Eng. Sci., 2022, **19**, 125–135

Engineered Science

DOI: https://dx.doi.org/10.30919/es8d645

Comparative Genomics Based Putative Drug Targets Identification, Homology Modeling, Virtual Screening and Molecular Docking Studies in *Chlamydophila Pneumoniae*

R. C. Shivamurthy,[1] Ravindranath B. S.,[2,*] Santhosh K. H.,[3] Murthy B. R. N.[1] and Sandeep Nambiar S.[1]

Abstract

Chlamydophila pneumoniae is an obligate intracellular Gram-negative human respiratory pathogen acquiring resistance to current antibiotics. Novel strategies including comparative genomics are vital to discover new drug targets to overcome global antibiotic resistance. The current study's primary objective is to identify novel drug targets based on path-genome (synteny) analysis and prioritize the chronic obstructive pulmonary disease (COPD) therapeutic targets. The orthologous syntons based on patho-genomes were compared using SynteView, and the classification of orthologous genes based on non-host prediction, essentiality analysis, protein-protein interactions study, toxicity, and druggability analysis were directional in identifying potential drug targets. A total of six hundred and thirty-nine orthologous syntons were identified and classified by prioritizing nrdA and rpoB as potential drug targets. Virtual screening of natural ligands led to the selection of twenty-one ligands, and further ADMET analysis was effective in considering nine ligands. Molecular docking analysis was crucial in analyzing the binding affinity and different binding modes of the nine ligands selected and prioritizing two ligands (CNP0057221 (-6.1 kcal/mol), CNP0372149 (-5 kcal/mol) as potential putative drugs subjected to Molecular dynamics simulation and in vitro based evaluation. The results reported may be vital in novel drug research in the current multi-drug resistance and extensive-drug resistance situation.

Keywords: Syntons; Orthologous; COPD, Virtual screening; Molecular docking.
Received: 30 October 2021; Revised: 19 January 2022; Accepted: 05 February 2022.
Article type: Research article.

1. Introduction

Comparative microbial genomics is a field of modern biology where a wide range of bacterial genomes are compared to derive the pathogenicity facilitating drug target discovery.[1,2] The conservation of genomic features is referred to as "synteny," which demonstrates the relative positions of genes. Exploiting the conserved synteny blocks from multiple bacterial genomes answers the classification of genes into orthologous regions by providing insight into the bacterial phylogenomic relationship.[3,4]

Many bacterial species have acquired the potential resistance against multiple antibiotics with currently marketed antibiotics through their competent ability, which has led to a global concern.[5] *Chlamydophila pneumoniae*, a Gram-negative human respiratory pathogen,[6] demonstrates a biphasic developmental form of replication, *i.e.*, elementary and reticulate bodies. elementary body (EB) denotes the small-sized (ca. 0.3 μm), circular, infectious form, electron-dense organism, whereas the Reticulate body (RB) refers to comparatively larger (ca. 1 μm), a non-infectious internalized form of elementary bodies. These elementary bodies undergo primary differentiation leading to granular cytoplasm, which diffuses fibrillar nucleic acids.[7] It displays a significant contribution to persistent and acute infection, leading to chronic obstructive pulmonary disease (COPD) and asthma.[8] Acute *C.* pneumoniae infection can trigger pneumonia, bronchitis, and emphysema.[9]

C. pneumoniae is also investigated for several systemic diseases, including atherosclerosis, Alzheimer's disease,[10]

[1] *Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal-576104, Karnataka, India.*
[2] *Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal-576104, Karnataka, India.*
[3] *Department of Biotechnology and Bioinformatics, Kuvempu University, Shankaraghatta - 577451, Karnataka, India.*
**E-mail:* ravindranath.bs@manipal.edu (B.S. Ravindranath)

multiple sclerosis, cerebrovascular stroke, coronary artery disease, and lung cancer.[11] It has also been associated with zoonotic respiratory infections in koalas, horses, and amphibians.[12] It indicates the need for novel, sophisticated diagnostic techniques and uncovering the druggable gene in the genome to discover putative drug targets and drug discovery using novel computational technologies.

The current study mainly focuses on identifying the potential drug targets based on orthologous genes across the Chlamydial species by aligning multiple genomes based on synteny and applying the *in silico* strategies, including synteny-based genomes comparison. Significant synteny blocks can be identified by comparing neighboring or distantly related bacterial species. The synteny blocks (genes) adjacently located are recognized as significant candidates for genome-based evolution and functional annotation studies. To perform the genome-based analysis, visualization of conserved synteny blocks among the compared bacterial genomes is vital.[13]

Prioritization and validation of the targets is a significant step in target-based drug discovery studies, as the prioritized target is the key to bacterial inhibition. Due to the lack of crystal structure, comparative homology modeling is an essential step to model the structure and virtual screening of natural compounds to identify the molecules with interactive potential. Chemical absorption, distribution, metabolism, excretion, and toxicity (ADMET) analysis and molecular docking simulation studies were significant to identify novel putative drug compounds from natural sources. The above protocol proposed in the current study successfully identified the potential drug targets by which the inhibition of these targets can lead to the inhibition of bacterial pathogenicity. Modeling the target protein helped predict the protein model and was utilized in docking studies. The identified natural ligand molecules with significant binding potential (docking scores) in this study may be directional in developing novel drugs leading to drug discovery against COPD targets.

2. Experimental section

2.1 Orthologous syntons-based patho-genomes comparison

SynteView 1.0 was engaged to predict the orthologous synteny blocks in all the other selected Chlamydial pathogens analyzed in the study. The web-service mode was applied to retrieve and investigate the genomic data in SynteBase.[13] Reference species were selected based on the literature survey, "Reference tab" and the "Compared tab" was used for the selection of reference species (*Chlamydophila pneumoniae* AR39) and compared species (*Chlamydophila pneumoniae* J138, *Chlamydophila pneumoniae* TW183, *Chlamydophila pneumoniae* CWL029, *Chlamydia trachomatis* D/UW-3/CX, *Chlamydia psittasi*, *Chlamydia muridarium* Nigg, *Chlamydophila caviae* GPIC, *Chlamydophila abortus,* and *Chlamydophila felis*) belonging to the same family. The "Start retrieval process tab" was operated for comparing genomes and predicting orthologous synteny blocks. SynteView gives the result by comparing proteomes based on the given condition, *i.e.*, E-value $\leq 10^{-5}$ and similarity percentage > 35%, and the alignment was stretched for nearly 80% length of all the shorter matching proteins.[13] The SynteView window is displayed as in Fig. S1(see supporting information).

2.2 Prediction of orthologous genes/proteins

SynteView displays the significant conservation of gene order between reference and compared species (Fig. S1). A unique colored rectangular box illustrates each gene present in reference and compared species. The yellow, grey and blue colored rectangular boxes represent positive, negative, and non-orthologous genes. The orthologous gene was carefully chosen and considered for further analysis depending on the color code. "Gene information panel" helped retrieve protein sequence and its information like GenInfo Identifier (GI) number, species name, gene name, its functionality, location, and sequence information. The protein sequence was further utilized to retrieve the protein information for further analysis.[14] The genome map depicts the features of the genome, the genes prioritized, and their loci on the genome (see Fig. S2).

2.3 Non-host prediction

BLASTP (Protein-protein BLAST) search analysis was performed between the host and pathogen sequence at http://blast.ncbi.nlm.nih.gov/, to predict non-host proteins in *C. pneumoniae*. The NCBI database was explicitly searched against the *Homo sapiens* proteome (Tax Id: 9606) using the available BLASTP parameters.[15] Specific protein hits meeting the given criteria E-value $\leq E^{-10}$ and sequence identity < 35% were selected carefully and stored for further analysis.

2.4 Essentiality and subcellular localization analysis

The orthologous protein sequences predicted in the *C. pneumoniae* were submitted to BLASTP to search explicitly against the Database of Essential Genes (DEG 7.0) at http://tubic.tju.edu.cn/deg/ for the prediction of essential genes. DEG database is the repository of the sequence information of essential genes/proteins, and it archives the related literature of experimentally derived essential genes or proteins from Gram-negative and Gram-positive bacteria. The database search criteria were E-value < E^{-10}, Bit score >100, and identity percentage ≥35% at the amino acid level was considered.[16] Prediction of subcellular localization of the filtered orthologous, non-host, and essential genes was performed using PSORTB v3.0.2[17] and was cross-verified using TMHMM and TOPCONS.[18,19]

2.5 Protein-protein interaction (PPI) analysis

Protein-protein interactions (PPI) analysis has been crucial in outlining the signaling cascades, predicting the protein functionality,[20] and also finding the associated proteins with the specific disease,[21] which enables identification of the

interacting drug and its activity. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) predicts and builds protein-protein interaction networks based on physical and functional associations by considering the experimental data, domain-fusion, phylogenetic profiling, and gene neighborhood analysis (neighborhood, gene fusion, co-occurrence, and co-expression).[22]

2.6 Toxicity prediction

Genes/proteins responsible for the bacterial toxicity have been predicted and further classified into exotoxins and endotoxins using the BTXpred tool (http: //www. imtech. res. in/raghava/btxpred/),[23] later the protein sequences were submitted to Interpro server 41.0 (http: //www. ebi. ac. uk/interpro/)[24] to extrapolate the domains, signatures, and functions.

2.7 Druggability analysis

For a drug target, the gene must possess the affinity to interact with a drug or a small molecule with an excellent binding affinity.[25] Druggability was analyzed by performing a sequence search in DrugBank. The druggability of the shortlisted genes/proteins producing toxins in the *C. pneumoniae* pathogenic system was investigated using the Drug Bank database (http://www.drugbank.ca/)[26] by using a Pharma search. BLASTp search was performed against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (http: //www. genome. jp/kegg/)[27] to analyze which pathway they are involved in. All the drug targets predicted were found to be members of distinctive metabolic pathways of *C. pneumoniae*.

2.8 Virtual screening of natural compounds

Ligand molecules interacting with the predicted drug target were searched using the COCONUT (Collection of Open Natural Products) database[28] and ChEMBL[29] using the drug target information. The COCONUT database is a publicly accessible database that includes compounds from a natural source. ChEMBL is a manually curated bioactive compound repository that exhibits drug-like properties that integrate chemical information with bioactivity to facilitate drug discovery.

2.9 Homology modeling (Ribonucleotide diphosphate reductase (RNR))

Due to the unavailability of the RNR crystal structure in the Protein data bank (PDB) database, prediction of the RNR three-dimensional (3D) structure was inevitable to understand the protein-ligand molecular interactions. Homology modeling-based modeling of the RNR structure was initiated by submitting the protein sequence to the Iterative Threading ASSEmbly Refinement (I-TASSER) server.[30] It uses a hierarchical approach to predict the protein structure by searching the templates from the PDB database through the multiple threading method. Modeled protein was refined by submitting it to MODREFINER.[31] Later multiple tools were employed to validate the modeled structure, including PROCHECK,[32] ERRAT,[33] and Verify 3D.[34] For energy minimization and to visualize the modeled structures, UCSF Chimera[35] and PyMOL[36] were utilized.

2.10 Molecular docking simulation (AutoDock vina)

A rigid docking simulation strategy was adopted to analyze the binding ability of the interaction of the ligands with the target protein. AutoDock vina[37] was employed to perform the docking simulation to identify the protein-ligand affinity and best binding mode. Based on the binding pocket predicted using CastP (Computer Atlas of Surface Topography of Proteins)[38] the grid box parameters were set as x = 72.317, y = 79.405 & z = 70.661. The grid box size (x = 22.358, y = 24.447 & z = 25.860) was resized based on x, y, and z coordinates to fit in the binding pocket residues of the target protein within the grid box, and the exhaustiveness was set to eight. The ligand molecules were sketched using Marvin Sketch[39] and were saved in Spatial Data File (.sdf) format. Open Babel software[40] was utilized to carry out the energy minimization of the ligand molecules. The best-scoring protein-ligand complex was prioritized based on the binding affinity and the lowest docking (binding) scores for post-dock analysis. PyMol was adopted to visualize the protein-ligand complex precisely for post-dock analysis, and two-dimensional (2D) plots were generated using Discovery Studio to analyze the specific type of interactions.

3. Results and discussion

3.1 Classification and prediction of orthologous genes

Orthologous synteny-based comparative genomics (OSBCG) explored the possible conserved and non-conserved regions with the reference genome and were aligned with one-to-many related genomes. This led to the classification and prediction of orthologous and non-orthologous genes. This data set can be further considered in identifying the potential drug targets with the modern approach. Orthologous regions in the genome were predicted compared to other Chlamydial organisms. Six thirty-nine genes are highly conserved in all organisms. This gives insight into the key regions corresponding to the functionality, and other vital processes were identified as orthologous synteny blocks excluding hypothetical proteins that are common in all the remaining Chlamydial organisms. Four hundred and fourteen non-orthologous synteny blocks were predicted consisting of genes with no orthology between the two genomes. They were classified based on synteny alignments, ontology, manual curation, and publicly available information in SynteBase from a high-quality and specific set of orthology predictions, as provided by SynteView (Fig. 1).

3.2 Non-homologous protein analysis

The protein sequences without significant homology to human

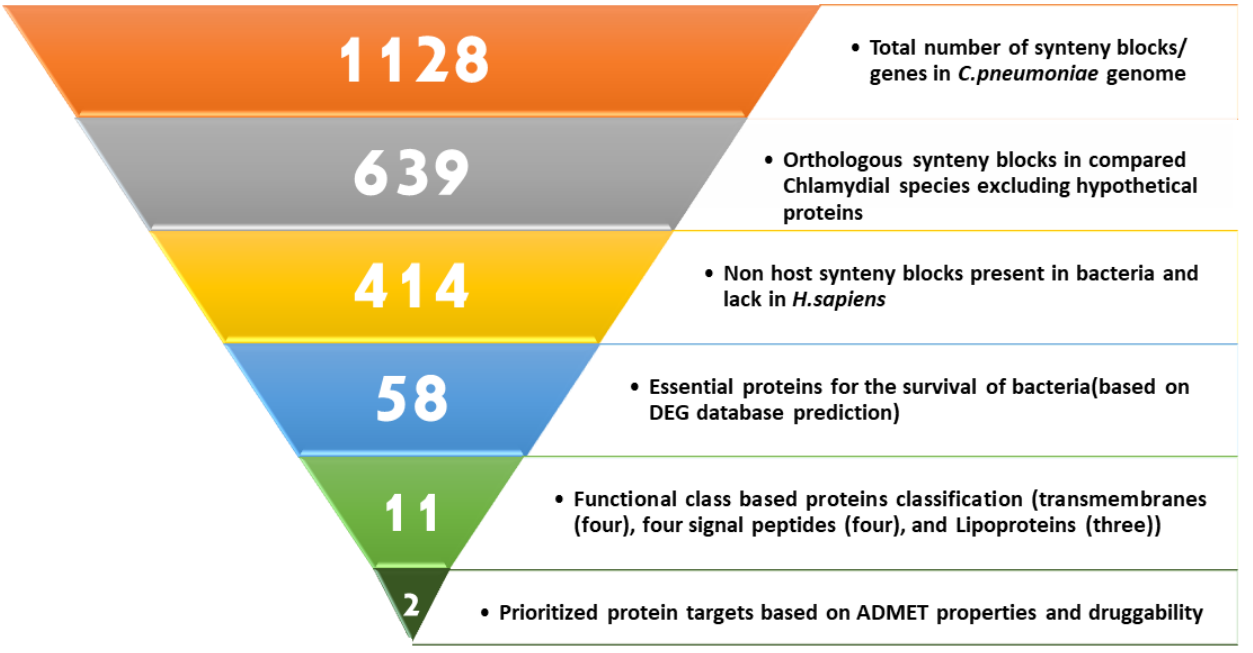


Fig. 1 Illustration of the step-wise sifting of synteny blocks (drug targets).

proteins were designated as non-homologous proteins. The set of proteins listed from the earlier step was searched using the BLASTP service available to perform a similarity search against the human proteome non-redundant database to identify non-homologous proteins. Four hundred fourteen proteins were predicted as non-homologous to the human proteome based on criteria like identity percentage ≥35%, E-value threshold cut off 10^{-10}, and amino acid length 100.

3.3 Essentiality analysis

The availability of databases like the DEG database of essential genes has paved the way to predict essential genes by comparing the proteins against experimentally verified essential proteins. The listed proteins from the earlier step of the analysis were submitted to Protein BLAST search against the DEG database with the criteria like E-value cut-off score of 10^{-10} and ≥35% identity percentage and Bit score 100, fifty-seven essential genes were predicted in the present study comprising ~4% of the total number of protein sequences in *C. pneumoniae*. These short-listed proteins were considered for further critical analysis in the next step.

3.4 Protein-protein interactions prediction

Protein-protein interaction analysis is crucial in functional annotations of the proteins has supportive of understanding the biological role of the protein within the cell environment. The development of novel tools, algorithms, and methods has enabled us to predict protein-protein interactions, which are experimentally proven to study the protein-protein interaction with a basic idea that proteins involved in similar functions will interact with each other to form a network. Only those proteins with a high confidence score of 0.900 were considered for listing out eleven potential protein interactions (Table. S1). The evidence-based interactions are demonstrated in Fig. 2.

3.5 Classification into functional class

To get an insight into the shortlisted essential proteins, they were further classified into transmembranes based on TMHMM analysis, lipoproteins using LipoP analysis,[28] and signal peptides prediction using SignalP[29] functional classes, which are the favorites to be the potential drug candidates, from the present analysis. They were classified into transmembranes (four), four signal peptides (four), and Lipoproteins (three). Weeded out the left-out proteins that do not fall into this functional class (Table S2).

3.6 Bacterial toxins prediction

Prediction of bacterial toxins has been an essential strategy in drug target discovery. These toxins are reported as the primary cause of pathogenicity in humans displaying visible symptoms during infection. BTXpred effectively predicts bacterial toxins based on Support Vector Machine (SVM) modules with an accuracy of 96.07%. Further, SVM-based modules discriminate endotoxins and exotoxins, using amino acid composition with an accuracy of 95.71%. In addition, modules Hidden Markov Models (HMM) and PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) in combination classify the exotoxins with higher accuracy.[41]

3.7 Druggability analysis and prioritization of therapeutic targets

It is pivotal in prioritizing the short-listed proteins as drug targets by assigning various criteria that a drug target must possess other than the earlier filtration steps followed after classifying identified essential proteins into four trans

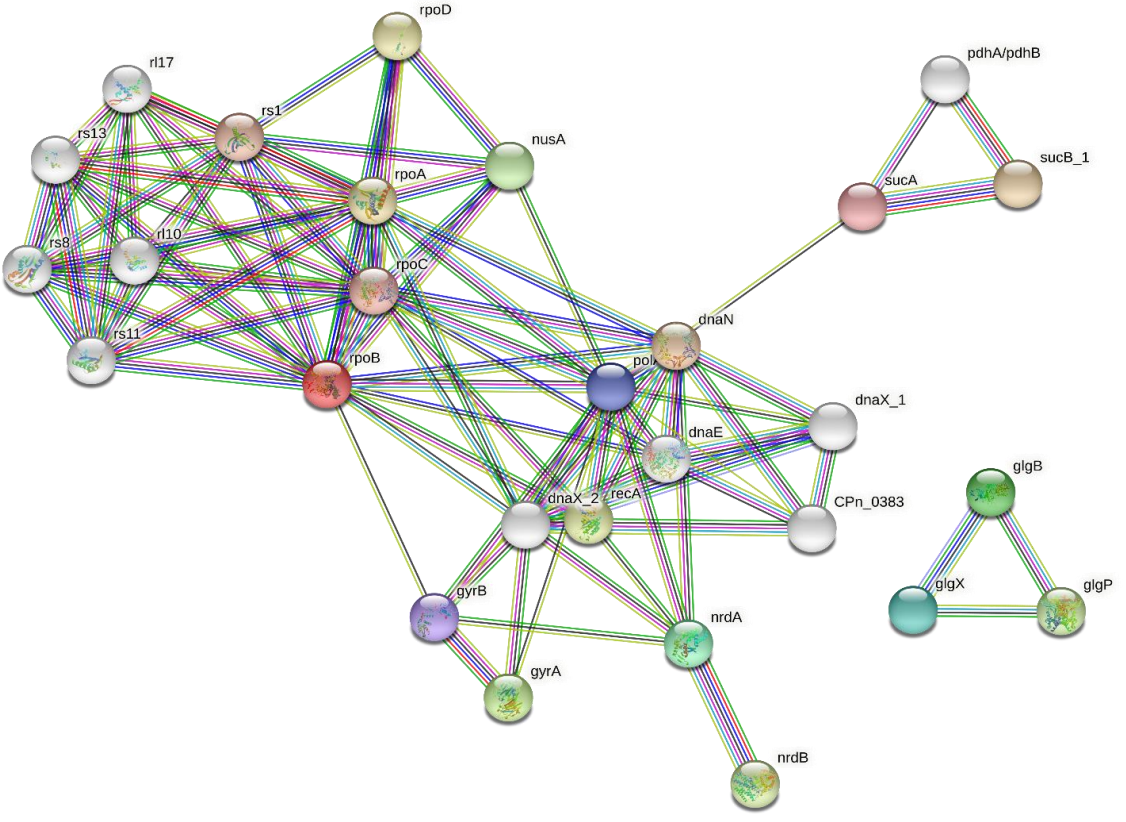


Fig. 2 Depicting the integrated protein-protein interaction (evidence-based interaction) network of proteins with high confidence scores of >0.900.

membranes, four signal peptides, and three lipoproteins. Further same were subjected to prioritization of drug targets with the criteria of subcellular localization, toxicity, molecular weight, availability of 3D protein structure, and FDA (Food and Drug Administration) drug availability for the possible drug targets identified. The short-listed essential proteins were classified into nineteen cytoplasmic proteins, two outer membrane proteins, and one extracellular protein using PSORTb v3.0.2 software in the subcellular localization step. Toxicity was predicted to identify toxins in the short-listed proteins and classified into nine exotoxins and three endotoxins among fifty-seven proteins using BTXpred software. Molecular weight was predicted using Compute pI/Mw[42] as the standard molecular weight of a target protein must be 110 KD remaining were weeded out.

Protein 3D structure search using RCSB PDB (Research Collaboratory for Structural Bioinformatics, Protein Data Bank) database yielded a single structure with PDB ID 3IYD_C belonging to DNA-directed RNA polymerase subunit beta (RNR) with poor identity percentage, and it was also analyzed that RCSB PDB database[43] holds only four experimentally solved structures belonging to *C. pneumoniae* which has been a limitation for the computational drug discovery. Druggability was predicted for the identified therapeutic targets in the COCONUT and CHEMBL databases, ribonucleotide diphosphate reductase (UniProtKB: Q9Z6S5.1) has been prioritized as the potential therapeutic target in this present study which meets all the criteria's to be a drug target. Further, molecular docking simulation studies were carried out using ribonucleotide diphosphate reductase (RNR) as the protein target to analyze the interactions (binding modes and binding affinity) between the compounds screened based on virtual screening.

3.8 Homology modeling and validation

To analyze the structural features and the drug designing purpose, the RNR protein was modeled based on Homology modeling by submitting query sequences to the I-TASSER server, the modeled protein structure was further subjected to model refining using MODREFINER. Refined model results showed the Root Mean Square Deviation (RMSD) of 1.488 and TM-score = 0.9796 from the initial to refined model, the same was visualized using PyMol (Fig. 3). Further, the modeled protein was validated using PROCHECK, ERRAT, and VERIFY3D. PROCHECK results yielded a Ramachandran plot indicating 90.8% residues in most favored regions, 7.6% residues in additional allowed regions, 0.6% in the generously allowed region, and only 1.0% residues in disallowed regions. VERIFY3D resulted in 81.21% of the residues having an average of 3D-1D score >= 0.2 ranked the model as PASS indicating the high quality of the model (Fig. S3) and ERRAT

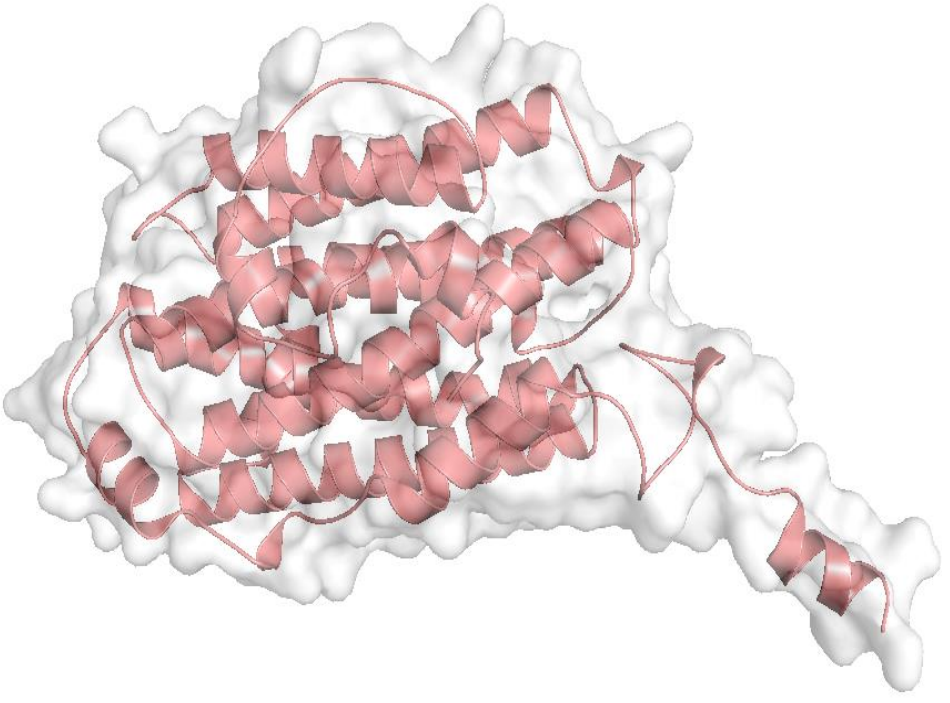


Fig. 3 Modeled protein structure of Ribonucleotide-diphosphate reductase (*C. pneumoniae*) is illustrated in the cartoon (deep salmon) and surface (white).

results indicated the model's overall quality factor with *Z*-score = 94.97% in comparison with similar-sized proteins based on the statistical scores of non-bonded interactions with different types of atoms. The validation tools were evident in validating the model's quality.

3.9 Natural ligands search

COCONUT database advanced search yielded twenty-one compounds interacting with query RNR protein which is a significant step in identifying the appropriate ligands based on specific protein-ligand interactions. ChEMBL database search reiterated the identified molecules as interacting molecules with RNR based on its compound report card. Further, the identified molecules were filtered based on ADMET property analysis including Molecular weight, TPSA (Topological Polar Surface Area) score, iLOGP, ESOL (Estimating Aqueous Solubility) Class, GI absorption, BBB (Blood Brain Barrier) permeant, Lipinski violations, Bioavailability (Table S3). Further, the selected (ten) compounds were subjected to *in silico* molecular docking simulation study.

3.10 *In silico* docking analysis (AutoDock vina 4.2)

Molecular docking studies were carried out with the screened ligand molecules to have possible interactions with the RNR target. The modeled RNR protein was considered a protein target to study the docking simulation analysis. It revealed the potential binding affinity with relatively higher binding energy scores against the commercially available drug (Cefotaxime) administered for *C. pneumoniae* infections, where the bacterial cell wall biosynthesis is inhibited by interacting with the penicillin-binding proteins.

Modeled RNR protein was subjected to docking analysis using AutoDock vina 4.2, the docking scores with significant interactions were recorded for the selected eleven compounds with the binding scores in the range of -6.1 to -4.8 kcal/mol, where the Cefotaxime exhibited -4.1 kcal/mol of binding energy score indicating relatively weaker binding affinity than all the molecules selected in this study. The binding energy scores with the RMSD upper bound and lower bound values are shown in Table 1. The molecules were prioritized based on ADMET properties and the docking scores, which exhibited a maximum docking score of -6.1 kcal/mol by interacting with binding site residues, including THR-336 (2.3 Å), ARG-302 (3.5 Å), and THR-333(2.7 Å) forming three H-bonds, conserved residue ARG-302 (2.7 Å) forming a carbon H-bond and a Vander Waals interaction with GLU-301 (2.8 Å) as shown in Fig. 4.

The binding modes of all the interacting selected molecules with the protein target (RNR) are shown in Fig. 4. Cefotaxime molecule formed interaction with residues (ARG-298 (3.4 Å) & ARG-302 (2.7 Å), forming two H-bonds. An unfavorable Positive-Positive interaction and a salt bridge interaction with ARG-302 residue were formed, including hydrophobic interaction with GLU-301(2.3 Å) but exhibited a relatively low binding score (-4.1 kcal/mol), leading to a weaker interaction. 2D plots were generated using Discovery studio to get better insights into the type of interactions formed between protein targets and interacting ligands (see Fig. 5).

Further, the post-dock analysis revealed the conserved binding sites (ARG-298, THR-336, THR-333, GLU-301, and

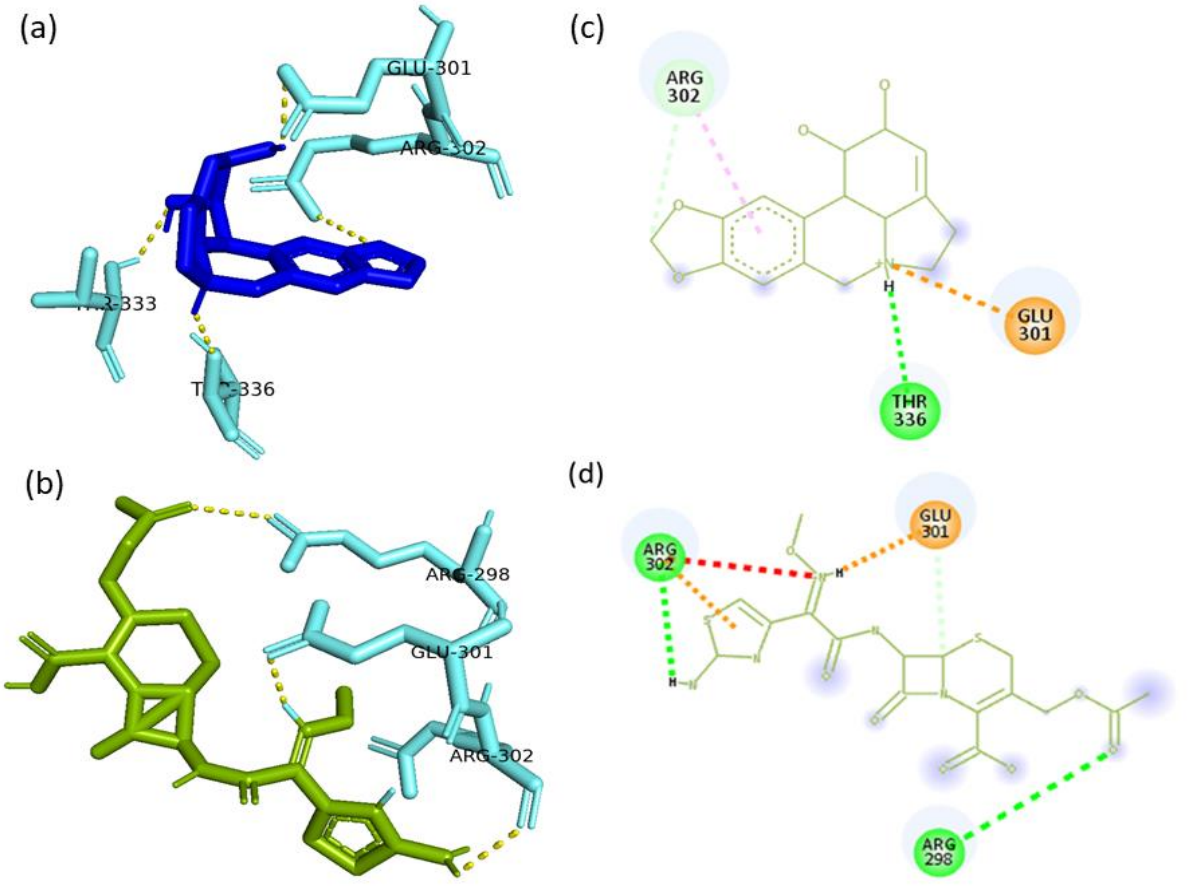


Fig. 4 (a) Illustration of docking poses with prioritized ligand (CNP0057221) (blue) of RNR (aquamarine) (b) Cefotaxime (standard drug) (split pea); (c, d) 2D plots.

Table 1. List of selected natural ligands with comparatively significant binding scores at zero RMSD.

Sl. No.	Ligands (COCONUT ID)	Binding energy (kcal/mol)	Hydrogen bond Interactions (Å)	Alkyl/Pi-alkyl	Carbon-hydrogen interactions	Pi-Sigma/Pi-cation/Pi-anion	Van der Wall's interaction
1.	CNP0064963	-4.8	THR-336 (3.0)	--	--	GLU-301, ARG-302	--
2.	CNP0057221	-6.1	THR-336 (2.3), (3.5), and THR-333 (2.7)	ARG-302	ARG-302	--	GLU-301
3.	CNP0160386	-5	THR-333 (2.8)	ARG-302	PHE-331	ARG-302	GLU-301
4.	CNP0417500	-5	ARG-298 (2.6, 3.4), THR-333 (2.8)	--	GLU-301	GLU-301	GLU-332, VAL-335, THR-336
5.	CNP0170992	-4.8	THR-333 (2.8)	ARG-298, VAL-335, ARG-302	GLU-301	ARG-302	--
6.	CNP0361821	-5.3	ARG-298 (2.8), ARG-302 (3.1)	ILE-235, ARG-302	--	GLU-301	--
7.	CNP0372149	-5	ARG-298 (2.6, 3.4) THR-333 (2.8)	--	GLU-301	GLU-301	THR-336, VAL-335, GLU-332
8.	CNP0108893	-5	THR-336 (2.1)	ARG-298, ILE-235, ARG-302	--	--	GLU-301
9.	CNP0412918	-4.8	--	VAL-335, PHE-331	--	ARG-302	--
10.	Cefotaxime	-4.1	ARG-298 (3.4), ARG-302 (2.7)	--	--	--	GLU-301, ARG-302

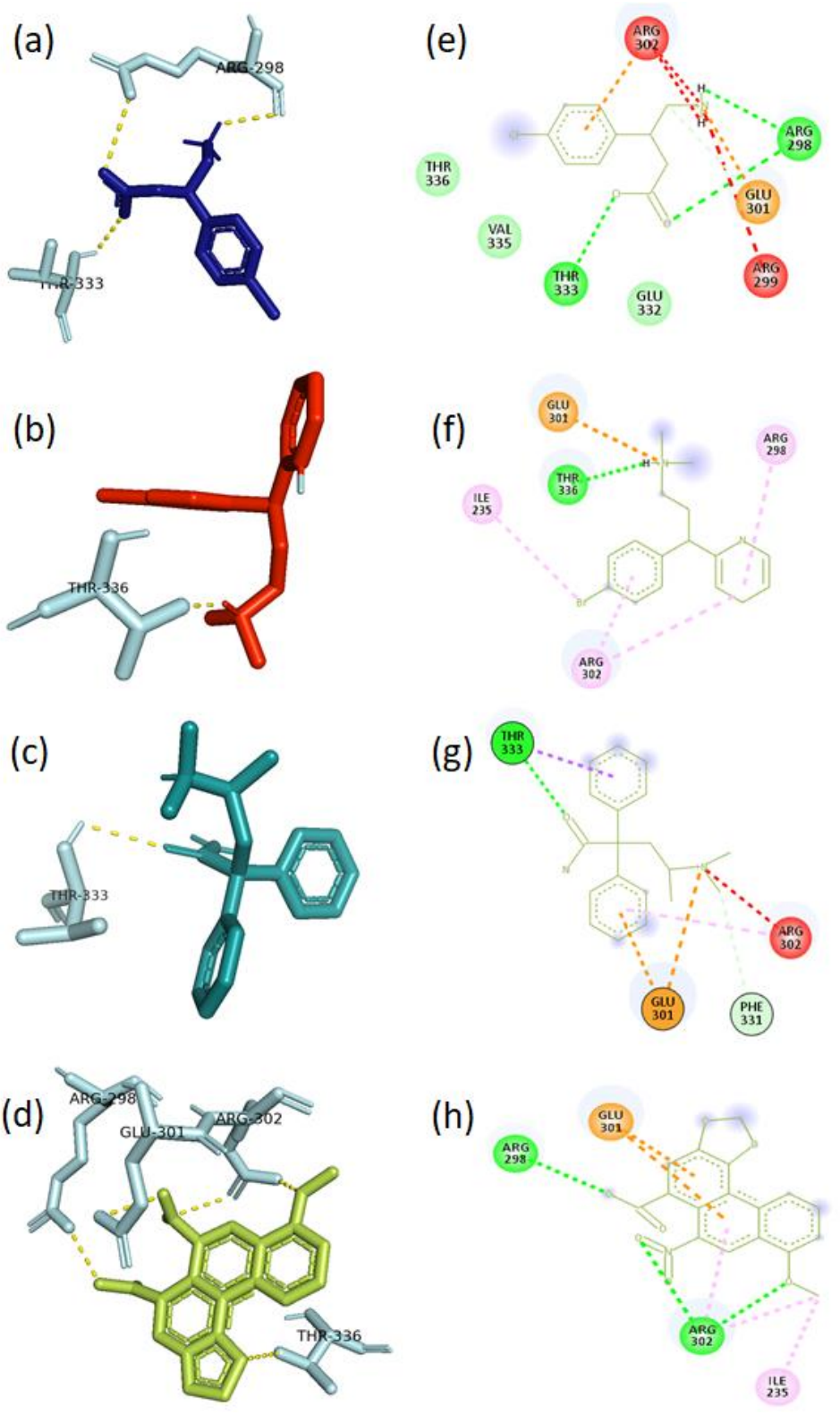


Fig. 5 (a, b, c, d) Ligand interactions with RNR binding pocket residues with ≥-5 kcal/mol; (e, f, g, h) binding modes of the ligands in 2D plots.

ARG-302). Though the binding site is falling 10 Å distant from the active site, it interacts with the active site with pi-pi stacking. CNP0057221 (Lycorine) is prioritized as the best potential inhibitor with a -6.1 kcal/mol docking score forming four hydrogen bonds (GLU-301, ARG-302, THR-333 & THR-336) indicating significant protein-ligand interactions. It has been extracted from the Amaryllidaceae plant family and is a natural alkaloid with antibacterial, antiviral, anti-inflammatory, and anticancer activity.[44] Alternatively, the comparison between the docking scores and binding modes of the Cefotaxime drug (-4.1 kcal/mol) clearly stated the poor binding affinity with RNR. Two natural compounds

(CNP0361821 (-5.3 kcal/mol) and CNP0417500 (-5 kcal/mol)) formed three H-bonds (ARG-298(2), ARG-302; ARG-298(2), THR-333) individually indicating significant interactions along with multiple hydrophobic interactions. The inhibition potential of the natural compounds based on docking studies mentioned above was found to be higher than the Cefotaxime drug.

4. Conclusions

In the current study, a critical annotation of the *C. pneumoniae* genome has been achieved based on genome-wide synteny comparison between the family members. The data set of predicted orthologous genes facilitates the prediction of the drug targets strategically. Non-homologous proteins prediction, essential genes/proteins analysis, subcellular localization prediction, and toxicity prediction were indispensable in identifying and classifying short-listed proteins. Further, prioritization based on druggability by databases search for the available natural compounds was imperative in identification and assigning biological activity. Molecular docking simulation studies revealed the docked natural ligands' higher binding scores compared to the commercially available drug Cefotaxime, indicating the more significant inhibition potential of the ligands identified. The outcome of the current study may provide a framework for the rational identification of drug targets in *C. pneumoniae* through synteny-based patho-genome comparison and novel understandings of interacting molecules with inhibitory potential of RNR, which is evident in developing and validating the novel therapeutic drugs against the RNR target.

Acknowledgments

The authors wish to express a deep sense of gratitude to the management of Manipal Institute of Technology, and Manipal Academy of Higher Education for the resources to carry out the experiments mentioned in this study.

Conflict of interest

There are no conflicts to declare.

Supporting information

Not Applicable.

References

[1] E. J. Strauss, S. Falkow, *Science*, 1997, **276**, 707-712, doi: 10.1126/science.276.5313.707.

[2] E. Donkor, *Genes*, 2013, **4**, 556-572, doi: 10.3390/genes4040556

[3] M. J. Pallen, B. W. Wren, *Nature*, 2007, **449**, 835-842, doi: 10.1038/nature06248.

[4] O. L. Champion, M. W. Gaunt, O. Gundogdu, A. Elmi, A. A. Witney, J. Hinds, N. Dorrell, B. W. Wren, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 16043-16048, doi: 10.1073/pnas.0503252102.

[5] K. Le Doare, J. Bielicki, P. T. Heath, M. Sharland, *Journal of the Pediatric Infectious Diseases Society*, 2015, **4**, 11-20, doi: 10.1093/jpids/piu014.

[6] K. Rajalingam, H. Al-Younes, A. Müller, T. F. Meyer, A. J. Szczepek, T. Rudel, *Infection and Immunity*, 2001, **69**, 7880-7888, doi: 10.1128/iai.69.12.7880-7888.2001.

[7] Y. M. AbdelRahman, R. J. Belland, *FEMS Microbiology Reviews*, 2005, **29**, 949-959, doi: 10.1016/j.femsre.2005.03.002.

[8] C. C. Kuo, L. A. Jackson, L. A. Campbell, J. T. Grayston, *Clinical Microbiology Reviews*, 1995, **8**, 451-461, doi: 10.1128/cmr.8.4.451.

[9] J. T. Grayston, M. B. Aldous, A. Easton, S.-P. Wang, C.-C. Kuo, L. A. Campbell, J. Altman, *Journal of Infectious Diseases*, 1993, **168**, 1231-1235, doi: 10.1093/infdis/168.5.1231.

[10] B. J. Balin, C. S. Little, C. J. Hammond, D. M. Appelt, J. A. Whittum-Hudson, H. C. Gérard, A. P. Hudson, *Journal of Alzheimer's Disease*, 2008, **13**, 371-380, doi: 10.3233/jad-2008-13403

[11] A. J. Littman, L. A. Jackson, T. L. Vaughan, *Biomarkers & Prevention*, 2005, **14**, 773-778, doi: 10.1158/1055-9965.epi-04-0599.

[12] L. Pospisil, J. Canderle, *Veterinární Medicína*, 2012, **49**, 129-134, doi: 10.17221/5686-vetmed.

[13] F. Lemoine, B. Labedan, O. Lespinet, *BMC Bioinformatics*, 2008, **9**, 536, doi: 10.1186/1471-2105-9-536.

[14] E. V. Koonin, *Annual Review of Genetics*, 2005, **39**, 309- 338, doi: 10.1146/annurev.genet.39.073003.114725.

[15] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Research*, 1997, **25**, 3389-3402, doi: 10.1093/nar/25.17.3389.

[16] H. Luo, Y. Lin, F. Gao, C.-T. Zhang, R. Zhang, *Nucleic Acids Research*, 2014, **42**, D574-D580, doi: 10.1093/nar/gkt1131.

[17] C.-S. Yu, C.-W. Cheng, W.-C. Su, K.-C. Chang, S.-W. Huang, J.-K. Hwang, C.-H. Lu, *PLoS One*, 2014, **9**, e99368, doi: 10.1371/journal.pone.0099368.

[18] D. Tsirigos Konstantinos, P. Christoph, S. Nanjiang, K. Lukas, E. Arne, *Nucleic Acids Research*, 2015, **43**, 401-407, doi: 10.1093/nar/gkv485.

[19] A. Krogh, B. Larsson, G. von Heijne, E. L. L. Sonnhammer, *Journal of Molecular Biology*, 2001, **305**, 567-580, doi: 10.1006/jmbi.2000.4315.

[20] J. Shen, J. Zhang, X. Luo, W. Zhu, K. Yu, K. Chen, Y. Li, H. Jiang, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 4337-4341, doi: 10.1073/pnas.0607879104.

[21] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, *Cell*, 2005, **122**, 957-968, doi: 10.1016/j.cell.2005.08.029.

[22] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, C. V. Mering, *Nucleic Acids Research*, 2016, **45**, 362–368, doi: 10.1093/nar/gkw937.

[23] S. Saha, G. P. Raghava, *Immunoinformatics*, 2007, **409**, 387-394, doi: 10.1007/978-1-60327-118-9_29.

[24] A. Mitchell, H. Y. Chang, L. Daugherty, M. Fraser, S. Hunter,

R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, *Nucleic Acids Research*, 2014; **43**, 213-221, doi: 10.1093/nar/gku1243.
[25] P. Imming, C. Sinning, A. Meyer, *Nature Reviews Drug Discovery*, 2006, **5**, 821-834, doi: 10.1038/nrd2132.
[26] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, *Nucleic Acids Research*, 2018, **46**, D1074-D1082, doi: 10.1093/nar/gkx1037.
[27] K. F. Aoki, M. Kanehisa, *Current Protocols in Bioinformatics*, 2005, **11**, 1-12, doi: 10.1002/0471250953.bi0112s11.
[28] M. Sorokina, P. Merseburger, K. Rajan, M. A. Yirik, C. Steinbeck, *Journal of Cheminformatics*, 2021, **13**, 2, doi: 10.1186/s13321-020-00478-9.
[29] A. Gaulton, A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez and A.R. Leach, *Nucleic Acids Research*, 2021, **45**, 945-954, doi: 10.1186/s13321-021-00566-4.
[30] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, *Nature Methods*, 2015, **12**, 7-8, doi: 10.1038/nmeth.3213.
[31] D. Xu, Y. Zhang, *Biophysical Journal*, 2011, **101**, 2525-2534, doi: 10.1016/j.bpj.2011.10.024.
[32] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *Journal of Applied Crystallography*, 1993, **26**, 283-291, doi: 10.1107/s0021889892009944.
[33] O. Dym, D. Eisenberg, T. O. Yeates, *International Tables for Crystallography*, 2012, doi: 10.1107/97809553602060000881.
[34] O. Dym, D. Eisenberg, T. O. Yeates. *Detection of errors in protein models*, 2006, doi: 10.1107/97809553602060000709.
[35] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera? A visualization system for exploratory research and analysis, *Journal of Computational Chemistry*, 2004, **25**, 1605-1612, doi: 10.1002/jcc.20084.
[36] S. Yuan, C.H.C Stephen, and H. Zhenquan, Using PyMOL as a platform for computational drug design, *WIREs Computational Molecular Science*, 2017, **7**, e1298. doi: 10.1002/wcms.1298.
[37] O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *Journal of Computational Chemistry*, 2010, **31**, 455-461, doi: 10.1002/jcc.21334.
[38] W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins, *Nucleic Acids Research*, 2018, **46**, W363-W367, doi: 10.1093/nar/gky473.
[39] P. Csizmadia, MarvinSketch and MarvinView: Molecule Applets for the World Wide Web, *The 3rd International Electronic Conference on Synthetic Organic Chemistry session Information and Compound Archives Management and Internet Application, 1999*, doi: 10.3390/ecsoc-3-01775.
[40] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: an open chemical toolbox, *Journal of Cheminformatics*, 2011, **3**, 1-14, doi: 10.1186/1758-2946-3-33.
[41] O. Rahman, S. P. Cummings, D. J. Harrington, I. C. Sutcliffe, Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria, *World Journal of Microbiology and Biotechnology*, 2008, **24**, 2377-2382, doi: 10.1007/s11274-008-9795-2.
[42] J. D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0., *Journal of Molecular Biology*, 2004 **340**, 783-95. doi: 10.1016/j.jmb.2004.05.028.
[43] S. K. Burley, H. M. Berman, C. Christie, J. M. Duarte, Z. Feng, J. Westbrook, J. Young, C. Zardecki, RCSB Protein Data Bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education, *Protein Science*, 2018, **27**, 316-330, doi: 10.1002/pro.3331.
[44] S. Jahn, B. Seiwert, S. Kretzing, G. Abraham, R. Regenthal, U. Karst, Metabolic studies of the Amaryllidaceous alkaloids galantamine and lycorine based on electrochemical simulation in addition to in vivo and in vitro models, *Analytica Chimica Acta*, 2012, **756**, 60-72, doi: 10.1016/j.aca.2012.10.042.

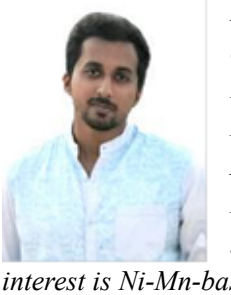
Author Information

Dr. Shivamurthy R. C. *is currently working as Assistant Professor (Selection Grade) in the Department of Mechanical & Manufacturing Engineering at Manipal Institute of Technology, Manipal, India. He is currently the Member of the Department IET Accreditation Committee and Program Chairman and Project Coordinator - Post Graduation Program - Tribology. He has authored various international journals and research conferences.*

Dr. Ravindranath B. S. *is currently working as an Assistant professor in Dept. Of Biotechnology, Manipal Institute of Technology, Manipal. He holds Master's and Ph.D. degrees in Bioinformatics from Kuvempu University, Shimoga. He has fourteen years of professional experience in teaching and eight of research experience. He is the author of multiple peer-reviewed international journals. He is an active member of scientific clubs and groups. He has chaired a jury in multiple scientific conferences and workshops. He is currently involved in research areas including computational data analysis, drug designing, immunoinformatics, and Network Pharmacology.*

Dr. Santhosh K. H. *is currently working as Faculty in Dept. of Biotechnology and Bioinformatics, Kuvempu University, Shimoga. He holds a Master's and a Ph.D. degree in Biotechnology from Kuvempu University, Shimoga. He has six years of research experience as a UGC-Post Doctoral Fellow and research topic entitled on "COX2 dependent and independent pathway evaluation by a novel triterpenoid" in the Department of Biotechnology, Kuvempu University. He has published multiple research articles in peer-reviewed National and International Journals. He has actively participated in National and International level conferences, and also chaired as jury in multiple workshops and faculty development programs related to his area of research.*

Dr. Murthy B. R. N. *BE, MTech Ph.D., is currently working as Associate Professor-Senior Scale, Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. He is the author of several articles in various International Journals. His research includes optimization and simulation of machining process on various types of fiber-reinforced polymer composites through Taguchi, Response Surface Methodology, Artificial Neural Networks, and System Dynamic Methods. His field of research also includes conduction of Austmepring heat treatment to improve the performance of Austempered Ductile Iron Balls which are recognized as the better grinding media for grinding iron ores in ball mills. Currently, he is working on many projects. A few of them are "Characterisation study of aluminum metal matrix composites reinforced with SiC, TiB2, and CNT (Carbon Nano Tubes)", "Development of Hesular alloy for the magnetic refrigeration", "Performance comparative study of hybrid polymer composites reinforced with various fibers and by varying the stacking sequence", etc.*



Mr. Sandeep Nambiar S.*, B.E, MTech, completed his Masters in Manufacturing Engineering and Technology at Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India, has been enrolled as Research Scholar at the same institute. His area of interest is Ni-Mn-based heusler alloys, Al-based metal matrix composite, machining of composites, and its heat treatment.*