



Data Article

Datasets comprising the quality validations of simulated protein-ligand complexes and SYBYL docking scores of bioactive natural compounds as inhibitors of *Mycobacterium tuberculosis* protein-targets



Sravan Kumar Miryala^a, Soumya Basu^a, Aniket Naha^a,
Reetika Debroy^a, Sudha Ramaiah^a, Anand Anbarasu^{a,**},
Saravanan Natarajan^{b,*}

^a Medical and Biological Computing Laboratory, School of Biosciences and Technology, Vellore Institute of Technology (VIT), Vellore 632014, Tamil Nadu, India

^b Department of Biochemistry, ICMR-National Institute for Research in Tuberculosis (NIRT), Chennai 600031, India

ARTICLE INFO

Article history:

Received 31 January 2022

Revised 29 March 2022

Accepted 4 April 2022

Available online 10 April 2022

Dataset link: [Supplementary data related to the quality validations of simulated protein-ligand complexes and SYBYL docking scores of bioactive natural compounds as inhibitors of Mtb protein targets \(Original data\)](#)

ABSTRACT

Docking scores and simulation parameters to study the potency of natural compounds against protein targets in *Mycobacterium tuberculosis* (*Mtb*) were retrieved through molecular docking and *in-silico* structural investigation. The molecular docking datasets comprised 15 natural compounds, seven conventional anti-tuberculosis (anti-TB) drugs and their seven corresponding *Mtb* target proteins. *Mtb* protein targets were actively involved in translation mechanism, nucleic acid metabolism and membrane integrity. Standard structural screening and stereochemical optimizations were adopted to generate the 3D protein structures and their corresponding ligands prior to molecular docking. Force-field integration and energy minimization were further employed to obtain the proteins in their ideal geometry.

DOI of original article: [10.1016/j.molliq.2021.117340](https://doi.org/10.1016/j.molliq.2021.117340)

* Corresponding authors: N. Saravanan, Head – Department of Biochemistry, National Institute for Research in Tuberculosis (NIRT), Indian Council of Medical Research (ICMR), Chetpet, Chennai.

** Corresponding authors: Anand Anbarasu, Professor, Medical and Biological Computing Laboratory, Vellore Institute of Technology (VIT), Vellore, India.

E-mail addresses: aanand@vit.ac.in (A. Anbarasu), saravanan.n@nirt.res.in (S. Natarajan).

<https://doi.org/10.1016/j.dib.2022.108146>

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Keywords:
 Docking
 Simulation
 Natural compounds
 Tuberculosis
 Therapeutics

Surflex-dock algorithm using Hammerhead scoring functions were used to finally produce the docking scores between each protein and the corresponding ligand(s). The best-docked complexes selected for simulation studies were subjected to topology adjustments, charge neutralizations, solvation and equilibrations (temperature, volume and pressure). The protein-ligand complexes and molecular dynamics parameter files have been provided. The trajectories of the simulated parameters such as density, pressure and temperature were generated with integrated tools of the simulation suite. The datasets can be useful to computational and molecular medicine researchers to find therapeutic leads relevant to the chemical behaviours of a specific class of compounds against biological systems. Structural parameters and energy functions provided a set of standard values that can be utilised to design simulation experiments regarding similar macromolecular interactions.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Subject area: Biological Sciences
Specific subject area	Sub-section: Structural biology <i>In silico</i> structural analyses of protein-ligand complexes with molecular dynamics based on chemical interactions
Type of data	Tables and Figures
How the data were acquired	We selected seven conventional drug targets in <i>Mycobacterium tuberculosis</i> whose 3D structures were downloaded from the Protein Data Bank (PDB) database (https://www.rcsb.org/). The structural and functional domains of the proteins were screened from UniProt (https://www.uniprot.org/), InterPro (https://www.ebi.ac.uk/interpro/) and Pfam (http://pfam.xfam.org/) databases. The ligand structures were acquired from the National centre for Biotechnology Information (NCBI) PubChem compound database (https://pubchem.ncbi.nlm.nih.gov/) and Drug bank database (https://www.drugbank.com/). The absence of a ligand structure in drug repositories was compensated by drawing the same with the ChemSketch tool followed by 3D structures generation using OpenBabel online server (http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html). The bond integrities were validated using the Avogadro tool. The datasets comprising molecular docking scores of natural compounds (ligands) and the classical anti-Tuberculosis drugs with their respective targets were generated using the SYBYL-Surflex-docking tool kit. The best-docked complexes were subjected to topology adjustments for individual proteins and ligands using CHARMM36-Mar2019 force-field-TIP3P water-model and CGenFF (https://cgenff.umaryland.edu/) online server with default parameters respectively. The integrated simulation suite GROMACS 2018.1 was utilized. The optimized macromolecular complexes were solvated by centering in an aqueous dodecahedron box of uniform edge distance of 1.0 nm. Subsequently, requisite counter ions (Na ⁺ or Cl ⁻) were added to balance the charges of the solvated system. Energy was minimized using integrated steepest descent algorithm for 50,000 steps and convergence-tolerance of 1000 kJ/mol nm ⁻¹ . System equilibration with standard NVT (constant Number of particles, Volume and Temperature) and NPT (constant Number of particles, Pressure and Temperature) ensembles were performed for

(continued on next page)

Data format	100 ps. A constant pressure of 0 (zero) bar and temperature of 300 K with uniform density of $\sim 1040 \text{ kg/m}^3$ was set for parameterization. Final molecular dynamics simulation (MDS) was carried out for 75 ns. Grace software was employed to visualize the trajectories of simulation parameters. A chronological list of commands and other associated parameter files to run simulation along with the entire MD-simulation files have been provided in the associated Mendeley dataset folder as mentioned in subsequent sections.
Description of data collection	Data is in raw and analysed form. The structural chemistry data was acquired from authorised databases and repositories, followed by necessary optimisations using licensed (academic and professional) software. The reported docking scores and simulation parameters are based on universally accepted terms/standards.
Data source location	<ul style="list-style-type: none"> • Institution: Vellore Institute of Technology, Vellore • City/Town/Region: Vellore, Tamil Nadu • Country: India
Data accessibility	Data is available within this article and the raw data files in excel format and other standard formats for simulation has been uploaded on public repository and datasets with active link below is provided as the supplementary data. Link: (https://data.mendeley.com/datasets/94rh86jfpk/3), DOI: 10.17632/94rh86jfpk.3
Related research article	The presented dataset is associated with our recent publication mentioned below[1]: S. K. Miryala, S. Basu, A. Naha, R. Debroy, S. Ramaiah, A. Anbarasu & S. Natarajan (2021). Identification of bioactive natural compounds as efficient inhibitors against Mycobacterium tuberculosis protein-targets: A molecular docking and molecular dynamics simulation study. <i>Journal of Molecular Liquids</i> , 341, 117,340. https://doi.org/10.1016/j.molliq.2021.117340

Value of the Data

- 1) There are four distinct types of datasets presented in this manuscript:
 - a) **Raw docking scores** like Crash score, G-score, PMF score, D-score, Chem scores and C-scores can help understand different chemical factors affecting ligand-protein binding.
 - b) **Optimized protein-ligand complexes used for simulation** will provide comprehensive idea about the fundamental format of input biomolecular structural complexes to run molecular dynamics simulations.
 - c) **The optimized molecular dynamics parameter files, output files and list of commands** will definitely facilitate in further analysis and performing essential dynamics studies besides guiding researchers to replicate similar experimental approaches and objectives.
 - d) **Trajectories of optimised conditions for simulation** regarding individual protein-ligand complexes can give a fair idea of the set of conditions required to simulate a specific type of biomolecule (protein) interacting with a certain class of compounds.
- 2) The datasets can be of interest to bioinformaticians, computational biologists, phytochemists and molecular medicine researchers, who can figure out leads relevant to the chemical behaviours of a certain class of compounds against biological systems.
- 3) The docking scores can further be exploited either based on individual compounds or collective understanding of a specific class of compounds or analysis of specific chemical parameters based on individual scoring algorithms.
- 4) The compounds that were not considered as per criteria presented in the main publication can further be explored similarly against other potent targets [1].
- 5) The optimised simulation parameters can readily guide researchers by providing a set of standard values that can be utilised to design simulation experiments regarding the same/similar macromolecules
- 6) The simulation profiles may encourage designing of efficient therapeutic agents by providing crucial interaction dynamics values.

1. Data Description

The presented datasets depict the feasibility of certain classes of natural compounds as therapeutic candidates against *Mtb* protein targets. **Supplementary Files-1–7 (Docking scores)** (<https://data.mendeley.com/datasets/94rh86jfpk/3>) portrayed the docking scores comprising crash score, G-score, PMF score, D-score, chem scores, polar, total score, consensus (C) score, number of Hydrogen-bonds of the natural compounds against *Mtb* targets [Arabinosyl transferase (PDB ID: 3PTY); DNA Gyrase subunit A (PDB ID: 4G3N); Ribosomal protein S1 (PDB ID: 4NNI); 2'-O-Methyltransferase (PDB ID: 5KYG); Enoyl (acyl-carrier protein) reductase (PDB ID: 5VRL); F-ATP synthase epsilon chain (PDB ID: 5YIO) and RNA polymerase subunit C (PDB ID: 5ZX3)] as compared to respective classical drugs (Ethambutol, Levofloxacin, Pyrazinamide, Capreomycin, Isoniazid, Bedaquiline and Rifampicin). The optimized protein-ligand complexes used as input files for parameterization and simulation has been provided as “**MDS_input_files**” (<https://data.mendeley.com/datasets/94rh86jfpk/3>). The molecular dynamics parameter files along with the set of commands to run MDS are available as “**MDS_parameter_files**” (<https://data.mendeley.com/datasets/94rh86jfpk/3>). **Fig. 1-3** represented the quality-check parameters after equilibrating certain protein-ligand complexes comprising classical and natural compounds prior to MD run depicting the electron Density, Pressure and Temperature levels [1]. The differences in molecular weight and number of atoms were reflected upon the electron density function of the individual protein-ligand complexes. The datasets for generating the figures has been provided explicitly in the **Supplementary Files-8–14 (Figure_datasets)** (<https://data.mendeley.com/datasets/94rh86jfpk/3>). The entire simulation dataset has been segmented appropriately based on the PDB IDs of the studied protein-targets. The different input and output files generated are made available under “**MD_simulation_files**” (<https://data.mendeley.com/datasets/94rh86jfpk/3>).

2. Experimental Design, Materials and Methods

Seven *Mtb* proteins were selected, which are already targets of conventional anti-TB drugs [2]. Their 3D structures were obtained from RCSB-PDB (<https://www.rcsb.org/>), while the functional domains/motifs were obtained from InterPro (<https://www.ebi.ac.uk/interpro/>), Pfam (<http://pfam.xfam.org/>), and UniProt (<https://www.uniprot.org/>) databases. The classical drugs and the natural compounds were retrieved from the DrugBank (<https://www.drugbank.com/>) and PubChem Compound (<https://pubchem.ncbi.nlm.nih.gov/>) databases. ChemSketch tool [3] was employed in the absence of ligand structures for 2D structure construction followed by generation of 3D coordinates using the OpenBabel Chemical File Format Converter (<http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html>). Further, the ligands were optimised with the Avogadro tool [4]. Molecular docking between the conventional anti-TB drugs and natural compounds with their respective targets was performed using the SYBYL-Surflex-docking tool kit (Tripos International, USA). The protein structures were refined to remove bound ligands and water molecules, fixing side chains, adding hydrogen atoms, followed by atomic-level charge designation using AMBER7 F99 force field. Thereafter, the proteins were energy minimised by Powell's method with Tripos force field followed by Protomol generation. Hammerhead functional scorings determined the polar, crash, entropic, hydrophobic and repulsive properties to yield the docked score datasets [5,6]. The MDS analyses for 75 nanoseconds (ns) was performed for each of the best-docked complexes with GROMACS 2018.1 suite [7]. Protein topologies were generated using CHARMM36-Mar2019 force-field mechanics and TIP3P model (for water cluster), while ligand topologies were built using CGenFF (<https://cgenff.umaryland.edu/>) online server with default parameters. The protein structures were placed within the center of the dodecahedron box of uniform edge distance of 1.0 nm, followed by solvation and addition of requisite counter ions (Na^+ or Cl^-) to the system. Steepest descent algorithm for 50,000 steps and convergence-tolerance of $1000 \text{ kJ/mol nm}^{-1}$ were utilised for energy minimisation following system equilibration under standard NVT (constant Number of particles, Volume and

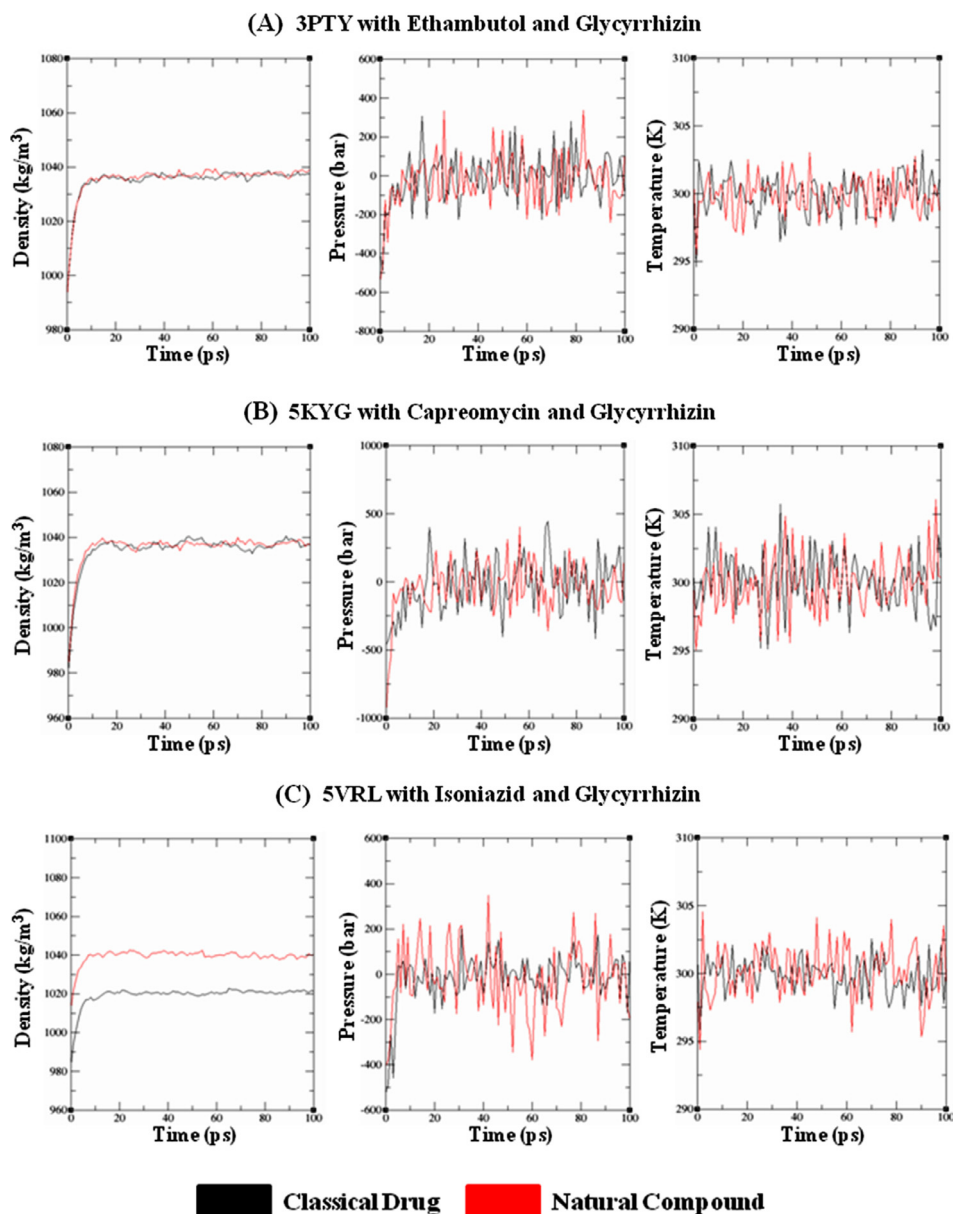


Fig. 1. Quality check parameters after equilibrating the respective protein-ligand complexes prior to MD run depicting the Density gradients, Pressure and Temperature levels. (A) 3PTY with Ethambutol and Glycyrrhizin, (B) 5KYG with Capreomycin and Glycyrrhizin, (C) 5VRL with Isoniazid and Glycyrrhizin.

Temperature) and NPT (constant Number of particles, Pressure and Temperature) ensembles for 100 ps [7–14]. The trajectories of simulation parameters were visualised using Grace software (<https://plasma-gate.weizmann.ac.il/Grace/>).

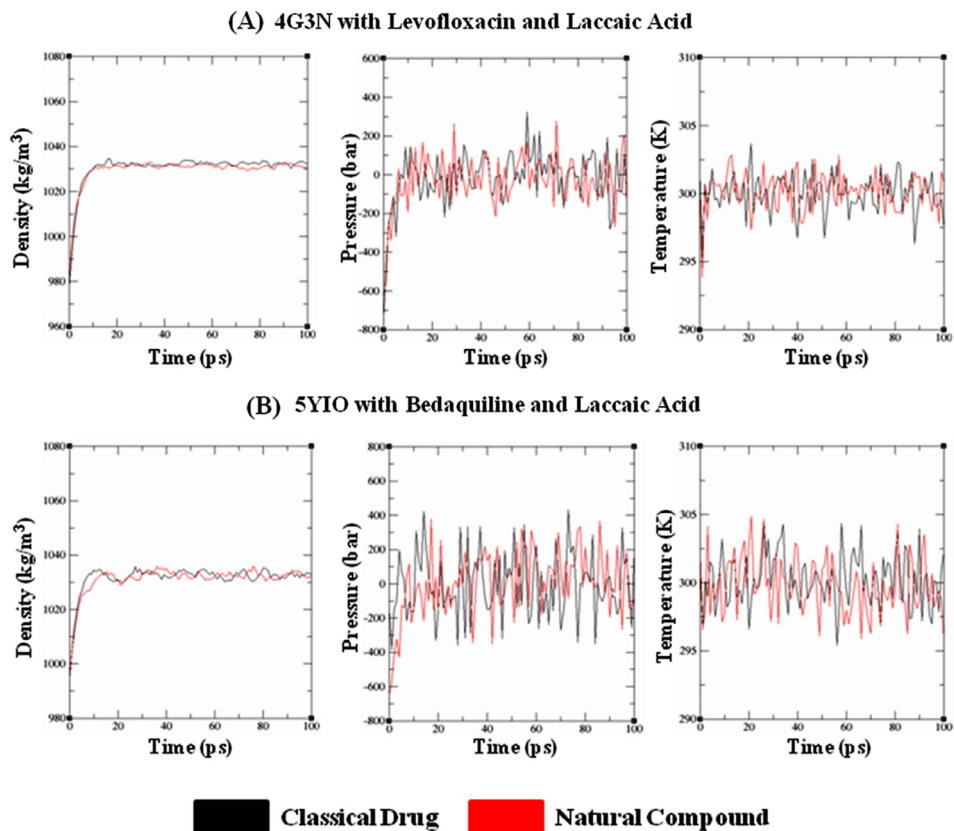


Fig. 2. Quality check parameters after equilibrating the respective protein-ligand complexes prior to MD run depicting the Density gradients, Pressure and Temperature levels. (A) 4G3N with Levofloxacin and Laccac Acid, (B) 5YIO with Bedaquiline and Laccac Acid.

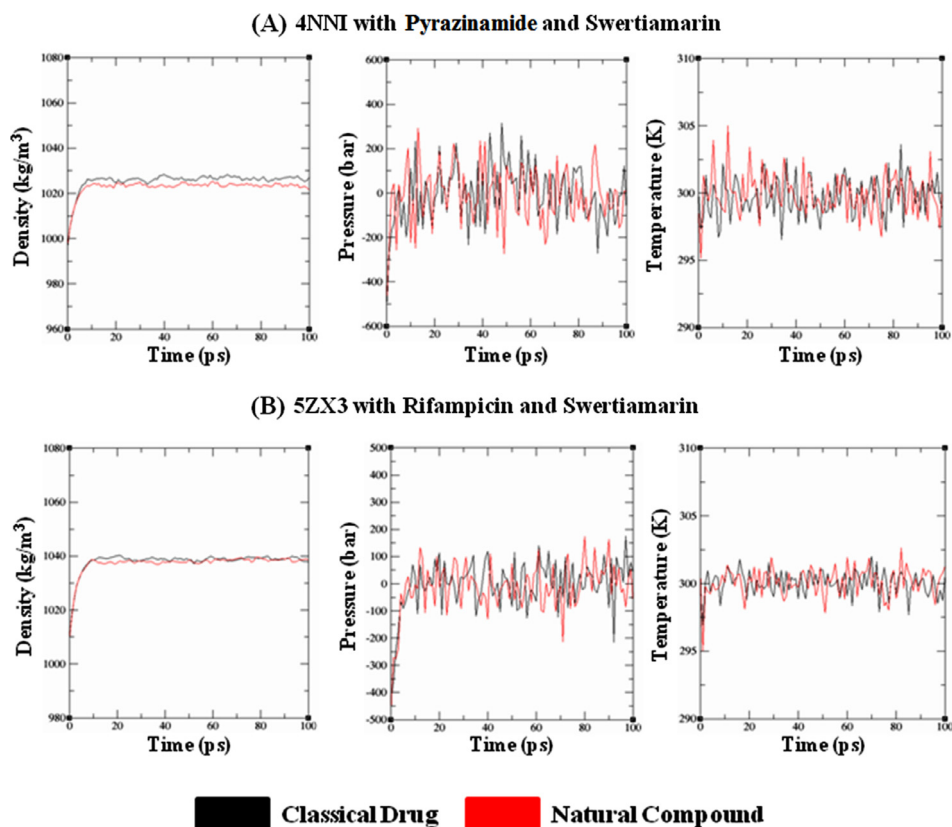


Fig. 3. Quality check parameters after equilibrating the respective protein-ligand complexes prior to MD run depicting the Density gradients, Pressure and Temperature levels. **(A)** 4NNI with Pyrazinamide and Swertiamarin, **(B)** 5ZX3 with Rifampicin and Swertiamarin.

Ethics Statements

The work did not involve any human subjects, animal experiments and data from social media platforms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Supplementary data related to the quality validations of simulated protein-ligand complexes and SYBYL docking scores of bioactive natural compounds as inhibitors of Mtb protein targets (Original data) (Mendeley Data).

CRediT Author Statement

Sravan Kumar Miryala: Data curation, Formal analysis, Visualization, Writing – original draft; **Soumya Basu:** Formal analysis, Visualization, Writing – original draft; **Aniket Naha:** Formal analysis, Visualization, Writing – original draft; **Reetika Debroy:** Formal analysis, Visualization, Writing – original draft; **Sudha Ramaiah:** Conceptualization, Methodology, Validation, Writing – review & editing; **Anand Anbarasu:** Funding acquisition, Conceptualization, Project administration, Supervision; **Saravanan Natarajan:** Funding acquisition, Conceptualization, Project administration, Supervision.

Acknowledgments

AA, SR, MSK, SB, AN, and RD would like to thank the management of VIT for providing the necessary facilities to carry out this research work. SN would like to thank the Director, ICMR-NIRT for providing the necessary supports to carry out this research work.

The authors gratefully acknowledge the Indian Council of Medical Research (ICMR), New Delhi, Government of India for the research grant [IRIS ID: 2020-0690] and ICMR-NIRT, Chennai, for the support in meeting the article publication charges. SB and AN thank ICMR for their research fellowships.

References

- [1] S.K. Miryala, S. Basu, A. Naha, R. Debroy, S. Ramaiah, A. Anbarasu, S. Natarajan, Identification of bioactive natural compounds as efficient inhibitors against mycobacterium tuberculosis protein-targets: a molecular docking and molecular dynamics simulation study, *J. Mol. Liq.* 341 (2021) 117340, doi:10.1016/j.molliq.2021.117340.
- [2] C.U. Ibeji, N.A.M. Salleh, J.S. Sum, A.C.W. Ch'ng, T.S. Lim, Y.S. Choong, Demystifying the catalytic pathway of mycobacterium tuberculosis isocitrate lyase, *Sci. Rep.* 10 (2020) 18925, doi:10.1038/s41598-020-75799-8.
- [3] Z. Li, H. Wan, Y. Shi, P. Ouyang, Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1886–1890.
- [4] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminform.* 4 (2012) 17, doi:10.1186/1758-2946-4-17.
- [5] K. Malathi, S. Ramaiah, Molecular docking and molecular dynamics studies to identify potential OXA-10 extended spectrum β -Lactamase non-hydrolysing inhibitors for pseudomonas aeruginosa, *Cell Biochem. Biophys.* 74 (2016) 141–155, doi:10.1007/s12013-016-0735-8.
- [6] A. Naha, S. Vijayakumar, B. Lal, B.A. Shankar, Genome sequencing and molecular characterisation of XDR acinetobacter baumannii reveal complexities in resistance : novel combination of Sulbactam-Durlobactam holds promise for therapeutic intervention, *J. Cell. Biochem.* (2021) 1–25, doi:10.1002/jcb.30156.
- [7] J. Lemkul, From proteins to perturbed hamiltonians: a suite of tutorials for the GROMACS-2018 molecular simulation package [article v1.0], *Living J. Comput. Mol. Sci.* 1 (2019) 1–53, doi:10.33011/livecoms.1.1.5068.
- [8] S. Basu, A. Naha, B. Veeraraghavan, S. Ramaiah, A. Anbarasu, In silico structure evaluation of BAG3 and elucidating its association with bacterial infections through protein-protein and host-pathogen interaction analysis, *J. Cell. Biochem.* (2021), doi:10.1002/jcb.29953.
- [9] M. Jayaraman, S.K. Rajendra, K. Ramadas, Structural insight into conformational dynamics of non-active site mutations in KasA: a mycobacterium tuberculosis target protein, *Gene* 720 (2019) 144082, doi:10.1016/j.gene.2019.144082.
- [10] S. Basu, B. Veeraraghavan, S. Ramaiah, A. Anbarasu, Novel cyclohexanone compound as a potential ligand against SARS-CoV-2 main-protease, *Microb. Pathog.* 149 (2020) 104546, doi:10.1016/j.micpath.2020.104546.
- [11] K. Vasudevan, S. Basu, A. Arumugam, A. Naha, S. Ramaiah, A. Anbarasu, B. Veeraraghavan, Identification of potential carboxylic acid-containing drug candidate to design novel competitive NDM inhibitors: an in-silico approach comprising combined virtual screening and molecular dynamics simulation, *Res. Prepr.* (2021), doi:10.21203/rs.3.rs-784343/v1.
- [12] M. Thillainayagam, K. Malathi, S. Ramaiah, In-Silico molecular docking and simulation studies on novel chalcone and flavone hybrid derivatives with 1, 2, 3-triazole linkage as vital inhibitors of plasmodium falciparum dihydroorotate dehydrogenase, *J. Biomol. Struct. Dyn.* 36 (2018) 3993–4009, doi:10.1080/07391102.2017.1404935.
- [13] M. Thillainayagam, S. Ramaiah, A. Anbarasu, Molecular docking and dynamics studies on novel benzene sulfonamide substituted pyrazole-pyrazoline analogues as potent inhibitors of plasmodium falciparum histidine aspartic protease, *J. Biomol. Struct. Dyn.* 38 (2020) 3235–3245, doi:10.1080/07391102.2019.1654923.
- [14] D.E. Elmore, D.A. Dougherty, Molecular dynamics simulations of wild-type and mutant forms of the mycobacterium tuberculosis Mscl channel, *Biophys. J.* 81 (2001) 1345–1359, doi:10.1016/S0006-3495(01)75791-8.