5.1. Summary and Conclusion:

5.1.1. Summary:

The 21st century is witnessing the dawn of Biotechnology, which is expected to surpass information technology as the new engine of the global economy. Its products will be more important than the fire, the wheel, or the car and will generate more knowledge in a short period of time than history's collective wisdom. Biotechnology is expected to alter healthcare, agriculture, commercial and industrial products. It is predicted, that by the middle of the 21st century, all companies will become biotech companies in some form or the other. One of the key driver and enabler of this emerging technology is bioinformatics. Large scale DNA arrays and other tools of genomics and proteomics offer unique opportunities, which will revolutionize medical research and development in this century. Bioinformatics is a highly interdisciplinary subject that calls for expertise in different disciplines, such as, biology, chemistry, physics, computer science, statistics and mathematics.

Pharmaceutical companies will be able to create drugs based on proteins, enzymes, and RNA molecules associated with genes and diseases. This will facilitate drug discovery and allow drug makers to produce a therapy more targeted to specific diseases. This accuracy not only will maximize therapeutic effects but also decrease damage to nearby healthy cells, instead of the standard trial-and-error method of matching patients with the right drugs, doctors will be able to analyze a patient’s genetic profile and prescribe the best available drug therapy from the beginning. Not only will this take the guesswork out of finding the right drug, it will speed up recovery time and increase safety as the likelihood of adverse reactions is eliminated. Bioinformatics provides the integrated approach of mining the voluminous data generated by the HGP to identify and understand human diseases and to develop corresponding drug and gene based
therapies. Bioinformatics will increasingly replace conventional methodologies in the area of drug discovery and drug delivery.

Tuberculosis (TB), a common, deadly infectious disease, cause of illness and death worldwide, caused by mycobacteria, mainly Mycobacterium tuberculosis (Mt). The proportion of infected children was found to be significantly higher in urban than in rural areas in all zones. The emergence of M. tuberculosis strains that are resistant to the first line drugs represents a serious challenge to tuberculosis control. BCG vaccination (with live but weakened tubercle bacteria) is no longer routinely given to all children of secondary school age. BCG induced tuberculin sensitivity was observed to be pronounced and akin to the sensitivity induced by natural tuberculous infection. There is a well documented association between TB and human immunodeficiency virus (HIV). For HIV-positive individuals, who have compromised immune systems and other risk factors, the lifetime risk exceeds 30 percent. India accounts for nearly 20% global tuberculosis burdens even today.

The emergence of M. tuberculosis strains that are resistant to the drugs represents a serious challenge to tuberculosis control. There is a well documented association between TB and human immunodeficiency virus (HIV). Intense attempts are underway to develop potent analogues of the current antituberculosis, as well as a search for novel drug targets. In Mycobacterium tuberculosis, purine metabolism enzyme sulfate adenylyltransferase subunit2 is a novel target for designing new inhibitors. Sulfate adenylyltransferase is the first enzyme of the two-step sulfate activation sequence. Resistant to currently available antituberculosis drugs have stimulated new efforts regarding the development of new chemotherapy to tuberculosis.
The amino acid sequence (Fasta sequence) of Mtb-SAT subunit2 was obtained from NCBI and the same has been used to search for homologous sequences using blast-P against PDB.

The derived homologous sequences of Mtb-SAT subunit2 protein were aligned using command line interface Clustal-W and graphical user interface Clustal-X, which calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be observed.

Evolutionary relationships can be seen via viewing cladograms or phylograms. The developed model Mtb-SAT subunit2 shows the evolutionary relationship with following templates 1ZUN, 2O8V, 1SUR, 1K6Y, 2PLJ, 1EX4, 2GJG, 1B9F, 1B92, 1BIZ, 1EXQ and 1B9D. Among these Mtb-SAT subunit2 and 1ZUN related with same sub family and also showing high identity.

A large number of techniques have been developed to predict 3-D structures of proteins among which homology modeling is one of the best, Modeller 9v3 was used for the building of 3D structure of Mtb-SAT subunit2, by using of known structure of 1ZUN and the built model was optimized using Swiss-PDBViewer. Validation analysis of stereo chemical quality of the proposed protein of Mtb-SAT subunit2 was submitted to PROCHECK, SWISS-PDB viewer (spdbv), WHAT IF, and ProSA.

Procheck analysis of Mtb- SAT subunit2 showed 94.2% amino acids in most favored, additional and generous allowed regions and 5.8% in disallowed regions. WHATIF analysis showed the average packing
quality the back bone conformations of Mtb- SAT subunit2 model are within range of good quality protein structure. ProSA analysis of Mtb-SAT subunit2 model showed that the Z-score value is reliable and within the range of native conformations.

- The structural superimposition of Ca backed of crystal structure (1ZUN) on to Mtb-SAT subunit2 model, revealed that the Ca RMSD value of 0.45Å, and backbone RMSD value 0.40Å has been characterized as good theoretical model for further analysis of binding sites for inhibitors.

- Analysis of Mtb-SAT subunit2 secondary structure showed that among the 332 amino acids contained strands 23 amino acids (6.9%) are involved in β strand, 74 amino acids (22.3%) in α helix, 4 amino acids (1.2%) in 3-10 helix and 231aminoacids (69.6%) in other structural conformation.

- The active site amino acid residues in Mtb-SAT subunit2 model were accomplished based on its alignment to the template 1ZUN. Alignment of Mtb-SAT subunit2 on to 1ZUN showed that the catalytic residues Leucine66, Serine68 & 73, Glycine70 &164, Lysine71, Histidine95, Valine96, Glutamic acid169 & 170 and Aspartic acid72. Interacting amino acids are conserved.

- The built Mtb-SAT subunit2 model has been accepted with less than 3% stereochemical check failures by Protein Molecular Data Base (PMDB) and available to public with PMDB Id: PM 0075315.

- Second line drugs are mostly used in the treatment of MDR-TB. Potent second line drugs for treatment of tuberculosis are Ethambutol, Isoniazid, Ethionamide, Cycloserine Pyrazinamide and P-aminosalicylic acid (PAS), these drugs were screened through AUTODOCK 4.0 tool which has
provided information that P-aminosalicylic acid (PAS), is possessing best lowest docking energy and thus it has been selected for further analysis.

► 100 lead molecules were designed by introducing with structural modifications of parent molecule of PAS using Molinspiration server. The PAS molecule was further modified structurally and optimized through HYPERCHEM 7.5, CHEM-OFFICE ultra and PRODRG server. All the designed lead molecules were tested for Lipinski’s Rule-of-Five using MOLINSPIRATION server.

► The end results of Lipinski’s rule have enabled for the selection of five best ranked lead molecules (PAS-1 to PAS-5) and were further docked on to Mtb-SAT subunit2 with suitable grid box.

► Docking of the five lead molecules for best interactions with Mtb-SAT subunit2 have given an insight that all the lead molecules have shown interactions with active site amino acids of Mtb-SAT subunit2.

► The analysis of the reports has shown that, the conducted study can be further utilized in designing of better anti tuberculosis drugs.

5.1.2. Conclusions:

The Function of a protein is directly linked to it’s 3-D structure. If the tertiary structure is changed, the protein normally loses it’s ability to perform biological function, since this function depends on the geometrical shape of the active site in the interior of the molecule (lock-key principle). The structures confirm the evolutionary changes in the primary structure of a given protein from related species, through random mutations. As these mutations lead to genetic disorder and diseases at the molecular level, clear understanding of the nature of these diseases is necessary. Also, when the structure of an enzyme is determined, a suitable inhibitor of the active site can be designed through
combinatorial chemistry, computer modeling and docking techniques. This "structure based drug design" promises efficient drugs for several diseases in a short time. Evidences have established the fact that proteins undergo confirmational changes during their participation in the biochemical events. Highly resolved structures can lead to very clear understanding of the functions of these molecules. The structure-function relationship is key to our knowledge of the bio-world. Given the large number of genes being discovered, the rate of new protein sequences is growing exponentially relative to the rate of protein structures being solved by experimental methods. About 7,500 protein structures have been determined experimentally by X-ray crystallography and NMR spectroscopy, while there are over 325,000 entries in GenPept sequence database alone. Therefore alternative strategies like automated computational methods have to be employed in order to obtain 3-D structural information of the proteins.

Comparative or homology modeling or knowledge-based prediction exploits the fact that evolutionarily related proteins with similar sequences have similar structures. The degree of similarity is very high in the so called "core regions" comprising of secondary structural elements ("helices and connecting secondary structures"). While the high precision structures required for detailed studies of protein-ligand interaction can only be obtained experimentally, theoretically predicted models provide molecular biologists with "low resolution" models which hold enough information about preferred spatial arrangements of important residues to guide the design of experiments. Thus even though the current methods are still in their infancy, prediction of structures for all protein sequences of complete genomes in conjunction with experimental work is a realistic goal. Structural analyses of proteins for further mutagenesis, substrate and inhibitor design, and enhanced function and stability are also possible.

Purine biosynthesis has greater scope as target sites for designing of new anti tuberculosis drugs, which are effective against the multidrug resistant strains.
A vivid screening of the targets in the pathways led to the identification of Mtb-SAT subunit2 as a potential target site. The systematic analysis of modeling and docking of Mtb--SAT subunit2 as seen in the various chapters of the present dissertation do enable to draw the following conclusions. The native modeling methods have several constrains, to overcome these computational studies have paved way to predict the 3-D model of Mtb--SAT subunit2 enzyme and for further structural designing of drugs. Based on the molecular interaction studies better drugs could be designed at a prime line level, lead molecule modification and designing, leads to identification of protein ligand interactions which do aid to rationalize effective anti TB drugs. For a virtual screening of protein and ligand molecules, Chemiinformatics and Bioinformatics tools have provided for enhanced experimental research. This paves way for researchers to further analyze the data obtained through computational studies. The systematic step wise analysis conducted in the present dissertation on homology modeling of pathogenic microbial proteins aided by CADD tools yields useful information to design better drug formulations by pharmacists abiding to pharmaceutical norms.