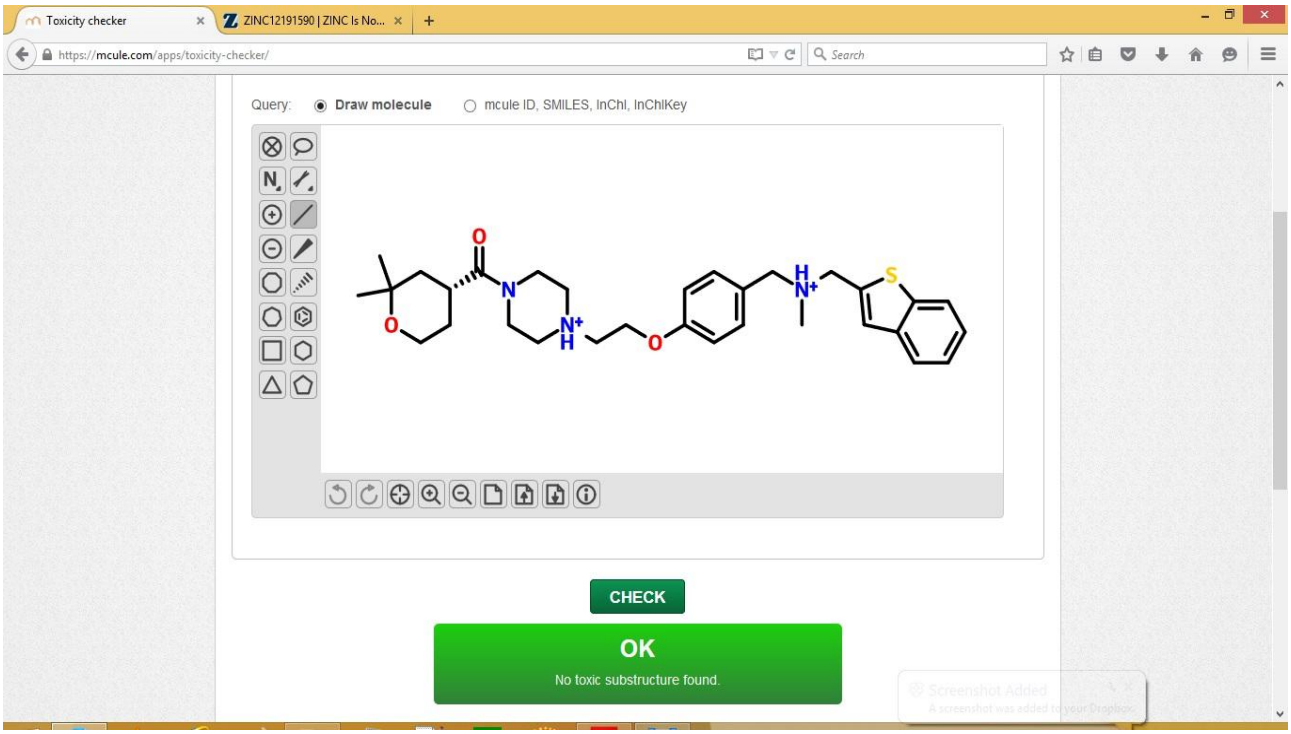

Fig16d. Toxicity checker result of ligand 4.

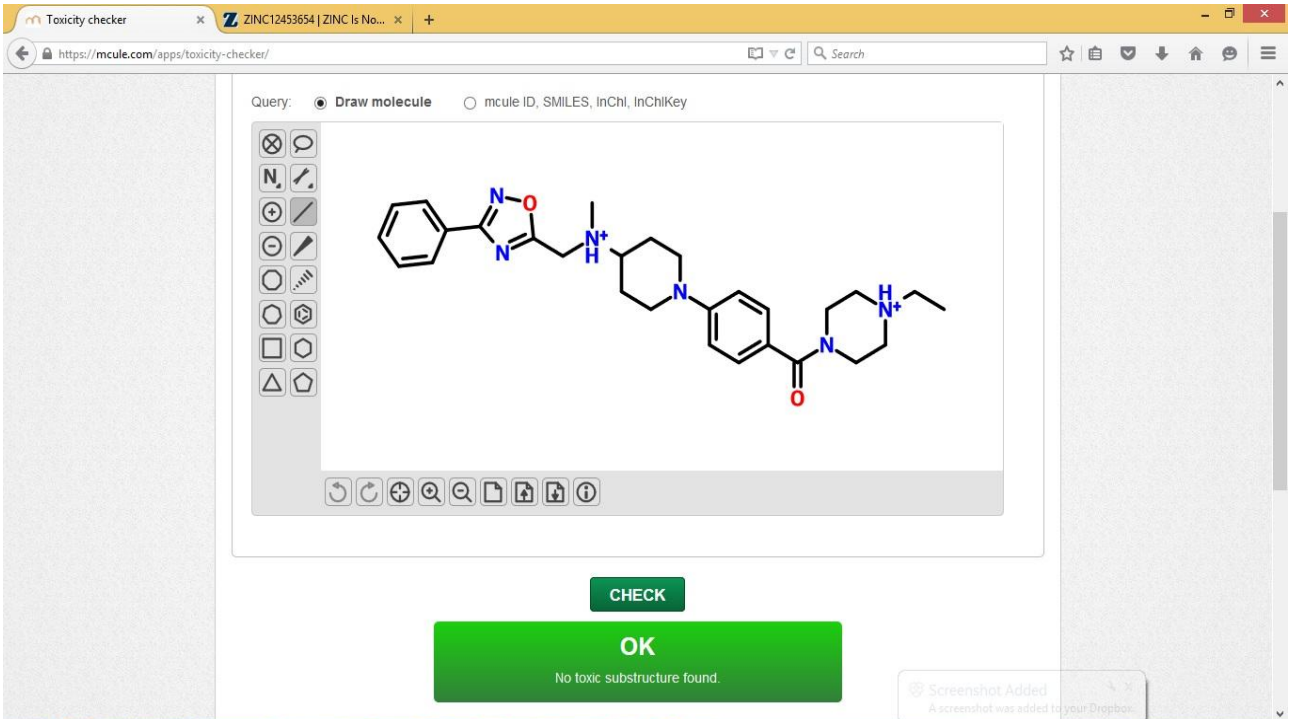

Fig16e. Toxicity checker result of ligand 5.

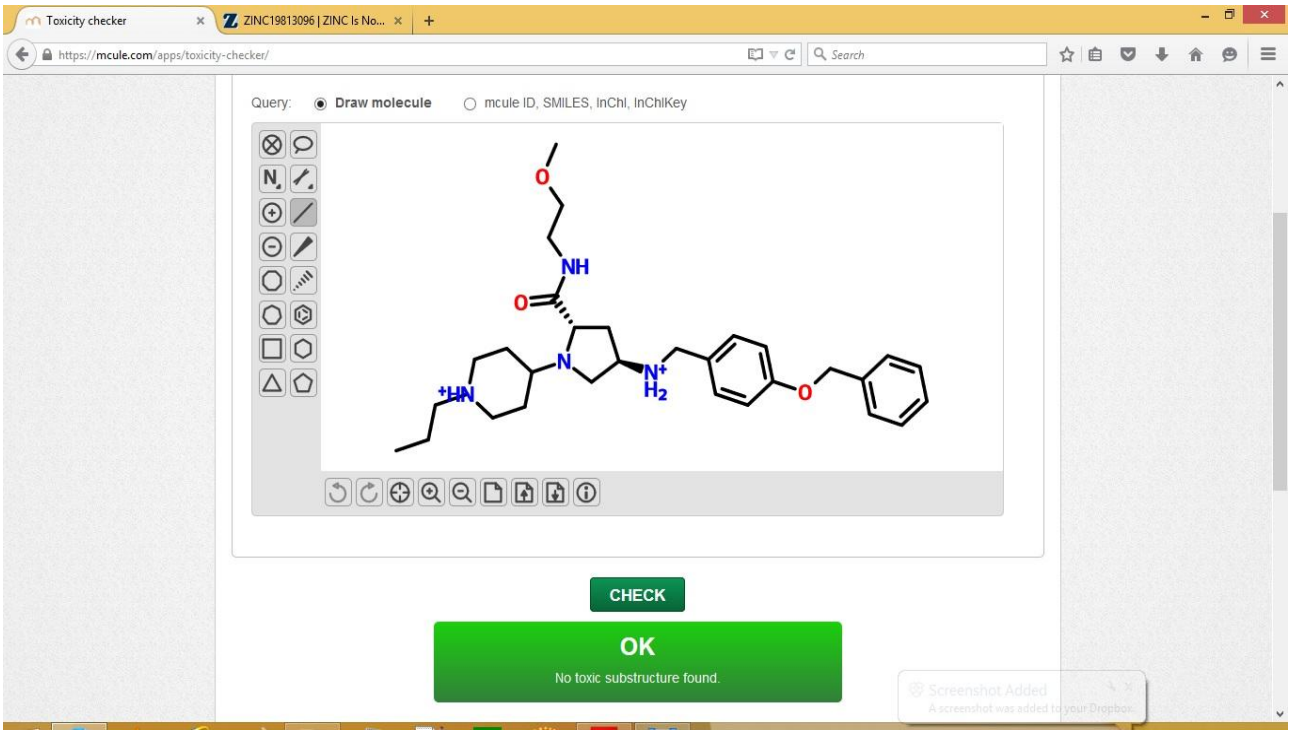

Fig16f. Toxicity checker result of ligand 6.

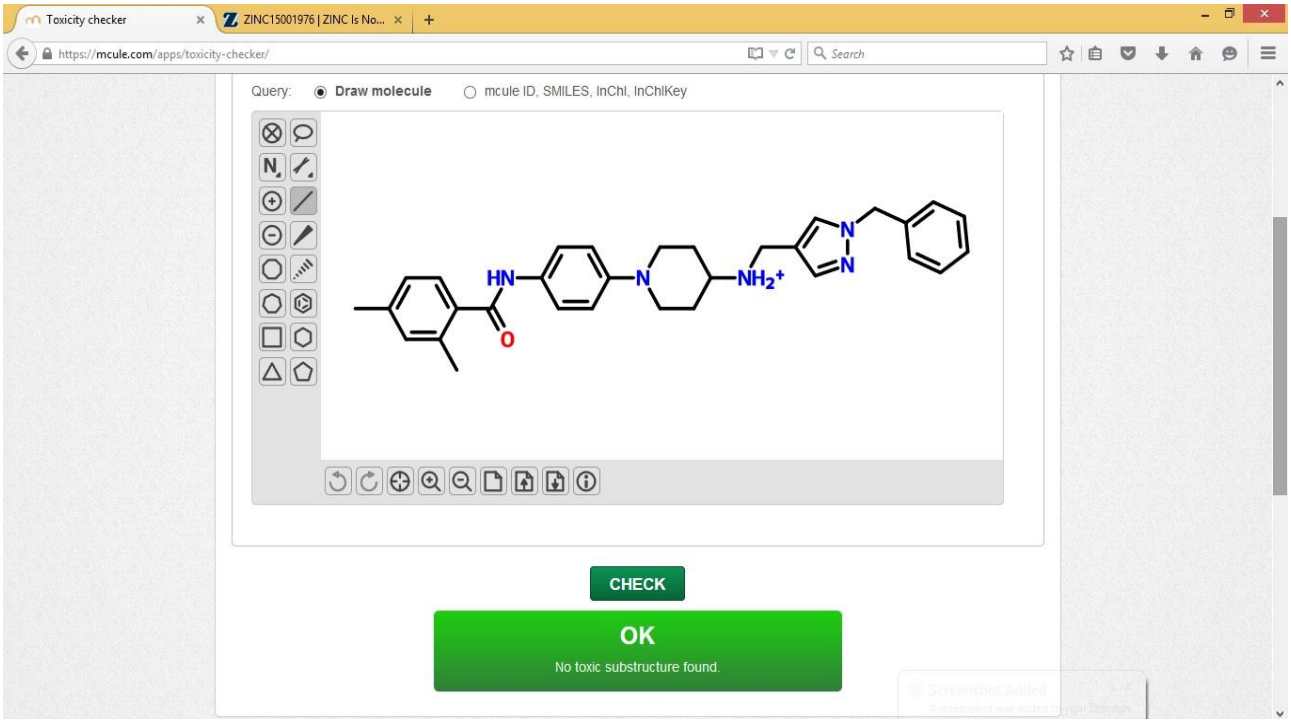

Fig16g. Toxicity checker result of ligand 7.

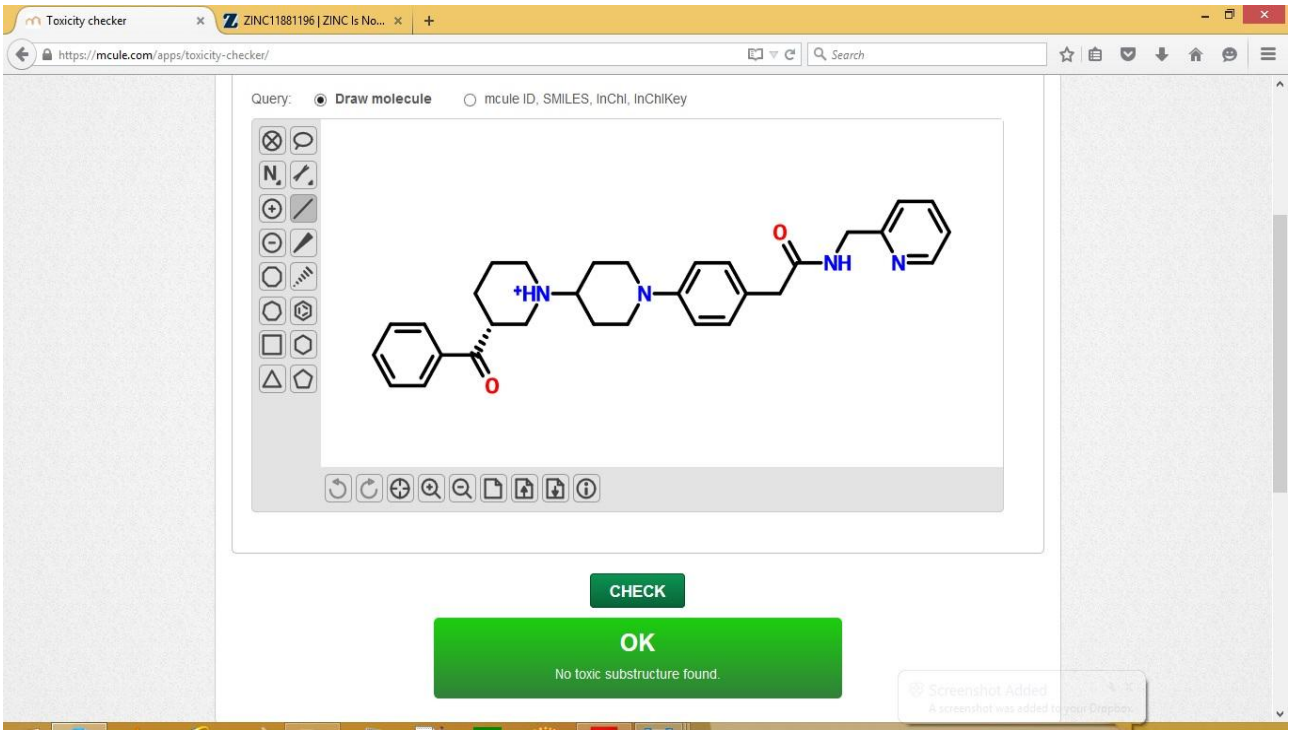

Fig16h. Toxicity checker result of ligand 8.

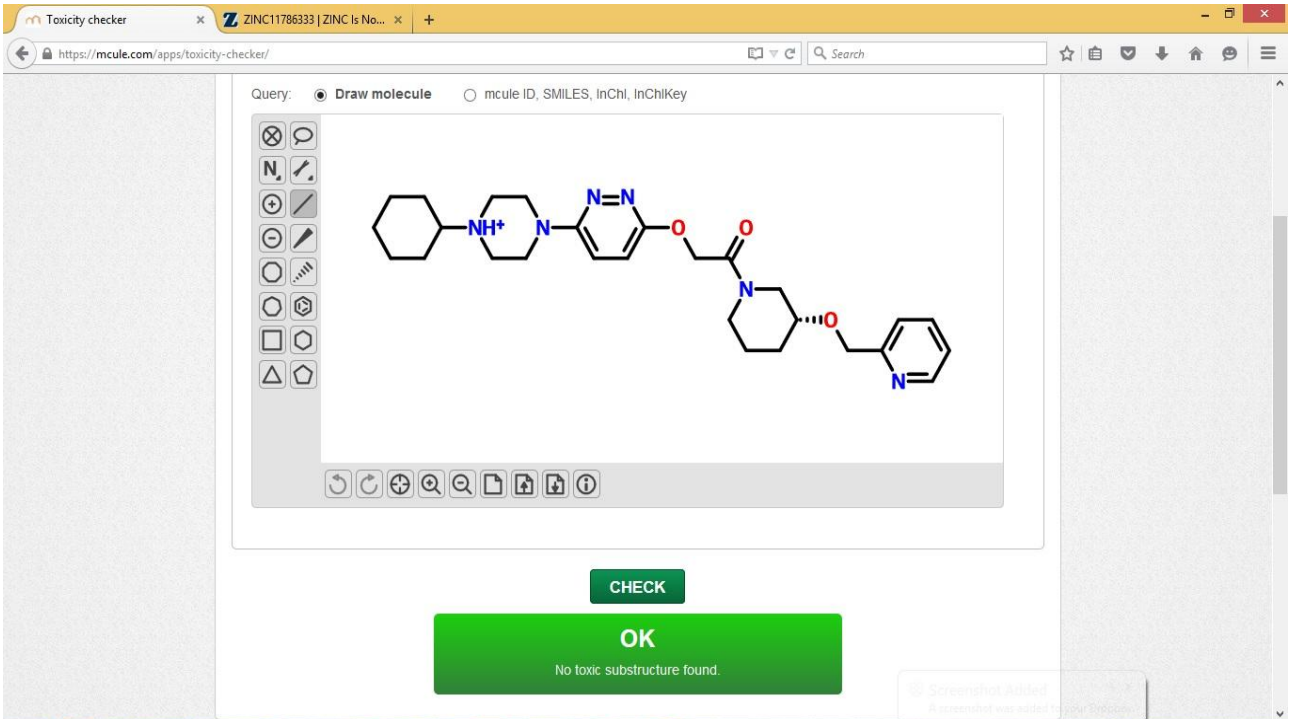

Fig16i. Toxicity checker result of ligand 9.

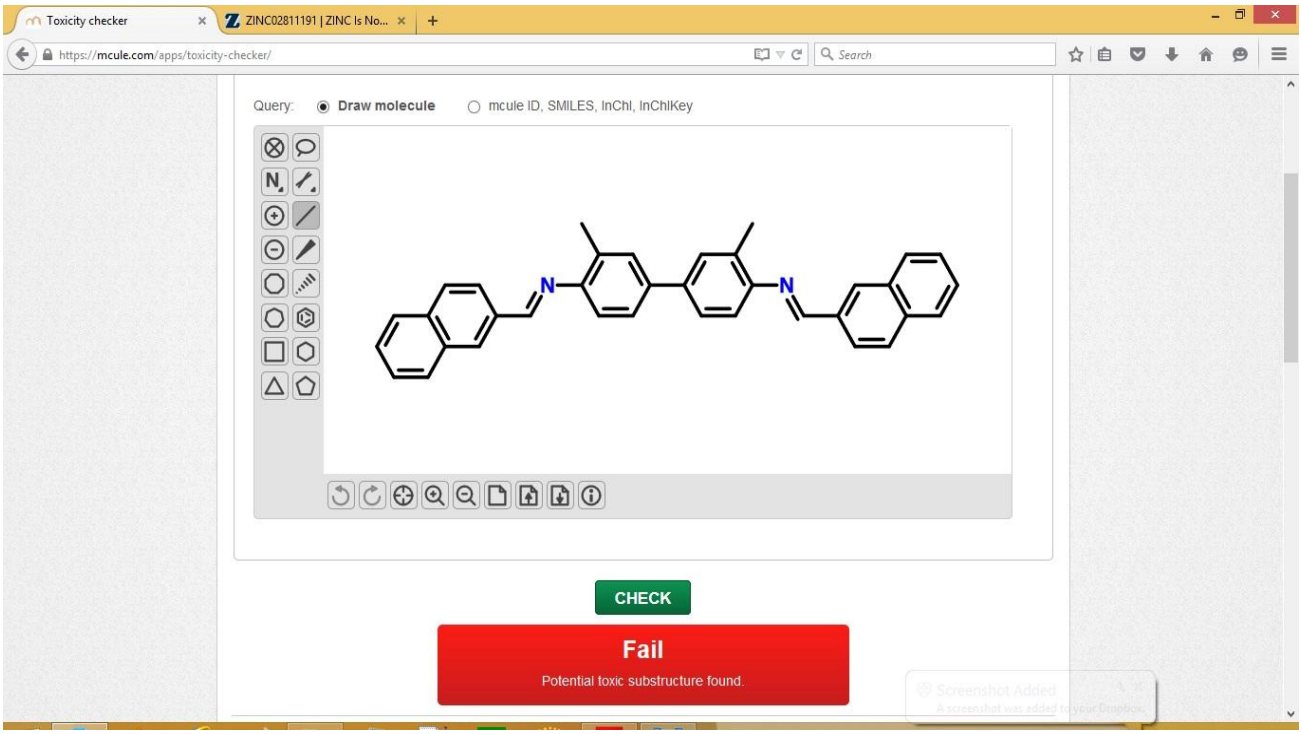

Fig16j. Toxicity checker result of ligand 10.

b) Using OSIRIS data warrior software

Toxicity as well as various physicochemical properties was again checked by OSIRIS data warrior software for top 10 ligand molecules as well as for tetracycline hydrochloride and doxycycline. These results are summarized in Table 4.

Ligand	tetracycline hydrochloride	doxycycline	ZINC04113772	ZINC11963881	ZINC12243261	ZINC12191590	ZINC12453654	ZINC19813096	ZINC15001976	ZINC11881196	ZINC11786333	ZINC02811191
Total Mol. Wt.	444.439	444.439	490.65	494	495	538	491	511	495	498	496	489
cLogP	-4.702	-4.54	-0.54	1.11	0.7	0.7	-0	-1	3	1.6	0.22	9.1
cLogS	-1.265	-1.37	-4.25	-4.76	-3.7	-4	-4	-3	-5	-5	-3.3	-11
H-Acceptors	10	10	8	8	6	6	8	7	6	6	9	2
H-Donors	6	6	4	2	3	2	2	3	2	2	1	0
Polar Surface Area	185.65	185.65	91.8	75	50.5	76	71	72	67	67	85.1	25
Druglikeness	3.0017	2.4239	1.9135	1.13	-0.2	2.9	5.3	4	6.7	4.2	-1.1	-0.4
Mutagenic	None	none	none	none	None	none	None	None	High	None	None	High
Tumorigenic	None	none	none	none	None	none	None	None	High	None	None	High
Reproductive Effect	High	high	Low	low	None	none	None	None	High	None	Low	None
Irritant	None	none	None	none	None	none	None	None	Low	None	None	High
Non-C/H Atoms	10	10	8	8	6	7	8	7	6	6	9	2

Stereo Centers	5	6	0	0	3	2	1	2	0	2	1	0
Rotatable Bonds	2	2	9	7	12	9	7	13	8	8	8	5
Rings	4	4	4	5	4	5	5	4	5	5	5	6
Aromatic Rings	1	1	2	2	2	3	3	2	4	3	2	6

Table4. Physicochemical properties and toxicity results for top 10 ligands and drug molecules.

CONCLUSION

The whole genome analysis of the pathogen and use of database or tools like DEG, BLAST, etc. can be helpful in the prediction of possible protein targets. And use of docking software like AutoDock may lead to novel or optimized drugs against the pathogen. *Chlamydophila psittaci* is a bacterial pathogen which causes psittacosis in humans. Available drugs against this disease are doxycycline and tetracycline hydrochloride which can't eradicate the pathogen completely. Here we present a computer aided prediction of target protein against the protein product of the essential gene of *C.psittaci* which is important for the survival of pathogen. In this study, we docked 100 ligands with this protein which are identified through RASPD tool based on their active site of protein, to identify the best ligand which can be the possible drug molecule. Further, we check different physicochemical properties like Lipinski filters, etc. and toxicity was also checked to find out whether any harmful substructure is present in our ligand molecules. In this study, we come up with top 9 ligands (based on their binding energy against target protein) which are having better binding energy than the available natural ligand as well as drug molecules, and also no toxic substructure is found in our ligand molecules.

N,N'-bis[(1-benzyl-4-piperidylidene)amino]butanediamide (ZINC ID 04113772) is the best ligand among the 9 ligand with the binding energy of -8.86 kcal/mol. So, it can be our candidate drug molecule.

DISCUSSION AND FUTURE PERSPECTIVE

The knowledge of full genome and the application of subtractive approach have been very useful for drug target identification against many bacterial pathogens. In this work we have tried to identify potential drug targets against the bacterial pathogen *Chlamydophila psittaci* through essential genes identification using the database DEG. Then through NCBI BLAST and PDB, we could get a non-human homologous protein structure common to all the three strains of the pathogen. So, our target protein is "bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II" protein. Biosynthesis of Riboflavin is carried out by this protein. The pathways and chemical reactions required in riboflavin biosynthesis leads to riboflavin (vitamin B2) formation. According to some researchers, the enzymes required in pathway of riboflavin biosynthesis may also be helpful in developing the antibacterial drugs for the cure of Gram-negative bacteria and yeasts infections. This is based on the fact that Gram-negative bacteria are unable to consume riboflavin from outside i.e. the external environment (Fischer and Bacher, 2008; Cushman *et al.*, 2001). Since, the Gram-negative bacteria are required to produce riboflavin by themselves, inhibiting the riboflavin synthase or any other enzymes required in riboflavin biosynthesis can be very helpful in the development of antibacterial drugs.

So, our next aim is to identify the potential drug molecule against our protein target molecule. The natural ligand is determined from PDB as Sulphate ion and docking of natural ligand with the target protein gives a cut-off value of binding energy for any ligand to be the drug molecule. Then based on the active sites present on our target protein, 100 ligands were identified as suitable ligands. Thus, virtual library of ligands were generated. And each ligand was separately docked with the target protein to determine top 10 ligands having better binding energy. And the best shows better binding energy than the natural ligand as well as available drugs. Then, various physicochemical properties and toxicity of the selected ligands were determined. Finally, we come up with ligand, N,N'-bis[(1-benzyl-4-piperidylidene)amino]butanediamide (ZINC ID 04113772) as the best ligand molecule and have the potential to be a drug molecule.

So, in future one can go for clinical trials with the best ligand and study their different properties like pharmacodynamic, pharmacokinetics, solubility etc.

REFERENCES

Anishetty S, Pulimia M, Pennathur G. Potential drug targets in Mycobacterium tuberculosis through metabolic pathway analysis. Comput. Biol. Chem. 2005;29:368–378.

Barh D, Kumar A. *In silico* identification of candidate drug and vaccine targets from various pathways in *Neisseria gonorrhoeae*. In Silico Biol. 2009;9:0019

Chhabra G, Sharma P, Anant A, Deshmukh S, Kaushik H, Gopal K, Srivastava N, Sharma N, Garg LC. Identification and modeling of a drug target for *Clostridium perfringens SM101*. Bioinformation. 2010;4:278–289

Chong CE, Lim BS, Nathan S, Mohamed R. *In silico* analysis of *Burkholderia pseudomallei* genome sequence for potential drug targets. In Silico Biol. 2006;6:0031.

Cushman M, Yang D, Kis K, Bacher A (December 2001). "Design, synthesis, and evaluation of 9-D-ribityl-1,3,7-trihydro-2,6,8-purinetrione, a potent inhibitor of riboflavin synthase and lumazine synthase". *J. Org. Chem.* **66** (25): 8320–7

Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, Bandyopadhyay D. *In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. In Silico Biol. 2006;6:43–47.

Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL (2004). "Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures". Journal of Computational Chemistry 25 (2): 265-84. doi:10.1002/jcc.10378. PMID 14648625.

Fischer M, Bacher A (June 2008). "Biosynthesis of vitamin B2: Structure and mechanism of riboflavin synthase". *Arch. Biochem. Biophys.* **474** (2): 252–65

Galperin MY, Koonin EV. Searching for drug targets in microbial genomes. Curr Opin Biotechnol. 1999;10:571–578

Goldman BB, Wipke WT (2000). "QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (QSDock)". Proteins 38(1): 79-94. doi:10.1002/(SICI)1097-0134(20000101)38:1<79::AID-PROT9>3.0CO;2-U. PMID 10651041.

Harding HB. The epidemiology of sporadic urban ornithosis. Am J Clin Pathol 1962;38:230-243.

Huynen M, Diaz-Lazcoz Y, Bork P. Differential genome display. Trends Genet. 1997;13:389–390.

Itaya M (1995) An estimation of minimal genome size required for life. FEBS Lett 362:257–260.

Jawetz E. Chemotherapy of chlamydial infections. Adv Pharmacol Chemother 1979;7:253-2822

Kaye, K., and T. R. Frieden. 1996. Tuberculosis control: the relevance of classic principles in an era of acquired immunodeficiency syndrome and multidrug resistance. Epidemiol. Rev. 18:52-63.

Kitchen DB, Decornez H, Furr JR, Bajorath J (2004). "Docking and scoring in virtual screening for drug discovery: methods and applications". Nature reviews. Drug discovery 3(11): 935-49. Doi:10.1038/nrd1549. PMID 15520816.

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P (2003) Essential *Bacillus subtilis* genes. Proc Natl Acad Sci USA 100:4678–4683

Lengauer T, Rarey M (1996). "Computational methods for biomolecular docking". Curr. Opin. Struct. Biol. 6 (3): 402-6. doi:10.1016/S0959-440X(96)80061-3. PMID 8804827.

MacFarlane JT, Macrae AD. Psittacosis. Med Bull 1983;39:163-167.

Meng EC, Shoichet BK, Kuntz ID (2004). "Automated docking with grid-based energy evaluation". Journal of Computational Chemistry 13 (4): 505-524. doi:10.1002/jcc.540130412.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function". Journal of Computational Chemistry 19 (14): 1639-1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 93:10268–10273

Perumal D, Lim CS, Sakharkar KR, Sakharkar MK (2007) Differential genome analyses of metabolic enzymes in *Pseudomonas aeruginosa* for drug target identification. In Silico Biol 7:0032

Rathi B, Sarangi AN, Trivedi N (2009) Genome subtraction for novel target definition in *Salmonella typhi*. Bioinformation 4:143–150

Sakharkar KR, Sakharkar MK, Chow VT (2004) A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa.* In Silico Biol 4:355–360

Sarangi AN, Aggarwal R, Rahman Q, Trivedi N. Subtractive genomics approach for *in silico* identification and characterization of novel drug targets in *Neisseria meningitidis* serogroup B. J Comput Sci Syst Biol. 2009;2:255–258.

Schaffner W, Drutz DJ, Duncan GW, Loenig MG. The clinical spectrum of endemic psittacosis. Arch Intern Med 1967;119:433-443

Sharma V, Gupta P, Dixit A (2008) *In silico* identification of putative drug targets from different metabolic pathways of *Aeromonas hydrophila*. In Silico Biol 8:0026

Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Chlamydophila psittaci* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA 94:9869-9874.

Stead, W. W. 1997. The origin and erratic global spread of tuberculosis. How the past explains the present and is the key to the future. Clin. Chest Med. 18:65-77.

Verweij PE, Meis JF, Eijk R, Melchers WJ, Galama JM. Severe human psittacosis requiring artificial ventilation: case report and review. Clin Infect Dis 1995;20:440-442

Yang, S.I. et al. (1991). Control of protein phosphatase 2A by simian virus 40 small-t antigen. Mol. Cell Biol. 11(4): 1988-1995.

Yung AP, Grayson ML. Psittacosis -- a review of 135 cases. Med J Aust 1988;148:228-233

Zhang R, Ou H-Y, Zhang C-T (2004) DEG: a database of essential genes. Nucleic Acids Res 32: D271–D272