ICMR-Biomedical Informatics Centre

The Bioinformatics facility at TRC has been widely used by many research scholars and students from TRC as well as other institutes. Fifteen students have carried out their academic project at TRC during this year for degree programs such as M.Sc., MCA, B. Tech., PGD in Bioinformatics and B.Sc., from various academic institutions. Further, the Centre conducted a workshop on "Perspectives of Biomedical Informatics Centre" which was attended by 25 participants including 18 faculties from different Medical Colleges in and around Chennai. The centre has contributed to five research projects of TRC.
[Contact person: Dr.Luke Elizabeth Hanna (E-Mail ID: hanna@trcchennai.in)]

Completed studies

Potential drug targets against drug-resistant tuberculosis

Background

M. tuberculosis has evolved into a drug-resistant species. *M. tuberculosis* resistant to INH and RMP is referred to as MDR-TB, while *M. tuberculosis* resistant to INH, RMP, one of the quinolones and one of injectable drugs is called extensively drug-resistant TB (XDR-TB). XDR-TB does not respond to most of the anti-TB drugs and MDR-TB has the potential to evolve into XDR-TB. In some parts of the world the incidence of MDR-TB is as high as 14% and XDR-TB cases have been confirmed in all regions of the world. The lethal combination of drug-resistant TB and HIV infection is a growing problem that presents serious challenges for effective TB control. Hence there is an urgent need to develop alternative drugs to combat TB.

To develop drugs for any disease, it is first important to identify candidate genes which could serve as potential drug targets. In this study a simple but significant strategy has been employed to identify such genes in *M. tuberculosis. M. tuberculosis* has 3989 known genes and 112 worked out metabolic pathways. The targets of all currently available anti-TB drugs are enzymes involved in 12 of

these pathways. We hypothesize that any other enzyme involved in these pathways could also be potential drug targets.

Aim

- To identify novel drug targets from the metabolic pathways involving drug-resistant genes

Methods

All proteins (372) involved in the 12 known drug resistance linked metabolic pathways were downloaded from the SWISPROT database. Among the 372 proteins, 38 were found to be involved in more than one pathway, bringing down the number of unique proteins to 334. Of these, 18 proteins are targets for currently used anti-TB drugs, reducing the number to 316. The 316 proteins were compared with the human proteome using BLASTP in order to identify and exclude mycobacterial proteins with human homologs (e value of 0.005). This process excluded 221 proteins, leaving out 95. Since these proteins are involved in proven drug targeted metabolic pathways of *M. tuberculosis* and have no significant homologs in humans, they possess the potential to qualify as drug targets. Since drug development is a very expensive process, narrowing down the number of candidates using stringent criteria would be highly beneficial. We have therefore applied additional filters, viz., essentiality (for survival of mycobacteria), virulence and drug regulation (genes regulated by anti-mycobacterial drugs), to identify drug targets with high potential.

Results & Conclusions

Among the 95 proteins identified by us, 63 proteins were already identified as drug targets by various groups while 32 proteins were predicted in this study for the first time as drug targets. Further, of the 95 probable drug targets, 34 were found to be essential for *M. tuberculosis* (database of essential genes), 3 were involved in the pathogenesis/virulence of *M. tuberculosis* (virulence factor database) and 3 were found to be regulated by the action of capreomycin (a second-line anti-TB drug) (microarray data), thereby adding significance to their value as potential drug targets for *M. tuberculosis* (tables 14 & 15).

Table 14: Potentials drug targets from metabolic pathways linked with drug resistance also regulated by anti-TB drugs

Sl. No.	Rv No.	Swissprot ID	Description	Metabolic pathway	E value	PDB ID
1	Rv0651	P66044	50S ribosomal protein L10	Ribosome metabolism mtu03010	5.8	NA
2	Rv1015c	P66121	50S ribosomal protein L25	Ribosome metabolism mtu03010	3.6	1DFU
3	Rv1547	P63977	DNA polymerase III subunit alpha	Purine metabolism mtu00230	0.35	NA

Table 15: Potential drug targets from metabolic pathways linked with drug resistance and also involved in virulence

Sl. No.	Rv No.	Swissprot ID	Description	Metabolic Pathway	E value	PDB ID
1	Rv0642c	Q79FX8	Methoxy mycolic acid synthase 4 (MMAA4)	Tryptophan metabolism mtu00380	0.058	NA
2	Rv2959c	Q50457	Rhamnosyl O-methyl transferase	Tryptophan metabolism mtu00380	0.097	NA
3	Rv3601c	P65660	Aspartate 1-decarboxylase	Alanine and aspartate metabolism mtu00252	5.3	2C45

[Contact person: Dr.Luke Elizabeth Hanna (E-Mail ID: hanna@trcchennai.in)]

Database for drug-resistant mutations in *M. tuberculosis*

Background

The emergence of drug-resistant TB is of great concern since there is no cure for XDR-TB and there is growing concern that it may spread around the world, stressing the need for additional control measures such as new diagnostics and

better drugs for treatment. The primary mechanism of drug resistance in *M. tuberculosis* is the accumulation of mutations in genes coding for drug targets or drug-converting enzymes. Therefore it is necessary to identify all mutations which cause resistance to anti-TB drugs. Molecular mutations causing resistance for most of the currently used anti-TB drugs have been identified. However, this information has not been consolidated and stored systematically for clinical applications as has been done for HIV.

Aim

- To consolidate all known mutations causing resistance against anti-TB drugs

Methods

DNA and protein sequences of *M. tuberculosis* $H_{37}Rv$ from GenBank (NC_000962) were used as reference. All DNA sequences with drug resistance mutations (DRM) deposited in GenBank were obtained using BLASTN. Their corresponding proteins sequences were also retrieved. All hits were compared individually with the reference sequence to identify the location of the mutations.

Results & Conclusions

One hundred and fifty nine sequences reported to contain mutations conferring resistance to anti-TB drugs were retrieved from GenBank. Majority of these sequences were katG mutants. Commonly occurring DRMs in each of these genes were also identified. The results are stored as a local database at the Biomedical Informatics Centre, TRC. This database would serve as a platform to carry out structural bioinformatics studies on drug-resistant TB and pave the way to computer-aided drug designing.

[Contact person: Dr.Luke Elizabeth Hanna (E-Mail ID: hanna@trcchennai.in)]

***In silico* modeling of mutant pyrazinamidase genes of *M. tuberculosis* and docking with pyrazinamide**

Background

Pyrazinamidase (PZAase) plays a key role in activating the prodrug, pyrazinamide (PZA), which is an important drug in the treatment of TB. Mutations in *pncA* gene coding for PZAase are a major mechanism of PZA resistance in *M. tuberculosis.* In this study, mutant PZAases were modeled using Discovery Studio.

Aims

- To predict the three dimensional structure of mutant PZAase
- To study the conformational changes due to the various mutations
- To predict the binding modes of PZA to wild type and mutated models of PZAase from docking calculations

Results & Conclusions

The template chosen was PZAase from *Pyrococcus hirokoshi* (1im5), which had 37% identity with the target protein. It comprises of 7 helices and 6 β strands. The amino acids identified at the active site were Lys96, Asp49, Asp8, Cys138, Trp68, Phe13, Ala134, and Thr135. Models for mutated PZAase at Cys138 to tyrosine and serine were predicted to study conformational changes leading to drug resistance. This study would help in better understanding of the mechanisms of drug resistance. In addition, identification of mutations in the active site of the target protein would give an insight on the drug-target interactions, leading to the rational design of more efficacious drugs that may shorten therapy.

[Contact person: Dr.Luke Elizabeth Hanna (E-Mail ID: hanna@trcchennai.in)]

Tape measure protein having MT3 motif facilitates phage entry into stationary phase cells of *M. tuberculosis*

Background

Tape measure protein (TMP) having MT3 motif in mycobacteriophage TM4 genome has been reported to enable the phage infection of *M. smegmatis* during stationary phase.

Aims

- To identify MT3 motifs in the genome of mycobacteriophages
- To corroborate the presence or absence of MT3 motif in TMP protein of these phages with the ability to infect stationary phase cells of *M. tuberculosis*

Results & Conclusions

TMP of eight additional mycobacteriophages were analyzed using *in silico* methods for the presence of MT3 motifs. Six were found to possess MT3 motifs. To validate the hypothesis that the absence of MT3 motif in Che12 and D29 makes them incapable of infecting stationary phase cells of *M. tuberculosis,* laboratory experiments were conducted to test the performance of respective LRP constructs developed from the parental phages, Che12, D29 and TM4.

[Contact person: Dr. Vanaja Kumar (E-Mail ID: vanajakumar@trcchennai.in)]

Three dimensional modeling of Rv2989 and identification of promoter region for Rv2991

Background

Isocitrate lyase regulatory protein (IclR) family of transcription factors are widely distributed in bacteria, and their products are involved in various functions such as glyoxylate shunt, MDR, quorum-sensing signals, sporulation, control of efflux pumps, etc. The IclR family is named after the well characterized IclR protein of *E. coli*, which controls the glyoxylate bypass. IclR family is defined based on the sequence similarities in the region between residues 151 and 229 of the *E. coli*-IclR primary sequence. *M. tuberculosis* $H_{37}Rv$ has three IclR types of transcriptional regulators (TRs): Rv1719, Rv1773c and Rv2989. These regulators

contain possible helix turn helix (HTH) motifs and share similar amino acid sequences. However, the role of the IclR type of TRs in *M.tuberculosis* has not been addressed so far. The putative IclR type TR transcriptional regulator Rv2989 is similar to SrpS efflux pump regulator from *Pseudomonas putida* (28.35% identity in 247 amino acids; http://genolist.pasteur.fr/TubercuList/). Characterization of this putative TR might provide a clue to its function in *M. tuberculosis*.

Aims

- To predict the three dimensional structure of Rv2989 and DNA binding region
- To predict the promoter region of Rv2991

Results & Conclusions

We identified a HTH motif in the IclR domain of Rv2989 using *in silico* approaches. Further, we predicted the three dimensional structure of Rv2989 protein, and identified its possible DNA-binding sequence. We also identified the conserved residues by performing multiple sequence alignment of Rv2989 with other known proteins having IclR and HTH domain. We have also predicted the promoter region of Rv2991.

[Contact person: Dr.Luke Elizabeth Hanna (E-Mail ID: hanna@trcchennai.in)]

Synonymous codon usage analysis of 32 mycobacteriophage genomes

Background

Modification of codons of the luciferase gene with respect to the optimal codons of the phage and the host should lead to better expression of the same, improving the sensitivity of mycobacterial detection by the LRP assay. The present venture to study and understand codon usage patterns of all the mycobacteriophages so far sequenced is thus aimed at optimizing the LRP assay.

Aim

- To study and compare the synonymous codon usage of 32 completely sequenced mycobacteriophage genomes

Results & conclusions

Synonymous codon usage of protein coding genes of 32 completely sequenced mycobacteriophage genomes was studied using multivariate statistical analysis. One of the major factors influencing codon usage was identified to be compositional bias. Codons ending with either C or G were preferred in highly expressed genes among which C ending codons were highly preferred over G ending codons. Translational selection was also identified to play a role in shaping the codon usage operative at the level of translational accuracy. High level of heterogeneity was seen among and between the genomes. The length of genes was also identified to influence codon usage in 11 of the 32 phage genomes. *Mycobacteriophage cooper* was identified as the most highly biased genome with better translational efficiency.

[Contact person: Dr. Vanaja Kumar (E-Mail ID: vanajakumar@trcchennai.in)]