Subtractive Genomics Approach to Identify Putative Drug Targets and Identification of Drug-like Molecules for Beta Subunit of DNA Polymerase III in *Streptococcus Species*

John J. Georrge · V. V. Umrania

Abstract The prolonged use of the antibiotics over the years has transformed many organisms resistant to multiple drugs. This has made the field of drug discovery of vital importance in curing various infections and diseases. The drugs act by binding to a specific target protein of prime importance for the cell's survival. *Streptococcus agalactiae*, *Streptococcus pneumoniae*, and *Streptococcus pyogenes* are the few gram positive organisms that have developed resistance to drugs. It causes pneumonia, meningitis, pharyngitis, otitis media, sinusitis, bacteremia, pericarditis, and arthritis infections. The present study was carried out to identify potential drug targets and inhibitors for beta subunit of DNA polymerase III in these three *Streptococcus species* that might facilitate the discovery of novel drugs in near future. Various steps were adopted to find out novel drug targets. And finally 3D structure of DNA polymerase III subunit beta was modeled. The ligand library was generated from various databases to find the most suitable ligands. All the ligands were docked using Molegro Virtual Docker and the lead molecules were investigated for ADME and toxicity.

Keywords *Streptococcus species* · BLAST · DEG · KAAS Server · Homologous proteins · Potential drug targets · Reverse vaccinology · Modeling · Docking · ADMET

Introduction

The complete genome sequences of many microbes were completed in the past decade [1]. Valuable information on finding the treatment of various infections caused by pathogens [2] can be retrieved using the comparative genomics and subtractive genomics approaches. The critical genes which are crucial for the survival of the pathogens and which are absent in the host [3] can be screened out by using the subtractive genomics approach. The chances of

J. J. Georrge (✉)
Department of Bioinformatics, Christ College, Rajkot 360005 Gujarat, India
e-mail: johnjgeorrge@yahoo.co.in

V. V. Umrania
Department of Microbiology, MVM Sc & HSc College, Rajkot, India

cross-reactivity and side effects [4] can be decreased by selecting such non-homologous proteins which are not found in humans. The genes and their products which can be used as a potential drug targets can be identified by analyzing these genes with the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [5].

The search for novel drug targets is relying on the genomics data. Comparative genomics approach can be used for selecting non-homologous genes coding for proteins which are present in pathogen but not in the host. For identifying such genes, Basic Local Alignment Search Tool (BLAST) against the human using BLASTP program can be performed. This will eliminate homologous genes present in the human. Thereafter, the critical genes required for the survival of the pathogen are identified using Database of Essential Genes (DEG) [6]. This approach will ensure that drug target is present only in the pathogen and not in the humans. Using this approach, novel targets have been identified successfully for various pathogens [7–17].

The involvements of combinatorial Chemistry [18], high throughput screening, and virtual screening, in silico absorption, distribution, metabolism, and excretion–toxicity screening, and de novo and structure-based drug design serve to expedite as well as economize the modern day drug discovery process. Structure-based computational drug design methods mainly focus on molecules for target site with known 3D structure followed by determination of their affinity for target, based on which hits are obtained. Knowledge of these targets and corresponding drugs, particularly those in clinical uses and trials, is highly useful for facilitating drug discovery [19].

Streptococcus is a genus of spherical Gram-positive bacteria belonging to the phylum Firmicutes [20]. Acute *Streptococcus pyogenes* infections may take the form of pharyngitis, scarlet fever (rash), impetigo, cellulitis, or erysipelas. Invasive infections can result in necrotizing fasciitis, myositis, and streptococcal toxic shock syndrome. Patients may also develop immune-mediated sequelae such as acute rheumatic fever and acute glomerulonephritis. *Streptococcus pneumoniae* causes pneumonia, meningitis, and sometimes occult bacteremia. *Streptococcus agalactiae* may cause meningitis, neonatal sepsis, and pneumonia in neonates; adults may experience vaginitis, puerperal fever, urinary tract infection, skin infection, and endocarditis. Viridans streptococci can cause endocarditis. Anaerobic streptococci participate in mixed infections of the abdomen, pelvis, brain, and lungs [21].

Several antibiotics have many side effects and developed resistant against *Streptococcus species*. Most of the selected pathogens have resistant mechanism by efflux pump [22, 23]. The genes cylA and cylB are responsible for the transport of drug molecules in *S. agalactiae* [24]. Penicillin-binding protein 1a, lb, 2a, and 2b, ABC transporter ATP-binding protein/permease, chloramphenicol-*O*-acetyltransferase, and MATE efflux family protein are other few proteins responsible for drug resistant in selected pathogens [25, 26]. Hence it is the need of the hour to explore into the possibility of novel drug target identification and drug designing for these pathogens. This can be achieved now due to the availability of the proteomes of these organisms. This study has analyzed the proteome of the *S. agalactiae*, *S. pneumoniae*, and *S. pyogenes* and identified the most suitable and efficient drug-like compounds.

Materials and Methods

Pathogens and Identification of Essential Genes in *S. Pyogenes* SSI-1

In the present investigation, 13 pathogenic proteomes were used: strain 2603V/R (Tax.ID 208435) and NEM316 (Tax.ID 211110) of *S. agalactiae*, strain 70585 (Tax.ID 488221), G54

(Tax.ID 512566), Hungary19A-6 (Tax.ID 487214), JJA (Tax.ID 488222), P1031 (Tax.ID 488223), TIGR4 (Tax.ID 170187), and Taiwan19F-14 (Tax.ID 487213) of *S. pneumoniae*, and strain MGAS5005 (Tax.ID 293653), MGAS6180 (Tax.ID 319701), MGAS8232 (Tax. ID 186103), and SSI-1 (Tax.ID 193567) of *S. pyogenes*. *S. pyogenes* SSI-1 (Tax.ID 193567) were used as a reference proteome, which is sequenced by Osaka University, Japan. The advantage with this proteome is the smallest among selected organisms. The 1859 reference proteins of *S. pyogenes* SSI-1 were downloaded from NCBI (www.ncbi.nlm.nih.gov) and this is subjected to BLASTP against above said strains, with E value of 10^{-4}.

The obtained shared proteins were used for further analysis. As stated by Dutta et al. [10], the proteins having sequence length less than 100 amino acids (they were less likely to represent essential genes) were not eliminated because our subjective investigation shows that many approved drugs targeting the proteins which have less than 100 amino acids [27].

The proteins which shared in all selected strains were analyzed using CD-HIT to identify the paralogous or duplicate proteins by clustering techniques [28]. Sequence identity cut-off was kept at 0.6 (60% identity) as sequence having identity >60% having similar/related structure and functions [29, 30]; global sequence identity algorithm was selected for the alignment of the amino acids; bandwidth of 20 amino acids and default parameters for alignment coverage were selected. These proteins were subjected to BLAST against Human [31] with expectation value (E value) of 10^{-4} and Refseq protein database was selected. To search for the homologous proteins between *Streptococcus species* and host, BLASTP program was used. The obtained homologous protein set was eliminated. The resultant dataset was found to be not homologous with human. DEG (http://tubic.tju.edu.cn/deg/) was performed to identify the essential genes necessary for the survival of the selected *Streptococcus species*. A random expectation value (E value) was kept as 10^{-4}; minimum bit-score cut-off of 100; BLOSUM62 matrix and gapped alignment mode were selected to screen out the essential proteins [6].

Metabolic Pathway Analysis

Metabolic pathway analysis of the essential proteins of *Streptococcus species* was done by KAAS server (www.genome.jp/tools/kaas/). KEGG Automatic Annotation Server (KAAS) provides functional annotations of genes by BLAST comparisons against the manually curated KEGG Genes database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways [32]. KEGG pathway studies were also conducted to analyze the occurrence of alternate pathways after which the proteins were selected as potential drug targets.

Identification of Novel Proteins and its Subcellular Localization Prediction

Screening of the potential drug targets was carried out by similarity search using protein sequence of all the potential targets against the DrugBank [27], TTD [33], PDTD [34], and HIT [35] to reach novel drug targets. Additionally, using computational methods, the subcellular localization of the protein by Psortb v3.0 [36] and outer membrane proteins by TransMembrane prediction using Hidden Markov Models (TMHMM) [37] were predicted to identify the surface membrane proteins which could be used as probable vaccine candidates. Psortb generates prediction results for four major localizations for Gram-positive bacteria (cytoplasmic, cytoplasmic membrane, cell wall, and extracellular); TMHMM is a program for predicting transmembrane helices based on a hidden Markov model and it reads a FASTA formatted protein sequence and predicts locations of trans-membrane, intracellular, and extracellular regions.

3D Structure Modeling and Validation of DNA Polymerase III Subunit Beta

After detailed analysis as a case study, 3D structure of the DNA polymerase III subunit beta (GI 28894914 and Accession NP_801264.1) was modeled by ab initio protein modeling tool I-TASSER as no template is available in PDB (Fig. 4).

The iterative threading assembly refinement (I-TASSER) server is an integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm. Starting from an amino acid sequence, I-TASSER first generates 3D atomic models from multiple threading alignments and iterative structural assembly simulations. The function of the protein is then inferred by structurally matching the 3D models with other known proteins. The output from a typical server run contains full-length secondary and tertiary structure predictions, and functional annotations on ligand-binding sites, enzyme commission numbers, and gene ontology terms. An estimate of accuracy of the predictions is provided based on the confidence score of the modeling [38].

The stereochemical quality of the models was verified with the program PROCHECK. It assess the stereochemical quality of a given protein structure. The aim of PROCHECK is to assess how normal, or conversely how unusual, the geometry of the residues in a given protein structure is as compared with stereochemical parameters derived from well-refined, high-resolution structures [39].

The energy was calculated at atomic level using ANOLEA server. The atomic empirical mean force potential (ANOLEA) is used to assess packing quality of the models. The program performs energy calculations on a protein chain, evaluating the "Non-Local Environment" of each heavy atom in the molecule. The y-axis of the plot represents the energy for each amino acid of the protein chain. Negative energy values (in green) represent favorable energy environment whereas positive values (in red) unfavorable energy environment for a given amino acid [40].

Active Site Identification

Active site determination was done for the modeled protein to further work on its docking studies. Active site determination gives us an idea to make a grid before docking. This was achieved by pocket finder [www.modelling.leeds.ac.uk/pocketfinder] which implements LIGSITE [41] which is based on the POCKET algorithm [42] and Q-site finder [www.modelling.leeds.ac.uk/qsitefinder] which works by binding hydrophobic (CH3) probes to the protein, and finding clusters of probes with the most favorable binding energy. These clusters are placed in rank order of the likelihood of being a binding site according to the sum total binding energies for each cluster [43].

Ligand Library Construction

The synthetic compounds which having 195,141 molecules were retrieved from Pubchem [http://pubchem.ncbi.nlm.nih.gov/], drug-like compounds of NCI database [http://ligand.info], inhibitors from BRENDA enzyme database [www.brenda-enzymes.org], and literature search. The herbal compounds which having 390 molecules were obtained from PubChem Bioassay, DR. DUKE'S PHYTOCHEMICAL library [www.ars-grin.gov/duke], BRENDA enzyme database, and literature search. The 1,436 molecules of FDA approved drugs were obtained from DrugBank [www.drugbank.ca]. The approved drugs were selected due to their FDA approval and thus which will not require pre-clinical trials. As they are already targeting bacterial proteins they are more likely to complete inhibition of an organism.
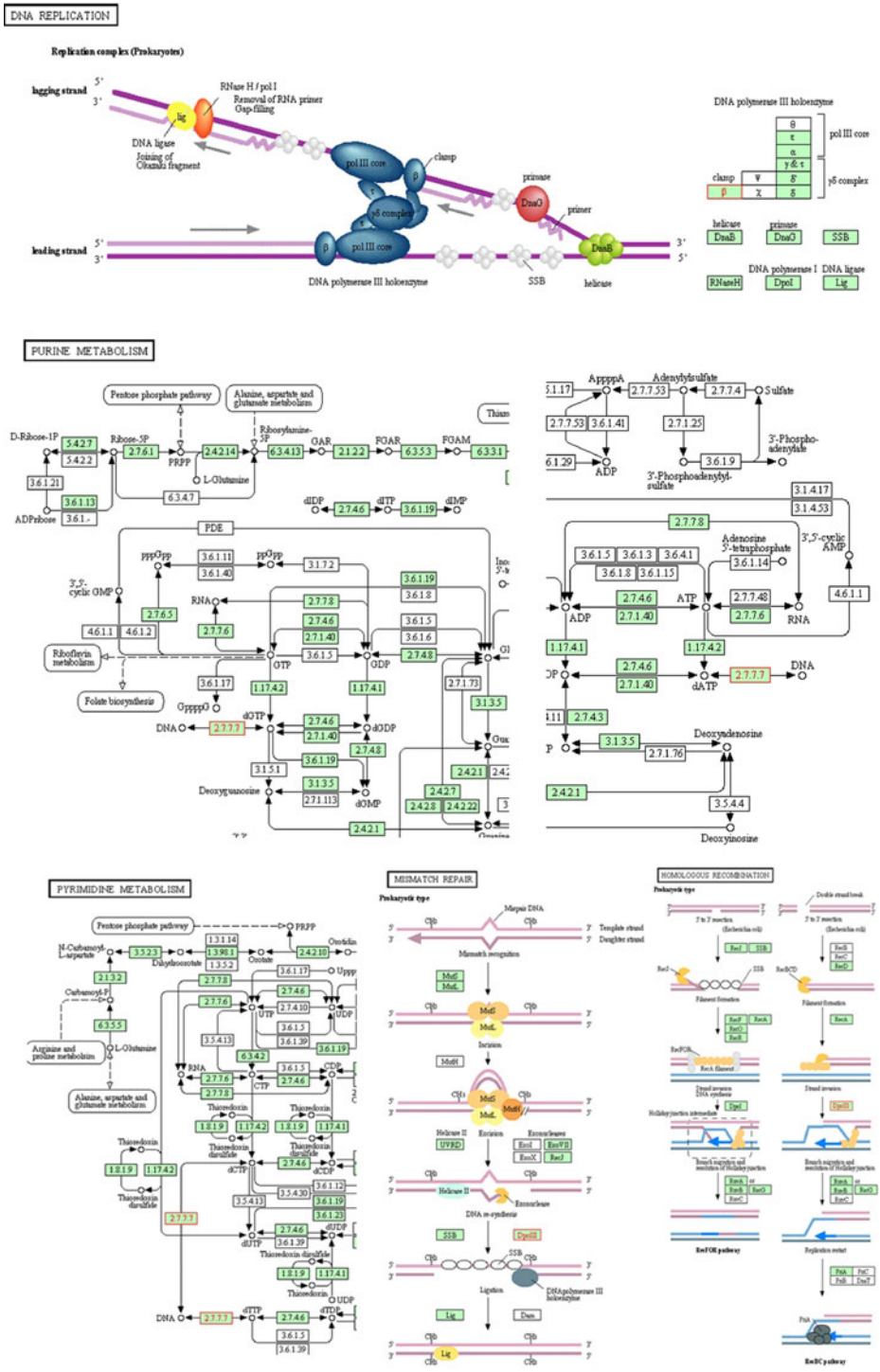
Fig. 1 Role of DNA polymerase III beta subunit in different metabolic pathways

Molecular Docking

Molecular docking studies were performed using Maestro version 9.0 [44] and High Throughput Virtual Screening Glide. This was done in order to screen the potential inhibitors from the ligand library. All ligands were docked flexibly to their respective targets. To prepare the system for docking, the proteins were then prepared for subsequent grid generation and docking using the Protein Preparation Wizard tool supplied with Glide. Using this tool, all hydrogen atoms were added and the entire protein was minimized. Next, a grid was prepared for docking into their respective targets using the Receptor Grid Generation tool in Glide. The molecules obtained from HTVS Glide were given as input for LigPrep-application with the OPLS_2005 force field. Next, a grid was prepared for re-docking and docking was performed using Glide XP mode.

For further verification of docking studies was carried out using Molegro Virtual Docker (MVD). The Molegro Virtual Docker has been shown to yield higher docking accuracy than other state-of-the-art docking products (MVD 87%, Glide 82%, Surflex 75%, FlexX 58%). It has two docking search algorithms, MolDock Optimizer and MolDock SE (Simplex Evolution). MolDock Optimizer is the default search algorithm in MVD. In order to dock the receptor and ligand, the receptor was prepared from the prepare molecule option provided. Than to search for grid, cavities were generated by using detect cavity option. And finally the ligands were provided in sdf file format for docking using docking wizard. During docking, the following parameters were fixed: number of runs 10, population size 50, crossover rate 0.9, scaling factor 0.5, maximum iteration 2,000, and grid resolution 0.30 [45].

Drug Likeliness and Toxicity Analysis

QikProp application was used to find the ADME property [46]. Thirty-one QikProp parameters were considered for each molecule. QikProp efficiently evaluates pharmaceutically relevant properties for over half a million compounds per hour, making it an indispensable lead generation and lead optimization tool. Toxicity was analyzed for genotoxicity, rat model, skin sensitization, skin irritations, eye irritations, rat dosage tolerance, etc. by Accelrys' Discovery Studio—Toxicity Prediction by Komputer Assisted Technology (TOPKAT) program [47]. It evaluates your compounds' performance in experimental assays and animal models. Compute and validate assessments of the toxic and environmental effects of chemicals solely from their molecular structure. TOPKAT employs robust and cross-validated Quantitative Structure Toxicity Relationship models for assessing various measures of toxicity and utilizing the patented Optimal Predictive Space validation method to assist in interpreting the results [48].

Results and Discussion

Infectious diseases are identified as the second leading cause [49] for death worldwide. In spite of having an increasing demand for new antimicrobial drugs, the new drugs identified are less due to many reasons like huge investment and less market and competition with newly developed agents [50]. Many new algorithms, tools, and databases have been developed as a result of the advancement in Bioinformatics which has facilitated the automation of microbial genome sequencing, comparison of genomes, identification of gene product function, and simplified the process of development of antimicrobial agents, vaccines, and rational drug design [51]. In silico subtractive genomics approach is a

powerful approach to identify the specific genes which are present in the pathogen but absent in the host, thus helps in the identification in novel genus specific genes which can be used as drug targets. In silico drug target identification mainly relies on the principle "a good drug target is a gene essential for bacterial survival yet cannot be found in host" [12]. Novel drug targets have been identified successfully for various pathogens with the help of subtractive genomics approach.

Identification of Essential Genes

In the current study, non-human homolog essential genes of *Streptococcus species* as well as their protein products have been identified using subtractive genomic approach, which are likely to lead to the development of drugs that strongly bind with the pathogen.

The *S. pyogenes* SSI-1 consists of 1859 reference proteins which were downloaded from NCBI (www.ncbi.nlm.nih.gov) and this is subjected to BLASTP against selected strains of *Streptococcus species*, with E value of 10^{-4}. From the obtained 1,050 proteins which shared in all selected strains, the 223 hypothetical proteins were eliminated as hypothetical protein is a protein whose existence has been predicted, but for which there is no experimental evidence that it is expressed in vivo. The screened 827 proteins were analyzed using CD-HIT to identify duplicate proteins, which were identified using 60% identity as threshold with the CD-HIT tool. Out of 827 proteins, 25 proteins were present twice, so 1 set of these 25 proteins were eliminated. Remaining 802 proteins were analyzed with the help of BLAST against human using BLASTP. This revealed 406 proteins hit back with no significant similarity with human proteins. By using the DEG, 283 essential proteins were identified.

Metabolic Pathway Analysis and Identification of Novel Drug Targets

The obtained 238 essential proteins were analyzed using KAAS server. The involvements of drug targets in metabolic pathways were analyzed. Comparative analysis of the metabolic pathways of the host and pathogen was performed to trace out drug targets involved in pathogen specific metabolic pathways. Detailed pathway analysis revealed that 129 proteins were such that even after targeting them the organism could survive. In other words, these proteins could not act as a drug target. Hence these proteins were removed and 63 proteins, which are very crucial for the survival of the organism, were identified. Screening of drug targets was carried out using DrugBank, TTD, PDTD, and HIT for the 63 proteins identified from the pathway analysis. Out of 63 proteins, 6 proteins were already used for as the drug target: thioredoxin reductase (NADPH) [EC:1.8.1.9] (involved in pyrimidine and seleno-compound metabolism), 3-oxoacyl-[acyl-carrier-protein] synthase III (fabH) [EC:2.3.1.180] (involved in fatty acid biosynthesis), UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase (murA) [EC:2.5.1.7] (involved in peptidoglycan biosynthesis, amino sugar, and nucleotide sugar metabolism), *S*-adenosylhomocysteine/5′-methylthioadenosine nucleosidase (mtnN) [EC:3.2.2.9] (involved in cysteine and methionine metabolism), large subunit ribosomal protein L32, and small subunit ribosomal protein S4. Among the six said proteins, Azelaic acid (DrugBank ID: DB00548) targets the first one, Cerulenin (DrugBank ID: DB01034) targets the second one, Fosfomycin (DrugBank ID: DB00828) targets the third one, Adenine-approved nutraceutical (DrugBank ID: DB00173) targets the fourth one, Troleandomycin (DrugBank ID: DB01361) targets the fifth one, and Doxycycline (DrugBank ID: DB00254), Lymecycline (DrugBank ID: DB00256), Clomocycline (DrugBank ID: DB00453), Oxytetracycline (DrugBank ID: DB00595), Demeclocycline (DrugBank ID:

DB00618), and Minocycline (DrugBank ID: DB01017) target the last one. Figure 2 depicted the brief summary of target identification.

The identified 57 novel drug targets involved in the 6 pathways/biological process which are unique to pathogens and 17 pathways/biological process which are communal in both host and pathogen. All 23 pathways/biological processes were classified into 10 classes: amino acid metabolism, carbohydrate metabolism, energy metabolism, glycan biosynthesis and metabolism, lipid metabolism, metabolism of cofactors and vitamins, metabolism of other amino acids, nucleotide metabolism, genetic information processing, and environmental information processing (Table 1). Figure 3 enlightens the percentage distribution of novel drug targets involved in different metabolic pathways/biological process

Novel Targets in Pathogens' Unique Pathways

Current study shows that 16 proteins uniquely involved in the pathogen specific 6 unique pathways are peptidoglycan biosynthesis, D-alanine metabolism, phosphotransferase system (PTS), bacterial secretion system, ABC transporters, and two-component system.

Among the cytoplasmic steps involved in the biosynthesis of peptidoglycan (KEGG Pathway: map00550), the 4 enzymes of *Streptococcus species* uniquely present in this pathways are murB; UDP-*N*-acetylmuramate dehydrogenase [EC:1.1.1.158], murC; UDP-*N*-acetylmuramate–alanine ligase [EC:6.3.2.8], murD; UDP-*N*-acetylmuramoylalanine–D-glutamate ligase [EC:6.3.2.9], murF; and UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate–D-alanyl–D-alanine ligase [EC:6.3.2.10]. The peptidoglycan is a macro-molecule made of long aminosugar strands cross-linked by short peptides. It forms the cell wall in bacteria surrounding the cytoplasmic membrane. The glycan strands typically comprise repeating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) disaccharides. Each MurNAc is linked to a peptide of three to five amino acid residues. Disaccharide subunits are first assembled on the cytoplasmic side of the bacterial membrane on a polyisoprenoid anchor (lipid I and II). Polymerization of disaccharide subunits by transglycosylases and cross-linking of glycan strands by transpeptidases occur on the other side of the membrane. The enzymes involved in peptidoglycan biosynthesis are among the best known targets in the search for new antibiotics [52, 53].

The subunit 2 of dltC, D-alanine–poly (phosphoribitol) ligase [EC:6.1.1.13], is reported as uniquely present in this pathogen specific D-alanine metabolism (KEEG Pathway: map00473). In the surface charge as well as the electrochemical properties and ligand
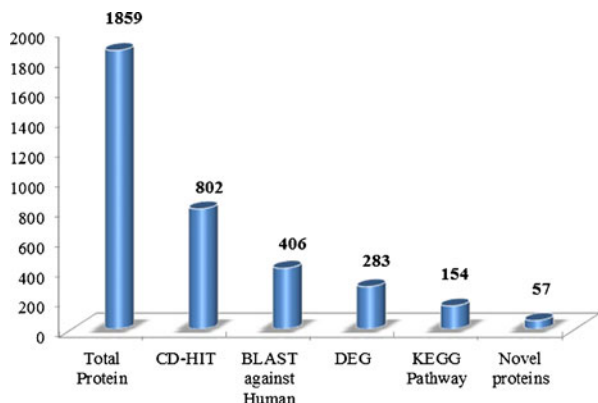
Fig. 2 Summary of target identification

Table 1 Novel drug targets involved in different metabolic pathways and other cellular activities

S. No. KEGG orthology number (ko), gene, and protein name

1. Amino acid metabolism (phenylalanine, tyrosine and tryptophan biosynthesis, cysteine and methionine metabolism, and lysine biosynthesis)

   1. ko:K00014 aroE; shikimate dehydrogenase [EC:1.1.1.25]

   2. ko:K00800 aroA; 3-phoshoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]

   3. ko:K00891 E2.7.1.71; shikimate kinase [EC:2.7.1.71]

   4. ko:K01929 murF; UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate–D-alanyl–D-alanine ligase [EC:6.3.2.10]

2. Carbohydrate metabolism (amino sugar and nucleotide sugar metabolism, and fructose and mannose metabolism)

   1. ko:K00075 murB; UDP-N-acetylmuramate dehydrogenase [EC:1.1.1.158]

   2. ko:K01809 E5.3.1.8; mannose-6-phosphate isomerase [EC:5.3.1.8]

   3. ko:K02794 PTS-Man-EIIB; PTS system, mannose-specific IIB component [EC:2.7.1.69]

3. Energy metabolism (oxidative phosphorylation and photosynthesis)

   1. ko:K01507 E3.6.1.1; inorganic pyrophosphatase [EC:3.6.1.1]

   2. ko:K02109 ATPF0B; F-type H+-transporting ATPase subunit b [EC:3.6.3.14]

4. Glycan biosynthesis and metabolism (peptidoglycan biosynthesis)

   1. ko:K00075 murB; UDP-N-acetylmuramate dehydrogenase [EC:1.1.1.158]

   2. ko:K01924 murC; UDP-N-acetylmuramate–alanine ligase [EC:6.3.2.8]

   3. ko:K01925 murD; UDP-N-acetylmuramoylalanine–D-glutamate ligase [EC:6.3.2.9]

   4. ko:K01929 murF; UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate–D-alanyl-D-alanine ligase [EC:6.3.2.10]

5. Lipid metabolism (fatty acid biosynthesis and glycerophospholipid metabolism)

   1. ko:K02371 fabK; enoyl-[acyl carrier protein] reductase II [EC:1.3.1.-]

   2. ko:K06131 cls; cardiolipin synthase [EC:2.7.8.-]

6. Metabolism of cofactors and vitamins (one carbon pool by folate and folate biosynthesis)

   1. ko:K00287 folA; dihydrofolate reductase [EC:1.5.1.3]

7. Metabolism of other amino acids (selenocompound metabolism, D-glutamine and D-glutamate metabolism, and D-alanine metabolism)

   1. ko:K01776 E5.1.1.3; glutamate racemase [EC:5.1.1.3]

   2. ko:K01924 murC; UDP-N-acetylmuramate–alanine ligase [EC:6.3.2.8]

   3. ko:K01925 murD; UDP-N-acetylmuramoylalanine–D-glutamate ligase [EC:6.3.2.9]

   4. ko:K14188 dltC; D-alanine–poly(phosphoribitol) ligase subunit 2 [EC:6.1.1.13]

8. Nucleotide metabolism (pyrimidine and purine metabolism)

   1. ko:K01589 purK; 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18]

   2. ko:K01591 pyrF; orotidine-5′-phosphate decarboxylase [EC:4.1.1.23]

   3. ko:K02337 DPO3A1; DNA polymerase III subunit alpha [EC:2.7.7.7]

   4. ko:K02338 DPO3B; DNA polymerase III subunit beta [EC:2.7.7.7]

   5. ko:K02340 DPO3D1; DNA polymerase III subunit delta [EC:2.7.7.7]

   6. ko:K03060 rpoZ; DNA-directed RNA polymerase subunit omega [EC:2.7.7.6]

   7. ko:K03763 DPO3A2; DNA polymerase III subunit alpha, Gram-positive type [EC:2.7.7.7]

9. Genetic information processing (transcription, translation, folding, sorting and degradation, and replication and repair)

   1. ko:K02314 dnaB; replicative DNA helicase [EC:3.6.4.12]

   2. ko:K02316 dnaG; DNA primase [EC:2.7.7.-]

   3. ko:K02337 DPO3A1; DNA polymerase III subunit alpha [EC:2.7.7.7]

   4. ko:K02338 DPO3B; DNA polymerase III subunit beta [EC:2.7.7.7]

Table 1 (continued)

S. No. KEGG orthology number (ko), gene, and protein name

5. ko:K02340 DPO3D1; DNA polymerase III subunit delta [EC:2.7.7.7]

6. ko:K02881 RP-L18; large subunit ribosomal protein L18

7. ko:K02888 RP-L21; large subunit ribosomal protein L21

8. ko:K02895 RP-L24; large subunit ribosomal protein L24

9. ko:K02902 RP-L28; large subunit ribosomal protein L28

10. ko:K02904 RP-L29; large subunit ribosomal protein L29

11. ko:K02907 RP-L30; large subunit ribosomal protein L30

12. ko:K02909 RP-L31; large subunit ribosomal protein L31

13. ko:K02913 RP-L33; large subunit ribosomal protein L33

14. ko:K02916 RP-L35; large subunit ribosomal protein L35

15. ko:K02933 RP-L6; large subunit ribosomal protein L6

16. ko:K02939 RP-L9; large subunit ribosomal protein L9

17. ko:K02961 RP-S17; small subunit ribosomal protein S17

18. ko:K02968 RP-S20; small subunit ribosomal protein S20

19. ko:K02970 RP-S21; small subunit ribosomal protein S21

20. ko:K02982 RP-S3; small subunit ribosomal protein S3

21. ko:K02990 RP-S6; small subunit ribosomal protein S6

22. ko:K03060 rpoZ; DNA-directed RNA polymerase subunit omega [EC:2.7.7.6]

23. ko:K03070 secA; preprotein translocase subunit SecA

24. ko:K03076 secY; preprotein translocase subunit SecY

25. ko:K03100 lepB; signal peptidase I [EC:3.4.21.89]

26. ko:K03470 rnhB; ribonuclease HII [EC:3.1.26.4]

27. ko:K03629 recF; DNA replication and repair protein RecF

28. ko:K03657 uvrD; DNA helicase II/ATP-dependent DNA helicase PcrA [EC:3.6.4.12]

29. ko:K03763 DPO3A2; DNA polymerase III subunit alpha, Gram-positive type [EC:2.7.7.7]

30. ko:K04066 priA; primosomal protein N' (replication factor Y) (superfamily II helicase) [EC:3.6.4.-]

10. Environmental information processing (membrane transport and signal transduction)

1. ko:K02759 PTS-Cel-EIIA; PTS system, cellobiose-specific IIA component [EC:2.7.1.69]

2. ko:K02313 dnaA; chromosomal replication initiator protein

3. ko:K02760 PTS-Cel-EIIB; PTS system, cellobiose-specific IIB component [EC:2.7.1.69]

4. ko:K02794 PTS-Man-EIIB; PTS system, mannose-specific IIB component [EC:2.7.1.69]

5. ko:K03070 secA; preprotein translocase subunit SecA

6. ko:K03076 secY; preprotein translocase subunit SecY

7. ko:K07652 vicK; two-component system, OmpR family, sensor histidine kinase VicK [EC:2.7.13.3]

8. ko:K07718 yesM; two-component system, sensor histidine kinase YesM [EC:2.7.13.3]

9. ko:K09811 ftsX; cell division transport system permease protein

10. ko:K09815 znuA; zinc transport system substrate-binding protein

11. ko:K11617 liaS; two-component system, NarL family, sensor histidine kinase LiaS [EC:2.7.13.3]

binding abilities of the Gram-positive cell wall is controlled by the D-alanylation of the lipoteichoic acid. The incorporation of D-Ala into lipoteichoic acid requires the D-alanine–D-alanyl carrier protein ligase (DltA) and the carrier protein (DltC) [18, 54].
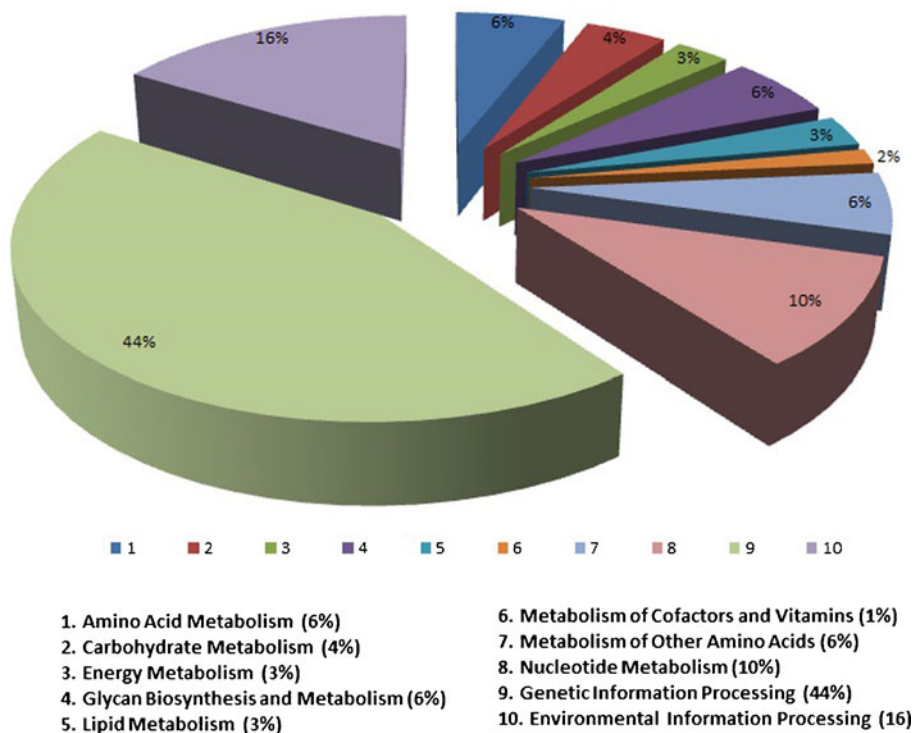
Fig. 3 Percentage distribution of novel drug targets involved in different metabolic pathways/biological process

The three enzymes, cellobiose-specific IIA component [EC:2.7.1.69], IIB component [EC:2.7.1.69], and mannose-specific IIB component [EC:2.7.1.69], were identified as unique enzymes present in PTS (KEGG Pathway: map02060). The phosphoenolpyruvate (PEP)-dependent PTS is a major mechanism used by bacteria for uptake of carbohydrates, particularly hexoses, hexitols, and disaccharides, where the source of energy is from PEP. The PTS catalyzes the uptake of carbohydrates and their conversion into their respective phosphoesters during transport [55, 56].

The two proteins, preprotein translocase subunit SecA and subunit SecY which are the parts of bacterial secretion system (Pathway: map03070), were identified as novel proteins. In Gram-positive bacteria, secreted proteins are commonly translocated across the single membrane by the Sec pathway or the two-arginine (Tat) pathway [57, 58].

The two transport proteins—cell division transport system permease protein (ftsX) and zinc transport system substrate-binding proteins (znuA) which are part of ABC transporters (KEGG Pathway: map02010)—are novel proteins. The ATP-binding cassette (ABC) transporters form one of the largest known protein families and are widespread in bacteria, archaea, and eukaryotes. Though this pathway is available in eukaryotes also, in this study, we consider as pathogen-specific pathway since in a typical eukaryotic ABC transporter, the membrane spanning protein and the ATP-binding protein are fused, forming a multi-domain protein with the membrane-spanning domain and the nucleotide-binding domain [59].

Total four novel proteins: dnaA; chromosomal replication initiator protein, vicK; two-component system, OmpR family, sensor histidine kinase VicK [EC:2.7.13.3], yesM; two-

component system, sensor histidine kinase YesM [EC:2.7.13.3] and liaS; two-component system, NarL family and sensor histidine kinase LiaS [EC:2.7.13.3], are present in two-component system (KEEG Pathway: map02020). Two-component signal transduction systems enable bacteria to sense, respond, and adapt to changes in their environment or in their intracellular state. Each two-component system consists of a sensor protein histidine kinase and a response regulator. It often enables cells to sense and respond to stimuli by inducing changes in transcription [60].

Novel Targets in Host–pathogen Common Pathways

Current study shows that total 47 proteins involved in 17 metabolic pathway, which are grouped as metabolism of amino acid, carbohydrate, energy, lipid, other amino acids, nucleotide, cofactors and vitamins, and genetic information processing (Table 1).

Thus, out of 63 proteins, 57 proteins were novel proteins in which 5 proteins were reported as a research target for various pathogens: 3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19] which involved in phenylalanine, tyrosine, and tryptophan biosynthesis [61, 62], 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18] involved in purine metabolism, orotidine-5′-phosphate decarboxylase [EC:4.1.1.23] involved in pyrimidine metabolism [63], mannose-6-phosphate isomerase [EC:5.3.1.8] involved in fructose and mannose metabolism, amino sugar, and nucleotide sugar metabolism [64], and signal peptidase I (lepB) [EC:3.4.21.89] which involved in protein export [65].

Subcellular Localization Prediction

Computational prediction of bacterial protein subcellular localization provides a quick and inexpensive means for gaining insight into protein function, verifying experimental results, annotating newly sequenced bacterial genomes, and detecting potential cell surface/secreted drug targets [66]. The protein localization study revealed that among 57 predicted novel drug targets, 41 proteins were present in cytoplasm, 13 were in cytoplasmic membrane, and 3 were unable to predict the protein localization. Reverse vaccinology is an emerging vaccine development approach that starts with the prediction of vaccine targets by Bioinformatics analysis of microbial genome sequences [67]. Subcellular location is considered as one main criterion for vaccine target prediction. However, more criteria have been added. For example, since it was found that outer membrane proteins containing more than one

Table 2 List of predicted outer membrane proteins

S.No	KEGG orthology number (ko), gene, and protein name
Outer membrane proteins with one transmembrane helix	
1.	ko:K01589 purK; 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18]
2.	ko:K03100 lepB; signal peptidase I [EC:3.4.21.89]
3.	ko:K07652 vicK; two-component system, OmpR family, sensor histidine kinase VicK [EC:2.7.13.3]
Outer membrane proteins with more than one transmembrane helix	
1.	ko:K06131 cls; cardiolipin synthase [EC:2.7.8.-]
2.	ko:K07718 yesM; two-component system, sensor histidine kinase YesM [EC:2.7.13.3]
3.	ko:K09811 ftsX; cell division transport system permease protein

transmembrane helix were, in general, difficult to clone and purify [68], the number of transmembrane domains for a vaccine target is often considered in Bioinformatics filtering. So, in this study only 6 outer membrane proteins were identified, out of which only 3 proteins were identified as having one or less than one transmembrane helix (Table 2): purK; 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18], lepB; signal peptidase I [EC:3.4.21.89] and vicK; and two-component system, OmpR family and sensor histidine kinase VicK [EC:2.7.13.3]. These proteins could be cloned and overexpressed to study the possibilities of the vaccine candidates.

3D Structure Prediction and Analysis

DNA polymerase III subunit beta involved in purine metabolism, pyrimidine metabolism, DNA replication, mismatch repair, and homologous recombination. A complex network of interacting proteins and enzymes is required for DNA replication. Generally, DNA replication follows a multistep enzymatic pathway. At the DNA replication fork, a DNA helicase (DnaB) precedes the DNA synthetic machinery and unwinds the duplex parental DNA in cooperation with the SSB. On the leading strand, replication occurs continuously in a 5 to 3

Fig. 4 I-TASSER predicted 3D structure of DNA polymerase III beta subunit

direction; whereas on the lagging strand, DNA replication occurs discontinuously by synthesis and joining of short Okazaki fragments. In prokaryotes, the leading strand replication apparatus consists of a DNA polymerase (pol III core), a sliding clamp (beta), and a clamp loader (gamma delta complex) [69–72]. Debmalya Barh and Anil Kumar suggested that the DNA polymerase III subunit beta can be the potential novel drug target for which till now no drugs are available. The beta subunit of DNA polymerase III involved pathways shown in Fig. 1 [15, 32].

The 3D structure of the DNA polymerase III subunit beta (GI 28894914 and Accession NP_801264.1) was modeled by ab initio protein modeling tool I-TASSER as no template is available in PDB. For threading alignments and iterative structural assembly simulations, the top ten proteins used were 2awaB, 2avtA, 3p16A, 2avtA, 2awaB, 2avtA, 2awaA, 3p16A, 1unnA, and 3p16A. The top model predicted by I-TASSER was considered for further work based on C-score (Fig. 4). C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [−5, 2], where a C-score of higher value signifies a model with a high confidence and vice versa. The predicted structure consists of 378 amino acids, 10 helices, 29 strands, 33 turns, 259 H-bonds, 3,001 bonds, and 2,951 atoms.
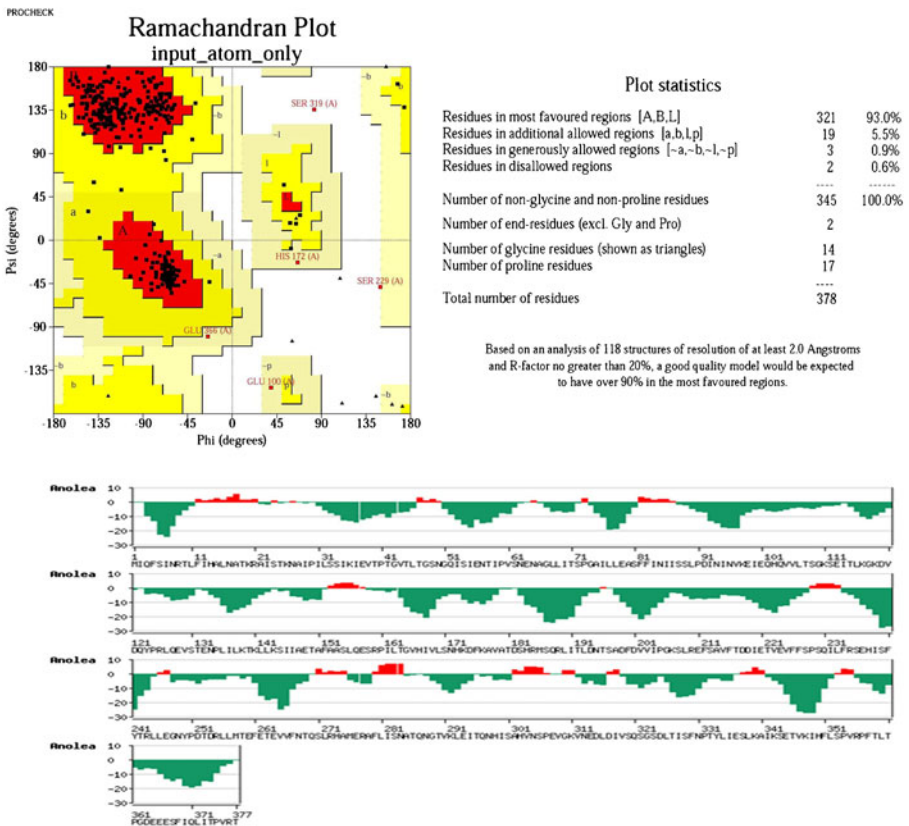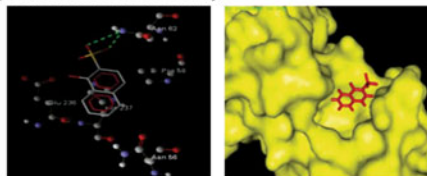

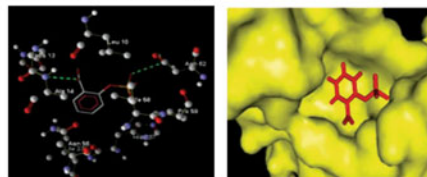Fig. 5 Predicted 3D structure quality analysis: Procheck and ANOLEA result

The quality of predicted structure was analyzed by PROCHECK and ANOLEA server. The PROCHECK program assessed using the Ramachandran plot. It is evident from the Ramchandran plot that the predicted model has 93.0%, 6.4%, and 0.6% residues in the most
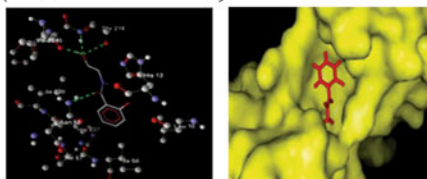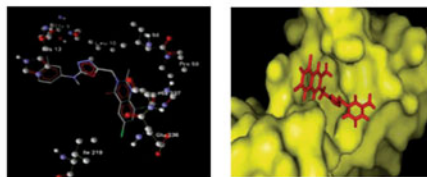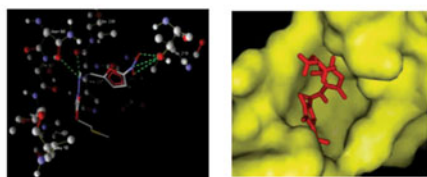

Fig. 6 **a** Docking result of top five synthetic and phytochemical chemical molecules with DNA polymerase III beta subunit. **b** Docking result of top five FDA approved drugs with DNA polymerase III beta subunit

Table 3 Docking energy of top five synthetic and phytochemical chemical molecules

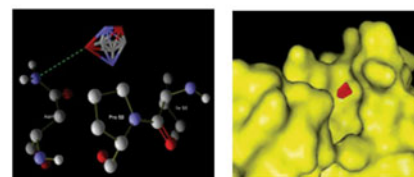| S. No	IUPAC Name of the Molecule	Pubchem CID	Glide	Mvd_ energy	Mvd rerank
1	1-Hydroxynaphthalene-2-sulfonic acid	11277	−6.730	−98.399	−77.561
2	2-Phosphonooxybenzoic acid	3418	−3.583	−98.399	−79.45
3	2-Hydroxy-N-(2-hydroxyethyl)benzamide	90407	−6.730	−94.011	−80.142
4	(E)-3-(4-Hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid	637775	−3.984	−91.943	−75.677
5	5-(Methylsulfanylmethyl)-3-[(E)-(5-nitrofuran-2-yl) methylideneamino]-1,3-oxazolidin-2-one	6433427	−2.86	−34.7763	−29.0985

favorable regions, the allowed regions, and the disallowed regions, respectively. Such a percentage distribution of the protein residues determined by Ramachandran plot shows that the predicted model is of good quality (a good quality model would be expected to have over 90% amino acids in the most favored region; Fig. 4). All Ramachandrans show 11 labeled residues out of 378 whereas chi1–chi2 plots show 8 labeled residues out of 227. The model shows all the main chain and side chain parameters to be in the "better" region. Another factor that is important for the predicted model to be reliable is G-factor, which is a log odds score based on the observed distribution of stereochemical parameters. For a reliable model, the score for G-factor should be above −0.50 [73]. The observed G-factor score for the present model was −0.09 for dihedrals bonds, −0.19 for covalent, and overall −0.12. The distribution of the main chain bond lengths and bond angles were 99.8% and 92.0% within limits, respectively. Also, all the planar groups were within the limits.

The ANOLEA result represents the graphical view of energy values of each amino acid. It shows that most of the amino acids having negative energy values (Fig. 5). The negative energy values (in green) represent favorable energy environment whereas positive values (in red) unfavorable energy environment for a given amino acid [40].

Virtual screening and docking

Molecular docking has played key role in the identification of efficient binding of receptor and ligand. Compounds identified from virtual screening with most favorable binding energy were considered as hits. From the docking studies, it was found that from ~195,500 synthetic and phytochemical chemical molecules only ~8,800 have the

Table 4 Docking energy of top five FDA approved drugs

S. No.	IUPAC name of the molecule	Generic name	DrugBank ID	Pubchem CID	Glide	Mvd_ energy	Mvd rerank
1	1-(Adamantan-1-yl)ethan-1-amine	Rimantadine	DB00478	5071	−4.663	−37.517	−62.635
2	[(2R,3S)-3-Methyloxiran-2-yl] phosphonic acid	Fosfomycin	DB00828	446987	−5.272	−36.514	−48.762
3	Phosphonoformic acid	Foscarnet	DB00529	3415	−3.984	−91.943	−38.385
4	Adamantan-1-amine	Amantadine	DB00915	2130	−6.005	9951.53	−11.622
5	4-Amino-1,2-oxazolidin-3-one	Cycloserine	DB00260	401	−3.890	−52.481	−43.104

complementary to binding sites; furthermore, only 700 were found to have efficient binding which was again reduced to 60 from ADME filtration using QikProp, and finally, only 11 molecules were predicted as non-toxic. The top five hits based on docking score of glide and MVD were shown in Fig. 6a. The top five FDA approved drugs were shown in Fig. 6b. The docking energy values are tabulated in Tables 3 and 4.

Current study contribute to the identification of five drug-like molecules: 1-hydroxynaphthalene-2-sulfonic acid, 2-phosphonooxybenzoic acid, 2-hydroxy-*N*-(2-hydroxyethyl)benzamide, (E)-3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid, and 5-(methylsulfanylmethyl)-3-[(E)-(5-nitrofuran-2-yl)methylideneamino]-1,3-oxazolidin-2-one which inhibits DNA polymerase III subunit beta.

1-Hydroxynaphthalene-2-sulfonic acid (CID 11277) has shown close structural similarity with experimental drug naphthalene-2,6-disulfonic acid which affects the target L-lactate dehydrogenase and bacterioferritin comigratory protein of bacteria [62].

2-Phosphonooxybenzoic acid (CID 3418) has shown close structural similarity with experimental drug 2-[(dioxidophosphino) oxy] benzoate, which affects the target D-alanyl–D-alanine carboxypeptidase of bacteria [62, 74].

2-Hydroxy-*N*-(2-hydroxyethyl)benzamide (CID 90407) was found to be non-nucleoside inhibitor of measles virus RNA-dependent RNA polymerase complex activity, and thus it can act as potential inhibitor for bacteria [75].

(E)-3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid (CID 637775) shows pharmacological action anti-infective agents this molecule is also found as experimental drug in DrugBank (db08586) which inhibits the protein feruloyl esterases [76]. The comparative docking studies of (E)-3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid to its native target feruloyl esterase and DNA polymerase III beta subunit have the lowest docking score than feruloyl esterase docking score −72.34.

From the literature studies, it was found that 5-(methylsulfanylmethyl)-3-[(E)-(5-nitrofuran-2-yl)methylideneamino]-1,3-oxazolidin-2-one (CID 6433427; common name: nifuratel) has a better spectrum of activity, without affecting lactobacilli, and it is a more effective inhibitor than the currently used antibiotics [77].

Conclusion

The in silico-based approach involves a series of screening of proteins that can be used as potential drug targets and vaccine candidates. The targets that were found are inevitable for the growth of the organisms and these proteins neither have a substitute protein nor an alternative pathway to accomplish the process. The current study carried out to design a drug that can block DNA polymerase III subunit beta proteins. It explore the possibilities of making new drugs from available chemical molecules and already approved drugs which used for other purposes also can be used to block DNA polymerase III subunit beta. The microorganisms are fast gaining resistance to the existing drugs, so designing better and effective drugs should be made faster. Thus, the current study can be the best replacement for current therapies available.

References

1. De Groot, A. S., et al. (2002). *Immunology and Cell Biology, 80*, 255–269.
2. Galperin, M. Y., & Koonin, E. V. (1999). *Current Opinion in Biotechnology, 10*, 571–578.
3. Koonin, E. V., et al. (1998). *Current Opinion in Structural Biology, 8*, 355–363.

4. Barh, D., et al. (2011). *Drug Development Research, 72*, 162–177.
5. Kanehisa, M., et al. (2010). *Nucleic Acids Research, 38*(Database issue), D355–D360.
6. Zhang, R., & Lin, Y. (2009). *Nucleic Acids Research, 37*, D455–D458.
7. Amineni, U., Pradhan, D., & Marisetty, H. (2010). *Journal of Chemical Biology, 3*(4), 165–173.
8. Rathi, B., et al. (2009). *Bioinformation, 4*(4), 143–150.
9. Koteswara Reddy, G., et al. (2010). *International Journal of Bioinformatics Research, 2*(2), 12–16.
10. Dutta, A., et al. (2006). *In Silico Biology, 6*(1–2), 43–47.
11. Sakharkar, K. R., et al. (2004). *In Silico Biology, 4*(3), 355–360.
12. Gupta, S. K., et al. (2010). *Journal of Antivirals & Antiretrovirals, 2*, 038–042.
13. Allsop, A. E., et al. (1995). *Bioorganic & Medicinal Chemistry Letters, 5*, 443–448.
14. Sharma, V., et al. (2008). *In Silico Biology, 8*(3–4), 331–338.
15. Barh, D., & Kumar, A. (2009). *In Silico Biology, 9*(4), 225–231.
16. Singh, S., et al. (2007). *Journal of Bioinformatics and Computational Biology, 5*(1), 135–153.
17. Anishetty, S., et al. (2005). *Computational Biology and Chemistry, 29*(5), 368–378.
18. May, J. J., et al. (2005). *FEBS Journal, 272*(12), 2993–3003.
19. Marrone, T. J., et al. (1997). *Annual Review of Pharmacology and Toxicology, 37*, 71–90.
20. Ryan, K. J., Ray, C. G. (2004) *Sherris Medical microbiology (4th ed. ed.)*. (McGraw Hill) ISBN 0-8385-8529-9.
21. Patterson, M. J. (1996). Streptococcus. In: S. Baron et al. (Ed.), *Baron's Medical Microbiology (4th ed.)*. Univ of Texas Medical Branch.
22. Albrich, W. C., Monnet, D. L., & Harbarth, S. (2004). *Emerging Infectious Diseases, 10*(3), 514–517.
23. Reinert, R. R., et al. (2004). *International Journal of Antimicrobial Agents, 24*(1), 43–47.
24. Gottschalk, B., et al. (2006). *Journal of Bacteriology, 188*(16), 5984–5992.
25. Nagai, K., et al. (2002). *Antimicrobial Agents and Chemotherapy, 46*(5), 1273–1280.
26. Byeon, W. H., & Weisblum, B. (1984). *Journal of Bacteriology, 158*(2), 543–550.
27. Knox, C., et al. (2011). *Nucleic Acids Research, 39*(Database issue), D1035–D1041.
28. Huang, Y., et al. (2010). *Bioinformatics, 26*(5), 680–682.
29. Wood, T. C., & Pearson, W. R. (1999). *Journal of Molecular Biology, 291*(4), 977–995.
30. Krissinel, E. (2007). *Bioinformatics, 23*(6), 717–723.
31. Online: www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606
32. Moriya, Y., et al. (2007). *Nucleic Acids Research, 35*(Web Server issue), W182–W185.
33. Chen, X., Ji, Z. L., & Chen, Y. Z. (2002). *Nucleic Acids Research, 30*(1), 412–415.
34. Gao, Z., et al. (2008). *BMC Bioinformatics, 9*, 104.
35. Ye, H., et al. (2011). *Nucleic Acids Research, 39*(Database issue), D1055–D1059.
36. Yu, N. Y., et al. (2010). *Bioinformatics, 26*(13), 1608–1615.
37. Krogh, A., et al. (2001). *Journal of Molecular Biology, 305*(3), 567–580.
38. Roy, A., et al. (2010). *Nature Protocols, 5*, 725–738.
39. Laskowski, R. A., et al. (1993). *Journal of Applied Crystallography, 26*, 283–291.
40. Melo, F., & Feytmans, E. (1998). *Journal of Molecular Biology, 277*(5), 1141–1152.
41. Hendlich, M., et al. (1997). *Journal of Molecular Graphics and Modelling*, 15359–15363.
42. Levitt, D. G., & Banaszak, L. J. (1992). *Journal of Molecular Graphics and Modelling*, 10229–10234.
43. Laurie, A. T., & Jackson, R. M. (2005). *Bioinformatics, 21*, 1908–1916.
44. Maestro, version 9.0, Schrödinger, LLC, New York, NY, 2009
45. Thomsen, R., & Christensen, M. H. (2006). *Journal of Medicinal Chemistry, 49*(11), 3315–3321.
46. QikProp, version 3.2, Schrödinger, LLC, New York, NY, 2009
47. TOPKAT, San Diego, CA: Accelrys, Inc
48. Greene, N. (2002). *Advanced Drug Delivery Reviews, 54*, 417–431.
49. World Health Organization (2002). The World Health Report.
50. Spellberg, B., et al. (2004). *Clinical Infectious Diseases, 38*, 1279–1286.
51. Bansal, A. K. (2005). *Microbial Cell Factories, 4*, 19.
52. Barreteau, H., et al. (2008). *FEMS Microbiology Reviews, 32*(2), 168–207.
53. Bouhss, A., et al. (2008). *FEMS Microbiology Reviews, 32*(2), 208–233.
54. Rogers, H. J., & Perkins, H. R. (1968). *Cell Walls and Membranes*. London: Spon.
55. Tchieu, J. H., et al. (2001). *Journal of Molecular Microbiology and Biotechnology, 3*(3), 329–346.
56. Kotrba, P., et al. (2001). *Journal of Bioscience and Bioengineering, 92*(6), 502–517.
57. Driessen, A. J., & Nouwen, N. (2008). *Annual Review of Biochemistry, 77*, 643–667.
58. Nakatogawa, H., et al. (2004). *Current Opinion in Microbiology, 7*(2), 145–150.
59. Tomii, K., & Kanehisa, M. A. (1998). *Genome Research, 8*(10), 1048–1059.
60. Gotoh, Y., et al. (2010). *Current Opinion in Microbiology, 13*(2), 232–239.
61. Jana, S., & Paliwal, J. (2007). *International Journal of Antimicrobial Agents, 30*(1), 4–10.

62. Overington, J. P., et al. (2006). *Nature Reviews. Drug Discovery, 5*(12), 993–996.
63. Bello, A. M., et al. (2009). *Journal of Medicinal Chemistry, 52*(6), 1648–1658.
64. Wills, E. A., et al. (2001). *Molecular Microbiology, 40*(3), 610–620.
65. Kulanthaivel, P., et al. (2004). *Journal of Biological Chemistry, 279*(35), 36250–36258.
66. Gardy, J. L., & Brinkman, F. S. L. (2006). *Nature Reviews Microbiology, 4*, 741–751.
67. Rappuoli, R. (2000). *Current Opinion in Microbiology, 3*(5), 445–450.
68. Pizza, M., et al. (2000). *Science, 287*(5459), 1816–1820.
69. Wijffels, G., et al. (2011). *Journal of Medicinal Chemistry, 54*(13), 4831–4838.
70. Berdis, A. J. (2008). *Biochemistry, 47*(32), 8253–8260.
71. Oberg, B. (2006). *Antiviral Research, 71*(2–3), 90–95.
72. Fang, J., et al. (2011). *Biochemistry, 50*(26), 5958–5968.
73. Morris, A. L. (1992). *Proteins, 12*(4), 345–364.
74. Imming, P., Sinning, C., & Meyer, A. (2006). *Nature Reviews. Drug Discovery, 5*(10), 821–834.
75. White, L. K., et al. (2007). *Antimicrobial Agents and Chemotherapy, 51*(7), 2293–2303.
76. Tarbouriech, N. (2005). *Acta Crystallographica Section D: Biological Crystallography, 61*(Pt2), 194–197.
77. Togni, G., et al. (2011) *Antimicrobial Agents and Chemotherapy.* 2011 Feb 14.