Indian Journal of Biochemistry & Biophysics
Vol. 62, December 2025, pp. 1351-1368
DOI: 10.56042/ijbb.v62i12.16040

Structure-based inhibition of *Rhizoctonia solani* in rice sheath blight Beta-glucosidase by natural terpenoids

Balakrishnan M[1]*, Vyshnavi M[1], Supriya P[1], Dhandapani A[1], Satendra Kumar Mangrauthia[2] & Srinivasa Rao CH[1]

[1]ICAR-National Academy of Agricultural Research Management, Hyderabad-500 030, Telangana, India

[2]ICAR- Indian Institute of Rice Research, Hyderabad-500 030, Telangana, India

Received 27 December 2024; *revised 13 March 2025*

Rice, a primary staple crop, is highly susceptible to fungal pathogens, with *Rhizoctonia solani* induced sheath blight being one of the most economically significant diseases, leading to substantial yield losses. The widespread leaning on synthetic fungicides has resulted in resistance development, environmental toxicity, and non-target effects, necessitating alternative eco-compatible disease management strategies. Natural terpenoids, with their broad-spectrum antifungal properties, present a promising biocontrol approach against *R. solani*. In this study, an *in silico* structure-based drug design approach was employed to target Beta-glucosidase, a key cell-wall degrading enzyme in *R. solani*. Due to the unavailability of its experimentally resolved 3D structure (Uniprot ID: A0A8H7H8J1), homology modelling was performed using SWISS-MODEL, MODELLER, MultiFOLD, and AlphaFold2. Structural validation through the SAVES v6.0 server indicated that the MODELLER-derived structure exhibited the highest structural fidelity. A molecular docking study involving 15 major terpenoid constituents of essential oils was conducted, wherein Beta-sitosterol, abietic acid, and palustric acid demonstrated the strongest binding affinities with Beta-glucosidase. Pharmacokinetic evaluation through ADME screening identified palustric acid, a diterpene, as the most stable compound with an optimal partition coefficient. However, ADMET toxicity profiling suggested potential hepatotoxic effects at elevated concentrations, while maintaining an overall favourable environmental safety profile. Given its efficacy and biocompatibility, palustric acid, in conjunction with biocontrol agents and improved agronomic practices, holds potential as an alternative strategy for mitigating *R. solani*-mediated sheath blight in rice.

Keywords: Anti-fungal, Biocontrol agents, Food security, Palustric acid, Rice sheath blight, Sustainability

Rice (*Oryza sativa* L.) serves as a fundamental food source for nearly half of the global population. Notably, China and India are the leading producers, together contributing approximately 50% of the total rice output[1]. With over 43 million hectares of land under cultivation, rice is the most significant and widely produced food crop in India. However, productivity of India (2.22 tonnes/ha) is lower than the world average yields (4.05 t/ha) and is much behind than the rice productivity of Egypt, Japan, China, Vietnam, USA and Indonesia[2,3]. In addition to its vital role in guaranteeing national food security, rice is regarded as a strategic commodity in the majority of Asian nations. Maintaining rice production self-sufficiency and ensuring price stability are still seen by most Asian countries as essential political objectives[4]. Most major rice-producing nations have witnessed yield and total output plateau for many years, and numerous barriers stand in the way of ongoing production expansion in order to meet growing demand brought on by an ever-increasing population[5].Climate change significantly threatens global food security, especially for monsoon-dependent agricultural systems common among the Indian diaspora. These systems are highly vulnerable to climate variability, which can disrupt food production and increase the risk of food insecurity[6]. The rising quantity of greenhouse gases in the atmosphere is expected to cause a 1.4–5.8°C increase in the global mean air temperature between 1990 and 2100. Scientists discovered that for every degree Celsius that the minimum temperature increased, the production of rice decreased by 10%[7,8]. The sustainability of agriculture has been put at risk Long-term intensive rice cultivation using traditional methodshas been linked to significant destruction of natural resources, declined factor productivity, multiple nutritional deficiencies, depletion of groundwater, lack of labor, and higher cultivation costs[9]. The System of Rice Intensification (SRI) is the use of specific management techniques, such as

*Correspondence:
E-mail: mbkrishnan72@gmail.com

adjustments to nursery management, transplanting schedules, watering, and weed control, that improve the root zone conditions of rice plants in comparison to those grown using traditional techniques designed to boost rice yields with fewer inputs[10]. Smart farming besides fungicides, palustric acid with biocontrol agents and better management practices can help combat disease spreada way forward to combat the disease spreadcan help the rice sector achieve resilient and sustainable production systems that could reduce environmental impact and ensure food security[11].

A significant necrotrophic fungal pathogen that causes sheath blight disease in rice is *Rhizoctonia solani* AG1-1A (Anastomosis Group 1 Subgroup 1A), and managing this disease is still difficult for sustainable rice farming. Sheath blight infection dramatically reduced rice yield; the rice field with the highest inoculum density showed a 40% yield decline[12,13]. The damages caused due to the infection such as blightening of the sheath at the water-level and poorly filled grains are depicted (Figs 1 & 2). The intensification of the rice-cropping system, which is marked by the creation of new, short-statured, high-tillering, high-yielding varieties and an increase in nitrogen fertilization, has been linked to the emergence of *R. solani* as a major rice pathogen and this has created a favourable microclimate for the pathogen[14]. Essential pathogenicity factors have been identified as a variety of secondary metabolites, carbohydrate-active enzymes, secreted proteins, and effectors in *R. solani*[15]. The development of sheath blight disease in rice is significantly influenced by the phytotoxins that the fungus produces. It was initially determined using a controlled artificial inoculation experiment that the crude toxin contributed to the pathogenicity and virulence of the fungal pathogen on rice[16]. Furthermore, due to its extensive host range and outstanding genetic variability, the disease can be challenging to control because no adequate degree of natural resistance has been found in the rice germplasm which is wholly accessible[17,18]. The application of agronomic techniques, chemical control, biological control, and genetic enhancement have all been addressed in research on disease management strategies[19]. Fungi possess a diverse array of secreted enzymes called cell wall–degrading enzymes (CWDEs) to break down the primary structural polysaccharide components of plant cell walls, such as pectin, cellulose, and hemicellulose[19]. An elevated abundance of several metabolites was identified through metabolomics investigation that induce plant immune responses and trigger signalling pathways. Hence, degradation of cell wall is the entry point for the fungus to gain control over the host's immune system[20].

Fig. 1 — Water-soaked parts develop a greenish-grey colour patches

Fig. 2 — Poorly filled grains and flag leaves that appear weak and brittle

Terpenes have the potential to be useful biocides; research indicates that they work by preventing the formation of ergo sterols, increasing lipid peroxidation, and possibly blocking enzymes. While the antifungal efficacy varies based on the specific terpene compound and fungal species, the studies demonstrate that natural terpenes exhibit favourable potential as biocides to combat oomycetes and harmful fungus, with effectiveness comparable or superior to traditional synthetic fungicides in many

cases[21]. In this study, the physicochemical properties of beta-glucosidase and molecular docking was carried out using 15 terpenic derivatives, diterpenes and triterpenes notable for their anti-fungal capacity. The main aim is to screen for a natural ligand or a combination of natural ligands that are good inhibitors of the fungal enzyme and to aid in plant defence responses.

Materials and methods

Physicochemical parameters and secondary structure prediction

Beta-glucosidase's amino acid sequence for *R. solani* was obtained from Uniprot[22,23] (https://www.uniprot.org/). A0A8H7H8J1 was the entry ID, and the protein known as beta-glucosidase was selected as the preferred protein for more examination. This amino acid sequence was used to estimate the following physicochemical parameters: molecular weight, aliphatic index (AI), theoretical isoelectric point (pI), total number of (+R) positive and (-R) negative residues, and grand average hydropathy (GRAVY). We utilize the Prot Param web server tool[24] (https://web.expasy.org/protparam/) from Expasy to calculate these parameters. Following that, the protein's secondary structure was examined using PROTEUS Structure Prediction Server 2.0[25] (http://www.proteus2.ca/proteus2/), which facilitates thorough protein structure prediction and structure-based annotation.

Molecular modelling of Beta-glucosidase

After finding a confident match to the protein of interest, homology modelling was performed using SWISS-MODEL[26], MODELLER[27], MultiFOLD[28] and AlphaFold2[29,30].

SWISS-MODEL

The FASTA sequence of beta-glucosidase was submitted to SWISS-MODEL (https://swissmodel.expasy.org) web server to develop a model with sufficient query sequence coverage and sequence identity. A 3D structure of the protein is generated on the basis of selection of a template with highest sequence identity.

MODELLER

For this aspect the bioinformatics toolkit[31] (https://toolkit.tuebingen.mpg.de/) developed by the Max Planck Institute for Biology which offers access to a great variety of public and in-house bioinformatics tools from protein sequence analysis to structure prediction through modeller. The beta-glucosidase FASTA sequence was input using HHpred, the first pairwise comparison of profile hidden Markov models (HMMs) and a fast server for remote protein homology detection and structure prediction. A large variety of databases, including the PDB, SCOP, Pfam, SMART, COGs, and CDD, can be searched using it. It takes a single query sequence or a multiple alignment as input and outputs the search results in a format that is easy to use, much like PSI-BLAST. The top-scoring templates were selected to 'Create a model using selection' which serves as an input to the modeller which is embedded within the toolkit.

MultiFOLD

MultiFOLD (https://www.reading.ac.uk/bioinf/MultiFOLD/), an automated prediction, quality assessment, and refining tool for tertiary and quaternary protein structure models, is utilized for protein sequences longer than 1000 amino acids. For homology modelling, the target protein was input into the server as a sample sequence in single-letter format. Five distinct 3D models were produced as a result, and they were arranged in decreasing order of global model quality scores. The global model quality scores ranged from 0 to 1.

AlphaFold2

DeepMind's AlphaFold2 (AF2) artificial intelligence (AI) system is capable of accurately predicting three-dimensional (3D) protein structures from amino acid sequences at the atomic level. By using multi-sequence alignments, it integrates biological and physical understanding about protein structure into the deep learning algorithm's design[29,30].A simplified online implementation of this breakthrough AI system called ColabFold v1.5.5[32] is available through Google Colabroratory (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) which provides GPU's free of cost to run computationally-intensive analysis. All operations take place in the cloud, so you can use the software without installing it on your own computer. Insert the amino acid sequence in the designated area and assign a 'jobname'. For the parameters, template mode was selected to 'none' to aid in *denovo* protein searchMMseqs2 was utilized for MSA since it has a speed that is several orders of magnitude quicker than BLAST while maintaining

sensitivity. It has a speed advantage of up to 400 times faster than PSI-BLAST, yet it can still execute profile searches. For monomer prediction, 'alphafold2_ptm' was chosen as the model type. The predicted protein structure will be downloaded once all cells are executed.

Protein model validation

The quality of models generated through modeller, SWISS-MODEL server, MultiFOLD and AlphaFold2 was validated. Using a Ramachandran plot, the models' stereo chemical investigation and assessment of their structures were carried out. The model that had the most residues in the preferred area and the fewest in the outlier area was chosen for additional examination. It was generated using SAVESv6.0 server along with that ERRAT[33], VERIFY3D[34], PROCHECK[35] tools were used to analyse the overall quality of the model. The Ramachandran plot is utilized to evaluate the quality of a modelled protein or an experimental structure, while the statistics from the plot offer insights into the total count of amino acid residues located in the favourable, allowed, and disallowed regions. Finally, one of the models generated by MODELLER was selected for subsequent studies.

Finding binding pocket and conserved domain

Binding sites on protein surfaces are crucial for their function. In this study, we utilized PrankWeb[36] (https://prankweb.cz/), a web server designed for predicting and visualizing ligand binding pockets. To identify conserved domains and motifs in beta-glucosidase, we employed additional resources for analysis such as the ScanProsite[37] part of ExPASy, (https://prosite.expasy.org/) is a tool that aids in identifying and annotating conserved regions within protein sequences, including protein families, domains, and motifs. It offers a web interface that allows users to find protein matches based on signatures from the PROSITE database.

Selection of ligands and screening

This study involved 15 terpenic derivatives, which are key components of essential oils, including carvacrol, 1,8-cineole, citral, eugenol, farnesol, geraniol, linalool, menthol, α-terpinene, and thymol, as well as diterpenes like abietic acid, palustric acid, and neoabietic acid, along with triterpenes such as beta-sitosterol and ursolic acid, all derived from natural sources[21]. These ligands were downloaded from the LOTUS database[38](https://lotus.naturalproducts.net/) which is an online resource for natural products.

Molecular docking of beta-glucosidase

Modelled 3D structure of Beta-glucosidase protein used for energy minimization by using UCSF Chimera[39] and prepared for docking.

This involves the following actions: removal of non-standard amino acids, addition of polar hydrogens and Kollmann charges. All these actions can be automated by using the 'Dock prep' option within the 'Actions' menu with which the protein is ready to be docked. For the docking procedure, the PyRx[40]was used which is a molecular docking and virtual screening software.

Analysis and Visualization

The interaction between the ligand and protein was examined and visualized using PyMOL[41] and LigPlot+[42]. PyMOL is a molecular visualization tool that creates high-quality 3D representations of small molecules and biological macromolecules, including proteins. LigPlot+ is a software that automatically generates 2D diagrams illustrating ligand-protein interactions.

ADMET screening and analysis

ADMET screening was conducted on the 15 selected ligands to assess their partition coefficients (Log D7.4, Log P), water solubility (Log S), and Synthetic Accessibility (GASA)[43]. This evaluation is crucial for empirically designed or visually screened molecules prior to chemical synthesis and biochemical testing. Similarly, ADMET analysis was performed to evaluate the pharmacokinetic parameters of drug-like compounds based on their molecular structure, focusing on Absorption, Distribution, Metabolism, Excretion, and Toxicity. This analysis highlights several important considerations that must be addressed even after identifying one or more bioactive lead compounds. For both assessments, the online server ADMETlab 3.038 (https://admetlab3.scbdd.com/) was utilized.

Results and discussion

Calculation of Physicochemical and Secondary Structure Parameters

The amino acid sequence of Beta-glucosidase from *Rhizoctonia solani* was obtained from UniProt. The selected sequence, identified by the entry ID A0A8H7H8J1, was used for analysing physicochemical

Table 1 — Quality assessment of the 3D structures generated through various software's with the help of SAVES server 6.0

S. No	Software used	ERRAT (Overall quality score)	VERIFY3D (3D-1D score)	PROCHECK (Ramachandran Plot)			
				Most Favoured	Allowed	General	Disallowed
1.	SWISS-MODEL	89.6104	88.97%	88%	10%	1.3%	0.7%
2.	MODELLER	71.8841	89.11%	89.7%	7.5%	1.9%	1.0%
3.	MultiFOLD (Model-I)	89.916	79.96%	84.7%	13.4%	1%	0.8%
4.	MultiFOLD (Model-II)	90.7117	80.83%	88.2%	10.5%	0.5%	0.8%
5.	MultiFOLD (Model-III)	86.6747	82.37%	77.5%	12.9%	6.2%	3.4%
6.	MultiFOLD (Model-IV)	86.5868	81.38%	76.1%	15.3%	5.5%	3.1%
7.	MultiFOLD (Model-V)	88.5127	78.42%	76.2%	15.1%	6.0%	2.7%
8.	AlphaFold2 (Model-I)	55.0685	60.13%	57.5%	23.6%	12.7%	6.2%
9.	AlphaFold2 (Model-II)	52.0408	51.26%	52.4%	25.0%	12.6%	10.1%
10.	AlphaFold2 (Model-III)	27.3092	42.38%	42.1%	26.9%	17.5%	13.5%
11.	AlphaFold2 (Model-IV)	40.64	51.81%	50.8%	23.8%	14.1%	11.3%
12.	AlphaFold2 (Model-V)	Failed	61.34%	37.0%	26.7%	20.0%	16.3%

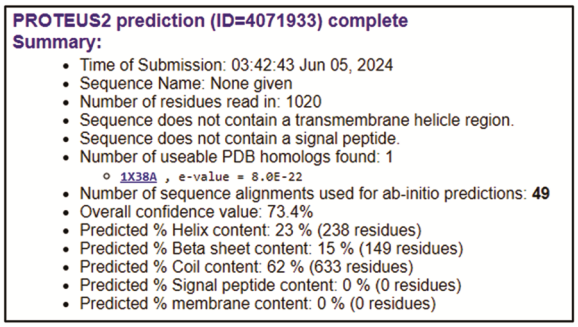


Fig. 3 — Summary of PROTEUS2 prediction of Beta-glucosidase

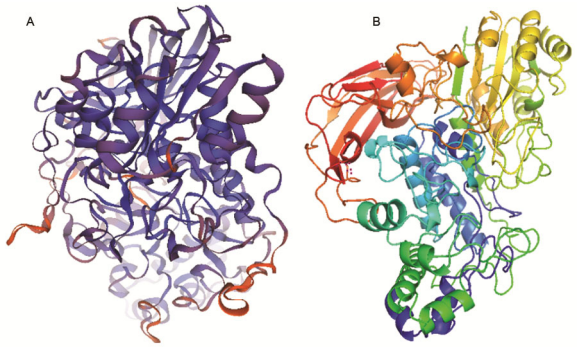


Fig. 4 — (A) 3D structure of protein generated by SWISS-MODEL; and (B) 3D structure of protein generated by MODELLER

properties and secondary structural features. This protein sequence consists of 1020 amino acids with an average molecular weight of 108,316.32 Daltons. It contains 79 negatively charged residues (aspartic acid and glutamic acid) and 78 positively charged residues (arginine and lysine). The isoelectric point (pI) for this sequence is 7.04. The aliphatic index (AI) is calculated to be 77.02, while the grand average of hydropathicity (GRAVY) is -0.214. The instability index (II) is determined to be 35.7, indicating that the protein is stable. The amino acid composition of the protein is detailed (Table 1). The analysis of the secondary structure of the constructed protein revealed an overall confidence score of 73.4% and a total of 1020 residues, which include alpha helices, beta turns, and random coils. Specifically, 238 amino acids (23%) are involved in alpha helix formation, 149 amino acids (15%) participate in beta turns, and 633 amino acids (62%) are classified as random coils, as illustrated (Fig. 3).

Construction of model and validation

The 3D protein structures constructed using various modelling paradigms such as SWISS-MODEL, MODELLER, MultiFOLD and AlphaFold2 are illustrated (Figs 4-6). The 3D structure generated by SWISS-MODEL used 6JXG.1. A beta-glucosidase from *Chaetomella raphigera* as a template. It showed a significant sequence identity of 51.52%, along with a GMQE score of 0.56 and a QMEANDisCo Global score of 0.77. Both GMQE and QMEANDisCo Global provide an overall assessment of model quality on a scale from 0 to 1, where higher scores suggest better expected quality. The GMQE score is dependent on coverage; for instance, a model that only covers half of the target sequence is unlikely to achieve a score above 0.5. In contrast, QMEANDisCo evaluates the model without considering coverage explicitly[44]. The structure was downloaded in PDB format and was further assessed. The query was supplied as a single protein sequence whose sequence structures were predicted and a MSA was generated against UniRef30 database using HHpred. 3ZYZ_A which had an E-value of $2.7e^{-98}$, score of 938.31 and SS of 74.8 which served as a template for protein structure generation using modeller. The 3D

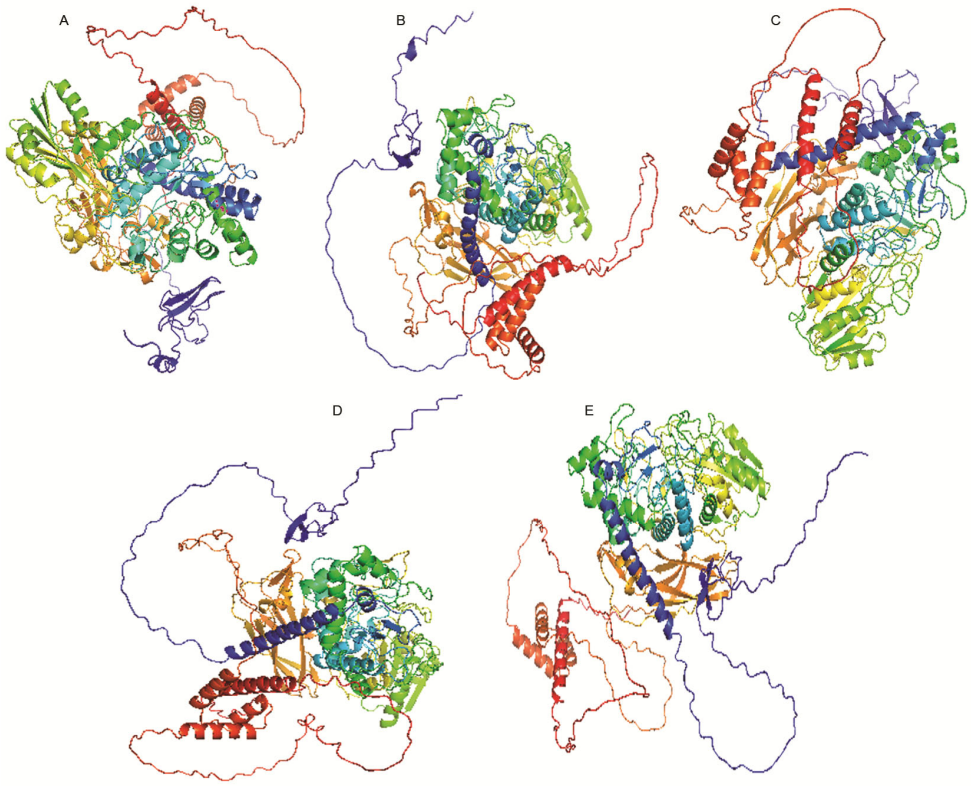


Fig. 5 — 3D structures generated by MultiFOLD

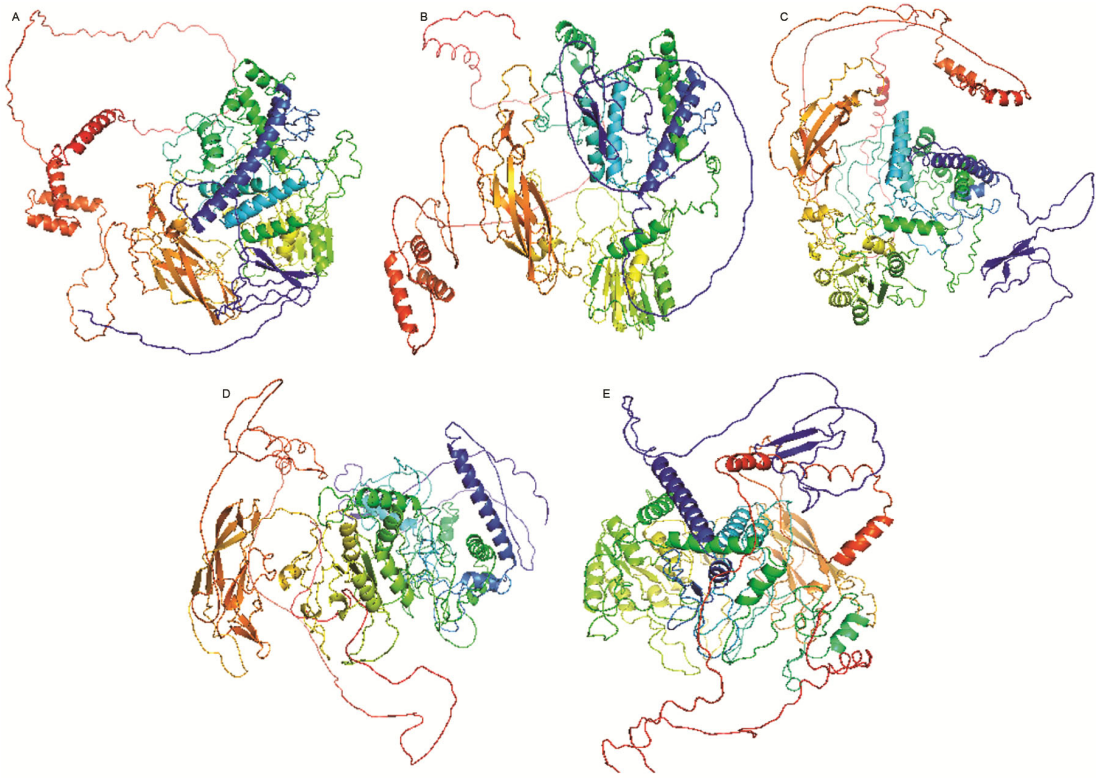


Fig. 6 — 3D structures generated by AlphaFold

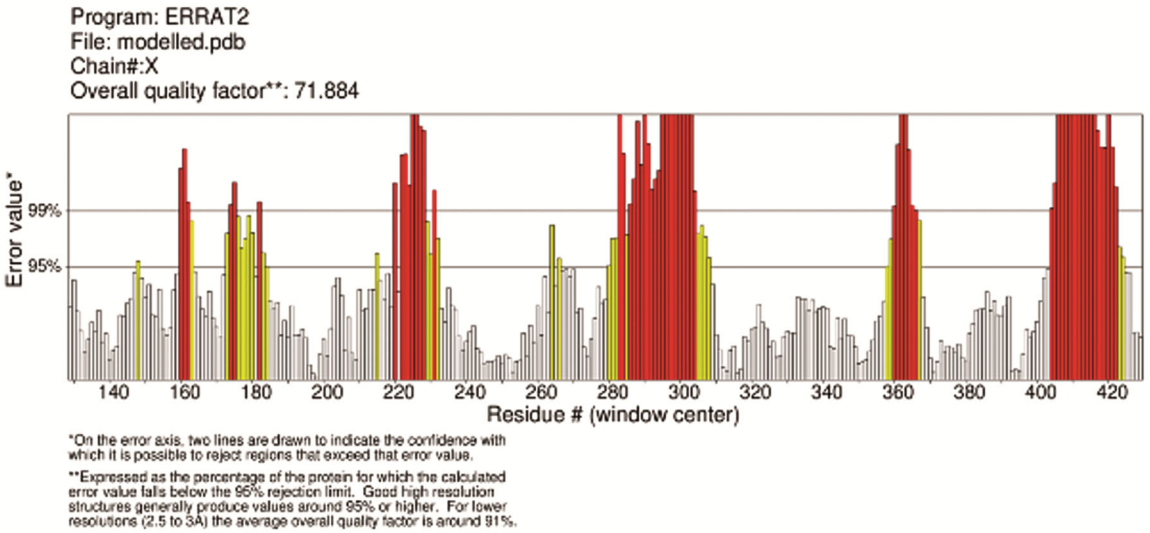


Fig. 7 — ERRAT quality score for the 3D structure generated by MODELLER

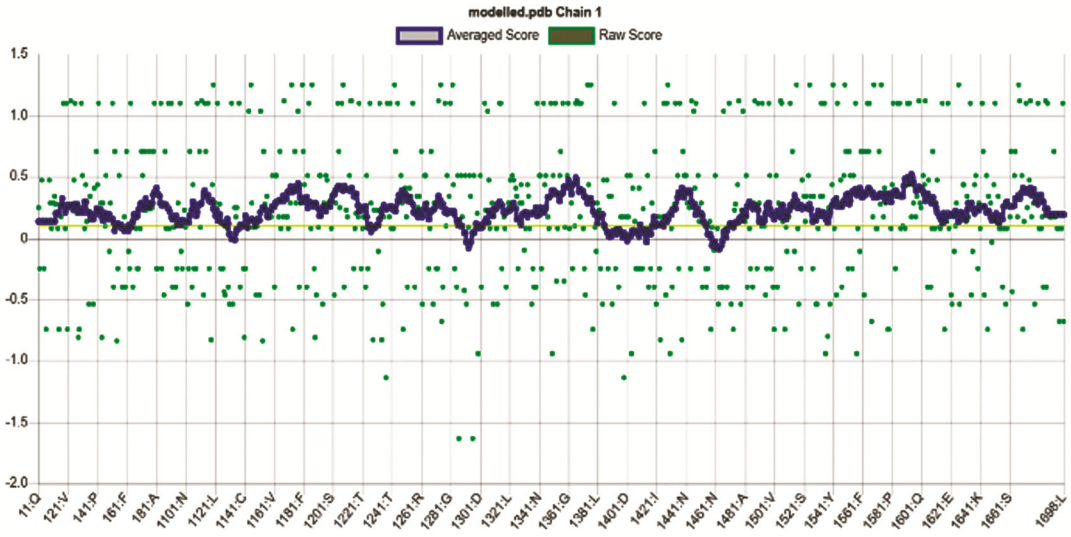


Fig.8 — VERFIY3D score suggests that more than 80% of the amino acids have averaged 3D-1D score >=0.1

structures generated by MultiFOLD were arranged in a decreasing order of pIDDT score and pTM score. The pIDDT score is an indicator of confidence, reflecting the per-residue score predicted by the model based on the lDDT-C metric. The pTM score, or template modelling score, measures the similarity between two protein structures. It is designed to provide a more accurate assessment of global similarity for full-length protein structures compared to the commonly used RMSD measure. Model I achieved pIDDT and pTM scores of 0.879 and 0.875,, respectively, while Model II followed closely with scores of 0.875 and 0.873. Models III and IV had pIDDT scores of 0.771 and 0.775, with corresponding pTM scores of 0.721 and 0.720. Model V lagged behind with a pIDDT score of 0.788 and a pTM score of 0.719. Regarding AF2, the models I-V were ranked in descending order based on their pLDDT and pTM scores.

All the structures generated by them were assessed and scrutinised using ERRAT, VERIFY3D and PROCHECK hosted by the SAVES6.0 server. The values of which are represented in (Table 1). ERRAT is a validation tool used to evaluate the quality of 3D protein structures obtained through crystallography. High-resolution structures typically yield values of around 95% or more, while lower-resolution structures average about 91%. Areas of the structure that can be rejected at the 95% confidence level are highlighted in yellow, as it is expected that 5% of a

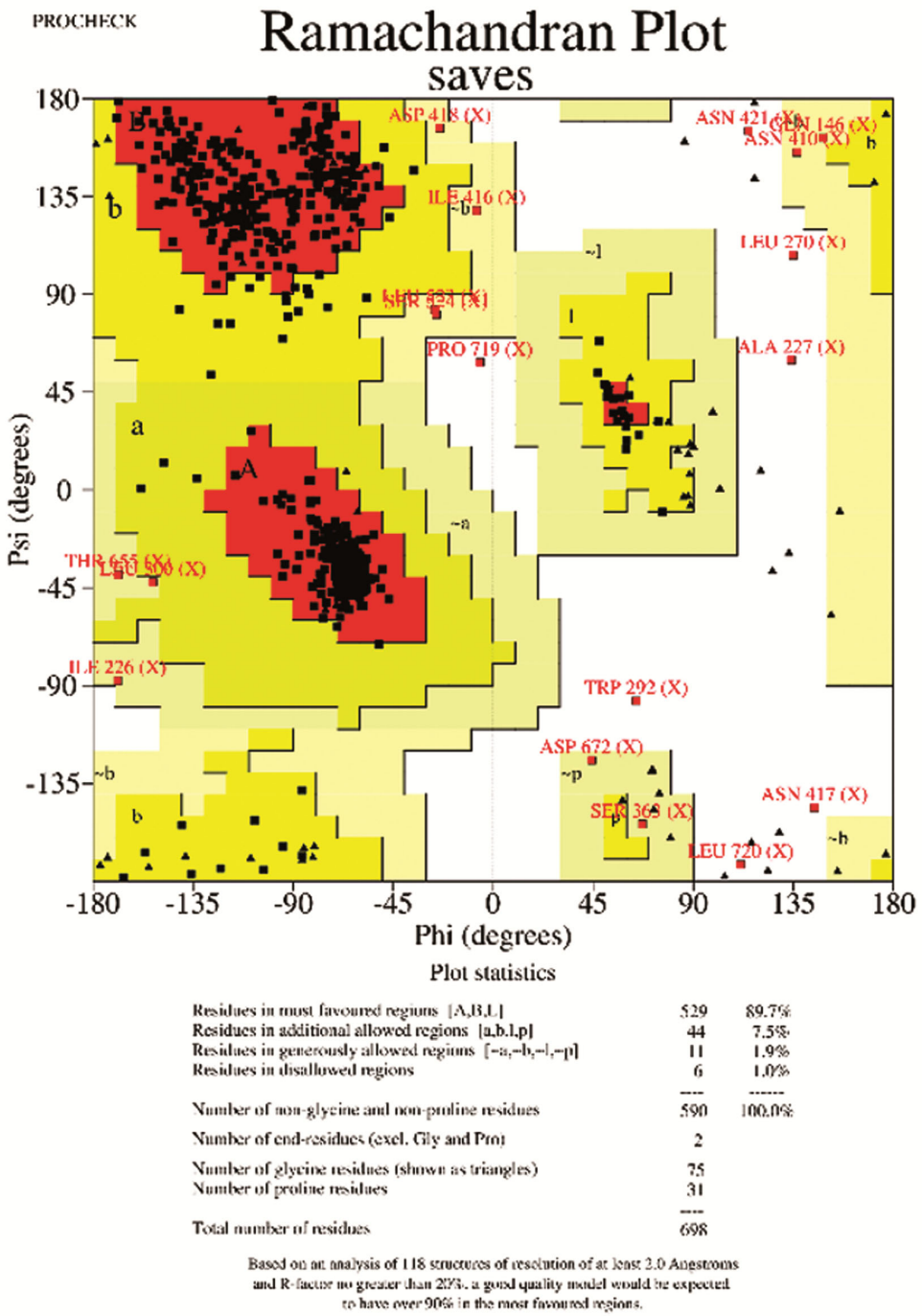


Fig. 9 — Ramachandran plot statistics show that 99 % of the amino acid residues are in the allowed regions and only 1% of the residues are in the disallowed regions implying the authenticity of the protein structure generated by modeller

well-structured protein will have an error value above this threshold. Regions that can be dismissed at the 99% confidence level are marked in red. VERIFY3D assesses how well a 3D atomic model aligns with its corresponding amino acid sequence by categorizing its structural class based on location and environment (such as alpha, beta, loop, polar, or nonpolar) and comparing these findings to established good structures. For a model to be considered acceptable, at least 80% of the amino acids should have a score of 0.1 or higher in the 3D/1D profile.

PROCHECK evaluates the stereochemical quality of a protein structure by analysing the geometry of individual residues and the overall structure using a

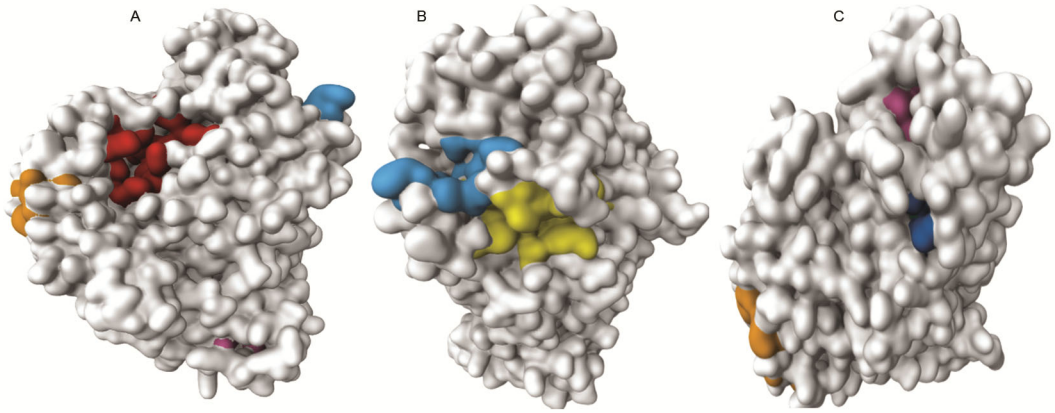


Fig. 10 — Top 7 ligand binding sites found by PrankWeb as they are arranged in decreasing order of probability, P2Rank score and number of binding residues. (A) This depicts the two binding sites in red and orange with the highest score of 12.42 and 9.96, respectively; (B) This represents the residues in blue and yellow with a median score of 3.56 and 1.90, respectively; and (C) This showcases the sites with the least score of 1.68 and 0.90 as seen in magenta and dark blue colours, another binding site in green is within the protein which cannot be visualised in a 2D format

Ramachandran plot. It employs torsional angles to assess the quality of the 3D conformation.

According to the findings confirmed by SAVES v6.0, the constructed 3D structure is reliable and appropriate for further research. The chosen model achieved an ERRAT quality factor score of 71.8841, indicating that its non-bonded interactions fall within a reasonable normal range. Results from VERIFY3D show that 89.11% of the residues have an average 3D-1D score of 0.1 or higher. PROCHECK analysis indicates that out of 1020 total residues, 88.9% are located in the most favored region, while 7.9% are in allowed regions. Only 1.0% of the modelled amino acids exhibit disallowed geometry. Generally, the percentage of phi (φ) and psi (ψ) angles in the most favored regions of a Ramachandran plot serves as an important measure of protein model quality; a percentage above 90% strongly suggests that the modelled 3D structure accurately reflects the true structure of the protein and is of high quality[40]. This high percentage validates the reliability and accuracy of the modelled 3D structure, ensuring that it closely resembles the true structure of the protein. The results are displayed (Figs 7-9).

Finding active binding sites of Beta-glucosidase

One of the most important steps in the drug development process is predicting the functional active site from a protein's tertiary structure. Evidence from experiments regarding a protein's active site is extremely valuable. An online tool called PrankWeb offers a user interface to P2Rank, a cutting-edge technique for predicting ligand binding sites. Based on the prediction of local chemical neighbourhood ligand ability based on points placed on a solvent-accessible protein surface, P2Rank is a template-free machine learning technique. The resulting ligand binding sites are then formed simply clustering points with a high ligand ability score. The software was executed using the 3D protein structure that the modeller had produced as the input. A total of seven ligand binding sites were found in the input structure provided. They are depicted (Fig. 10) in decreasing order of P2Rank score and number of pocket residues.

Characterization of conserved domain

To locate the domain in the beta-glucosidase amino acid sequence, the ExPasY ScanProsite program was utilized. We discovered three hits in the beta-glucosidase sequence (two hits by profile and one hit by pattern) shown (Figs 11 & 12). A CBM1 (carbohydrate binding type-1) domain profile (21-57 amino acid) having 12.364 score. Additionally, a homeobox domain profile was also found expanding from 874 to 917 amino acid range with a 9.993 score.

Molecular docking

Based on previous studies, the selected ligands were downloaded from the LOTUS database. Information related to their origin and 2D structures are depicted (Table 2). Most of the ligands selected are components of our food hence safe for human consumption and less hazardous to the environment.

For any successful docking study, the first and foremost step is the preparation of the protein. A sophisticated molecular visualization program called UCSF ChimeraX was created for the analysis of molecular structures and related information.

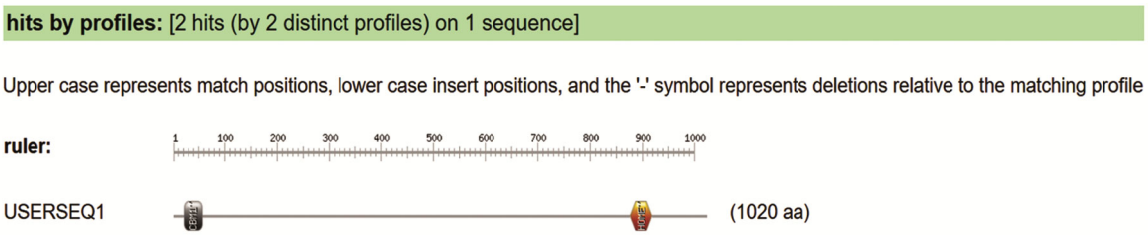


Fig. 11 — Domains found in the amino acid sequence by the ScanProsite tool. The black rectangle represents carbohydrate binding domain and the orange hexagon represents homeobox domain profile

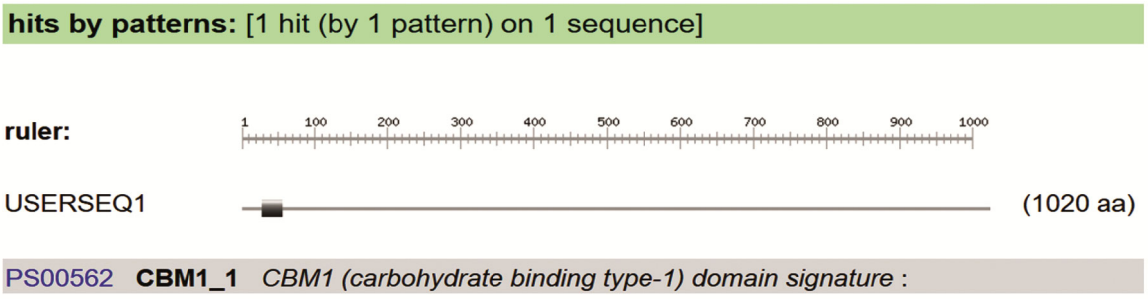
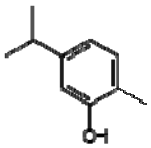
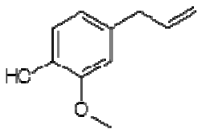
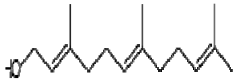


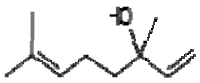
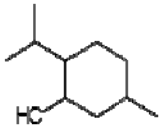
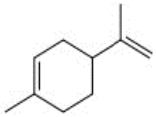
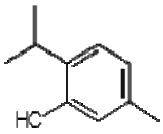
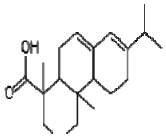
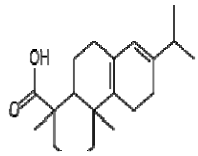
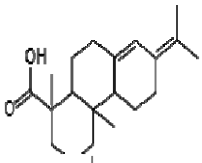
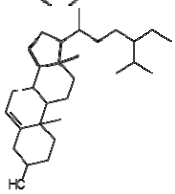
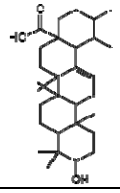
Fig. 12 — Patterns found in the amino acid sequence by ScanProsite tool. The black rectangle represents carbohydrate binding type 1 domain signature

Table 2 — Structure and origin of the selected ligands

S. No	LOTUS ID	Ligand Name	Origin	2D structure
1.	Q225543	Carvacrol	*Juglans nigra* (black walnut), *Pinus halepensis* (aleppo pine), *Pinus cembra* (Swiss stone pine), *Morus alba* (white mulberry), *Mentha*(mint)	
2.	Q105302939	1,8-cineole (Eucalyptol)	*Zingiber officinale* (ginger), *Eucalyptus, Cinnamomum camphora*(camphor), *Paeonia lactiflora*(Chinese peony)	
3.	Q410888	Citral	*Malus domestica* (apple), *Rosa gallica* (apothecary rose), *Syzygium aromaticum* (clove)	
4.	Q423357	Eugenol	*Cinnamomum zeylanicum* (Cinnamon), *Syzygium aromaticum* (Clove), *Ocimum tenuiflorum* (Tulsi)	
5.	Q60194891	Farnesol	*Lonicera japonica* (Japanese honeysuckle), *Melaleuca leucadendra* (cajeput), *Nicotiana alata* (Persian tobacco), *Elettaria cardamomum* (true cardamom)	
6.	Q410836	Geraniol	*Cymbopogon winterianus* (Java citronella), *Eucalyptus, Citrus limon*	

(*Contd.*)

Table 2 — Structure and origin of the selected ligands (*Contd.*)

S. No	LOTUS ID	Ligand Name	Origin	2D structure
7.	Q105247687	Linalool	*Eucalyptus*, *Opuntia ficus-indica* (Indian-fig), *Prunus dulcis* (almond), *Brassica napus* (rape seed), *Piper guineense (*pepper)	
8.	Q407418	Menthol	*Malus domestica* (apple), *Mentha arvensis* (wild mint), *Origanumvulgare* (oregano), *Litchi chinensis* (lychee), *Cymbopogon citratus* (lemon grass)	
9.	Q27089405	Limonene	*Malus* (apple), *Matricaria discoidea*, *Lagoecia cuminoides* (Common Wild Cumin), *Piper anisum*	
10.	Q408883	Thymol	*Vitex agnus-castus* (lilac chastetree), *Juglans nigra* (black walnut), *Salvia caespitose*, *Xylopia sericea*	
11.	Q82006107	Abietic acid	*Pinus massoniana* (southern red pine), *Pinus heldreichii* (Bosnian pine), *Pseudotsuga menziesii* (Douglas-fir)	
12.	Q105166481	Palustric acid	*Larix gmelinii var. gmelinii*, *Pinus nigra* (Austrian pine), *Pinus resinosa* (red pine), *Pinus heldreichii* (Bosnian pine)	
13.	Q27110207	Neoabietic acid	*Pinus heldreichii*(Bosnian pine)*, Uvaria microcarpa, Viola hondoensis, Marasmius graminum, Russula queletii*	
14.	Q121802	Beta-Sitosterol	*Asterococcus muratae, Rubus pileatus, Atractylodes lancea, Aster indicus, Leitneria floridana*	
15.	Q416260	Ursolic acid	*Cetraria islandica subsp. Islandica, Isodon excisus, Stylidocleome brachycarpa, Duranta erecta, Salvia palaestina*	

Generally, it is essential to remove the non-standard residues from the protein as they may interfere with docking or tamper with the authenticity of a dock.As the protein structure through modeller is devoid of non-standard amino acids, the ‘dock prep’ step is followed. During this step, solvent molecules, non-metal-complex ions, non-current alternative locations are deleted. Moreover, 5057 hydrogens were added at the chain-final residues. Ultimately, charges were assigned to the residues using Amber 20, which is recommended for the default charges and atom types for standard residues. Following this step, the protein is ready to be docked with ligands.

PyRx provides a comprehensive platform for exploring the complex relationships between proteins and multiple ligands. By leveraging advanced

Table 3 — Binding affinities of various ligands obtained through docking

S. No	Ligand	Binding Affinity (Kcal/mol)
1.	Beta-Sitosterol	-11.7
2.	Abietic acid	-11.3
3.	Palustric acid	-11.1
4.	Neoabietic acid	-10.8
5.	Ursolic acid	-8.3
6.	Farnesol	-7.4
7.	Carvacrol	-7.0
8.	Thymol	-6.8
9.	Menthol	-6.8
10.	Limonene	-6.7
11.	1,8-cineole (Eucalyptol)	-6.6
12.	Eugenol	-6.3
13.	Citral	-6.3
14.	Geraniol	-6.2
15.	Linalool	-5.9

computational algorithms and visualization techniques, one can gain valuable insights into the binding affinities, conformational changes, and potential therapeutic implications of these interactions. The protein prepared through UCSF Chimera is uploaded and converted into 'PDBQT' format using the 'Make macromolecule' feature. Later, ligands are added through the Open Babel section, their energies are minimized and all of them are converted into 'PDBQT' format. One of the most coveted features of PyRX is that it provides a no-code interface for all the steps necessary for docking, all the softwares such as Autodock Vina, Open Babel, Vina Wizard are provided at the same location.

During the blind docking process, the ligands are able to interact with the whole protein and attach themselves at the most optimal position. This technique involves using methods to explore potential binding sites on a proteins surface without depending on any preconceived notions or assumptions, about where the binding site might be located. Hence, for this experiment blind docking procedure was followed. 'Auto Grid' was set to maximum so that the grid covers the entire protein surface.The AutoDock Vina program operates using a straightforward scoring system and a fast gradient-optimization method for conformational searching. A total of 129 conformations or best docking poses of varying RMSD/Upper bound and RMSD/Lower bound were generated for the 15 selected ligands. The binding affinity of the ligands with RMSD = 0 is displayed (Table 3).

Consequently, the ligand with the least binding affinity is best suitable candidate *i.e.*, beta-sitosterol. However, abietic acid and palustric acid with their similar binding affinities also serve as suitable candidates. The docked structures of beta-glucosidase with the docked ligands obtained through PyMOL are illustrated (Fig. 13), the ligands are depicted in red colour and beta-glucosidase in cyan.

The ligand-protein interactions of beta-sitosterol, abietic acid and palustric acid acquired through LigPlot+ are depicted (Fig. 14). The pictures display patterns of hydrophobic and hydrogen-bond interactions between the ligand(s) and side- or main-chain protein components. Residues like arginine 674 and tyrosine 787 are important in the creation of hydrogen bonds between the ligand and beta-glucosidase. Other amino acids such as Proline 243 &719, Tyrosine 244 & 251, Glycine 247, Valine 298, Lysine 295, Alanine 763, Tryptophan 292 & 799, Glutamine 725 interact with the ligand through hydrophobic bonds more or less across all ligands. However, Leucine 764 residue is specific to abietic acid.

ADMET screening and analysis

All the selected ligands were screened for their ADMET properties in terms of Log D7.4, Log P, Log S and Synthetic accessibility (GASA) to estimate the physiological activity of the ligands if ingested by a human. Log D7.4 is a logarithmic assessment of the estimate the n-octanol/water distribution coefficients of a chemical at physiological pH (7.4). Compounds in the range from 1 to 3 log mol/L will be considered proper. Beta-sitosterol has a log D7.4 of 5.434 log mol/L meaning that it has higher lipophilicity and is able to penetrate the biological membranes effectively.Among the top 3 ligands, palustric acid has a log D7.4 of 2.902 log mol/L making it a viable option compared to the rest. In terms of membrane permeability and hydrophobic interactions with macromolecules, including the target receptor and other proteins like transporters, metabolic enzymes, or plasma proteins, log P plays a definitive role. At a glance, beta-sitosterol has a log P value of 8.106 log mol/Lwhich is fairly high as compounds within the range 0-3 (log mol/L) are considered good. Eucalyptol, eugenol and linalool have the least log P values compared to the rest of the ligands. Log S is the logarithmic estimate of the solubility of a compound in water, compounds with

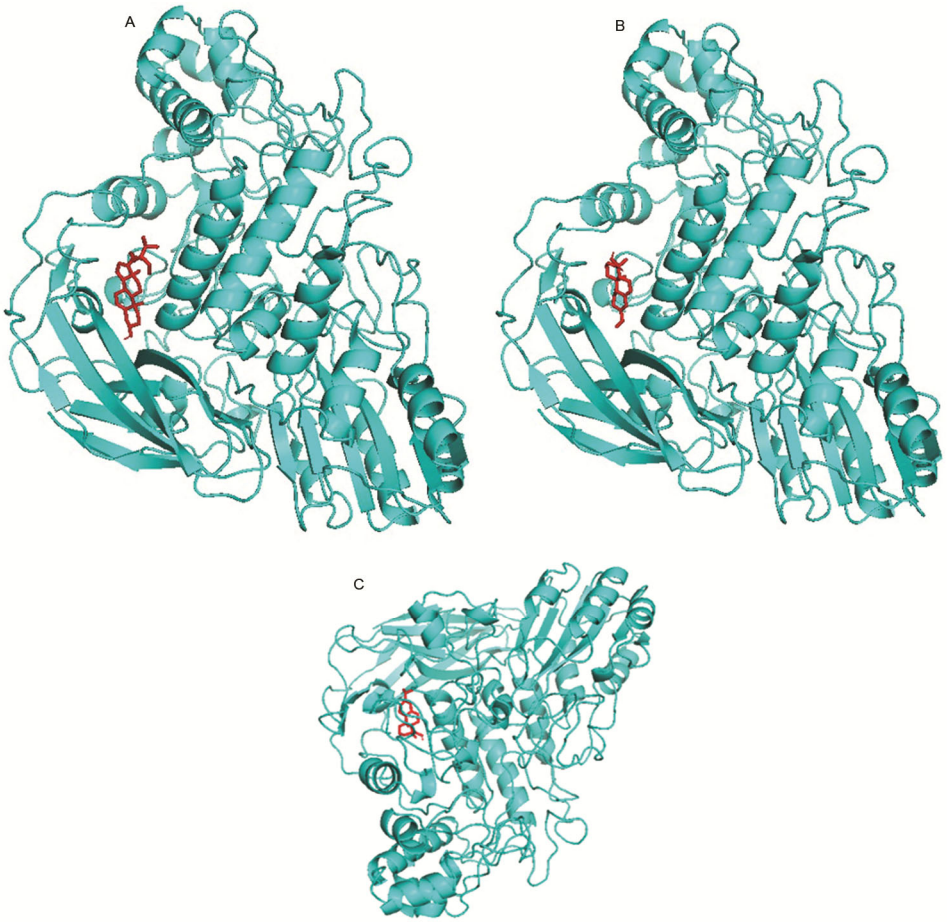


Fig. 13 — (A) Beta-glucosidase docked with Beta-sitosterol; (B) Beta-glucosidase docked with abietic acid; and (C) Beta-glucosidase docked with palustric acid

ranges within -4 and 0.5 log mol/L will be regarded as appropriate. All ligands except beta-sitosterol and abietic acid fall within the favourable range of this crucial property. Synthetic accessibility (GASA/SA score) is a measure of how simple it is to synthesize organic molecules based on their synthetic complexity, which takes into account both structural complexity and starting material information. Diterpenes and triterpenes are considerably difficult to synthesize compared to monoterpene derivatives, thus pointing to the fact that the former are better-suited for fungal control. An outline of the logarithmic coefficients and SA score of all the ligands is represented (Table 4).

After the careful screening of all the ligands, it is reasonable to establish that palustric acid is a safer and viable option compared to other top ligands for fungal control. Hence, an ADMET evaluation of it was performed to understand its pharmacokinetics and pharmacodynamics in terms of absorption, distribution, metabolism, excretion and toxicity is displayed (Table 5).

The absorption features detail that palustric acid has the ability to cross the intestinal cell membranes, enter the gastrointestinal tract via passive diffusion and get greatly absorbed. The distribution facet outlines that it gets poorly absorbed into the brain, blood plasma and has very low distribution at steady state. It depicts reduced bioavailability and increased clearance from the body due to poor values of OATP1B1 inhibitor, OATP1B3 inhibitor, MRP1 inhibitor.The metabolic aspect of palustric acid proves that it does not take part in the CYP-mediated metabolism as a substrate and interacts with CYP-inhibitors. Adverse drug-drug interactions and the generation of potentially dangerous metabolites are possible consequences of CYP-mediated metabolism. HLM is 0.843 propounding that it has arelatively long half-life and has good metabolic stability in human liver and is unlikely to be rapidly cleared in vivo.

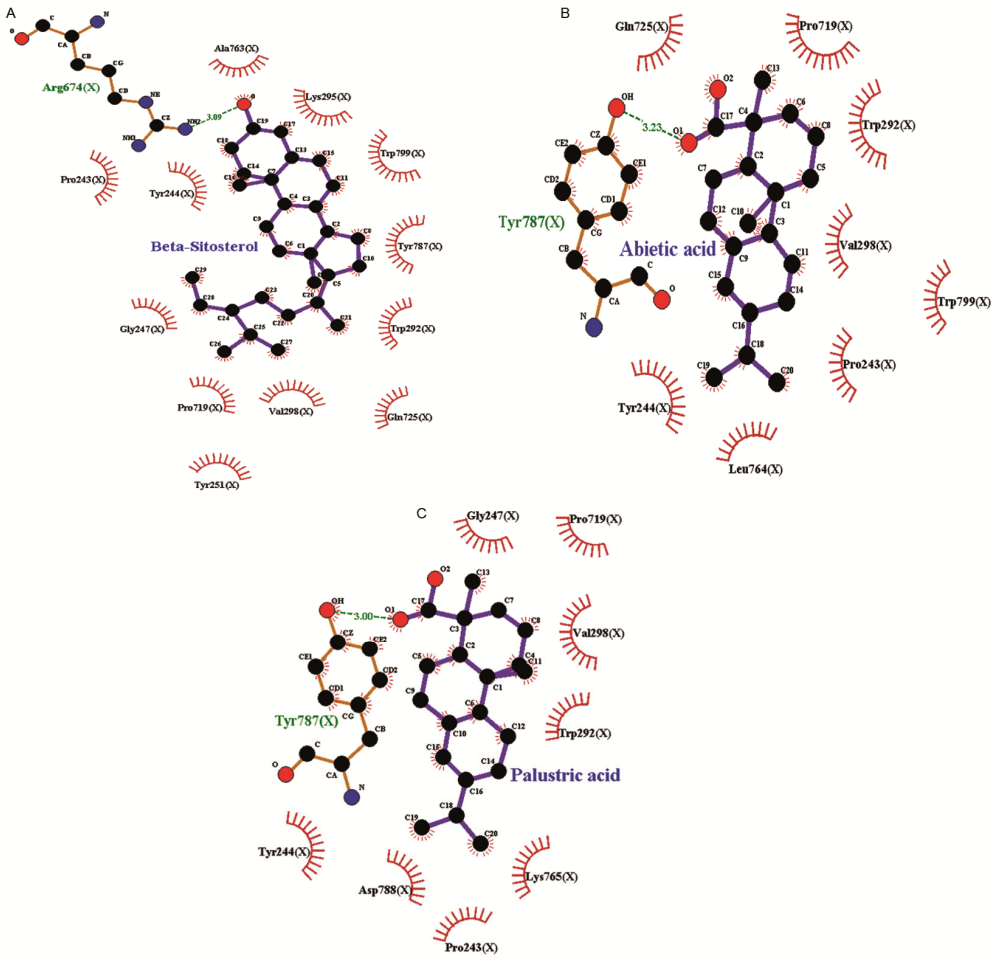


Fig. 14 — (A) The interacting residues between Beta-glucosidase and Beta-Sitosterol; (B) The interacting residues between Beta-glucosidase and Abietic acid; and (C) The interacting residues between Beta-glucosidase and Palustric acid

Table 4 — ADMET screening of the selected ligands

S. No	Ligand	LogD7.4 (log mol/L)	Log P (log mol/L)	Log S (log mol/L)	Synthetic Accessibility (GASA)
1.	Beta-Sitosterol	5.434	8.106	-8.165	Hard
2.	Abietic acid	3.427	4.372	-4.312	Hard
3.	Palustric acid	2.902	3.585	-3.820	Hard
4.	Neoabietic acid	3.103	3.934	-4.106	Hard
5.	Ursolic acid	3.692	4.206	-5.035	Hard
6.	Farnesol	3.468	5.211	-4.219	Easy
7.	Carvacrol	2.971	3.218	-2.115	Easy
8.	Thymol	3.007	3.161	-2.067	Easy
9.	Menthol	3.024	3.259	-2.578	Hard
10.	Limonene	3.278	4.541	-4.123	Easy
11.	1,8-cineole (Eucalyptol)	2.175	2.168	-1.383	Hard
12.	Eugenol	2.146	2.321	-1.891	Easy
13.	Citral	2.885	3.327	-2.996	Hard
14.	Geraniol	2.799	3.315	-2.758	Easy
15.	Linalool	2.336	2.512	-1.806	Easy

Table 5 — An overall description of ADMET properties of Palustric acid

S. No	Property	Value	Inference
		Absorption	
1.	Caco-2 Permeability	-4.778 log cm/s	Excellent – It has the ability to cross intestinal cell membranes.
2.	MDCK Permeability	-4.82 x 10^{-6}cm/s	Poor - It has low possibility of crossing the blood-brain-barrier.
3.	PAMPA (Parallel Artificial Membrane Permeability Assay)	0.04 x 10^{-6} cm/s	Excellent – It can enter the gastrointestinal tract via passive diffusion.
4.	Pgp inhibitor	0.586	Medium – It has moderate binding affinity to Pgp inhibitor.
5.	PGP substrate	0.001	Excellent- It has the ability to function as a PGP substrate
6.	HIA (Human Intestinal Absorption)	0.0	Excellent- It gets greatly absorbed in the human intestine
7.	$F_{20\%}$ (20% Bioavailability)	0.008	Excellent- When orally administered, only 20% if it reaches systemic circulation in an efficient manner
8.	$F_{30\%}$(30% Bioavailability)	0.0015	Excellent - When orally administered, only 30% if it reaches systemic circulation in an efficient manner
9.	$F_{50\%}$(50% Bioavailability)	0.261	Poor – At higher concentrations like 50%, it cannot reach blood circulation.
		Distribution	
10.	PPB (Plasma protein binding)	97.635	Poor – It has poor plasma protein binding ability
11.	VDSS (Volume of Distribution at Steady State)	-0.063 L/kg	Poor – It means that palustric acid cannot distribute effectively throughout the body
12.	Blood-Brain-barrier penetration	0.016 cm/s	Excellent – It has very limited ability to cross the blood-brain barrier
13.	F_U (Fraction Unbound in plasma)	2.399	Poor – Its ability to bind with proteins in the blood is minimal.
14.	OATP1B1 inhibitor	0.972	Poor – It does not get carried by hepatic uptake transporters to the liver.
15.	OATP1B3 inhibitor	0.933	Poor - It does not get carried by hepatic uptake transporters to the liver.
16.	BCPR inhibitor (Breast cancer resistance protein)	0.007	Excellent-It has very low ability to inhibit BCRP inhibitor
17.	MRP1 inhibitor	0.999	Poor – It cannot be absorbed by organs or physiological barriers
		Metabolism	
18.	CYP1A2 inhibitor	0.0	Excellent- It shows great affinity to CYP1A2 inhibitor
19.	CYP1A2 substrate	0.994	Poor – It does not take part in CYP-mediated metabolism
20.	CYP2C19 inhibitor	0.002	Excellent – It interacts with CYP2C19 inhibitor and repels CYP-mediated metabolism
21.	CYP2C19substrate	0.992	Poor – It acts as a non-substrate aiding in drug elimination
22.	CYP2C9 inhibitor	0.01	Excellent – It inhibits CYP-mediated metabolism by interacting with CYP2C9 inhibitor
23.	CYP2C9 substrate	0.983	Poor - It acts as a non-substrate aiding in drug elimination
24.	CYP2D6 inhibitor	0.0	Excellent - It inhibits CYP-mediated metabolism by interacting with CYP2D6 inhibitor
25.	CYP3A4 inhibitor	0.016	Excellent - It inhibits CYP-mediated metabolism by interacting with CYP3A4 inhibitor
26.	CYP3A4 substrate	0.984	Poor - It acts as a non-substrate aiding in drug elimination
27.	CYP2C8 inhibitor	0.883	Poor – It may interact with CYP2C8 inhibitor and aid in undesirable effects
28.	Human Liver Microsomal (HLM) stability	0.843	Poor – There is a high likelihood that it may not be eliminated by liver
		Excretion	
29.	CL_{plasma}	3.177mL/min/kg	Excellent – The drug elimination rate of palustric acid is quite high.
30.	$T_{1/2}$	0.779 hours	Poor – This denotes that palustric acid has an ultra-short half-life.

(*Contd*.)

Table 5 — An overall description of ADMET properties of Palustric acid (*Contd.*)

S. No	Property	Value	Inference
		Toxicity	
31.	Drug Induced Liver Injury (DILI)	0.294	Excellent – It cannot be effectively cleared by liver and hence may cause injury of the liver.
32.	hERG Blockers	0.037	Excellent – It interacts well with hERG blockers
33.	AMES Mutagenicity	0.105	Excellent – It can be mildly mutagenic
34.	Rat Oral Acute Toxicity	0.233	Excellent – It can cause acute oral toxicity in mammals if ingested orally.
35.	Skin Sensitization	0.472	Medium – If it is encountered with skin, it may cause contact dermatitis
36.	Carcinogenicity	0.336	Medium – It can be mildly carcinogenic but not gravely concerning
37.	Eye Corrosion	0.133	Excellent – It is an eye corrosive
38.	Eye Irritation	0.853	Poor – It is not an eye irritant
39.	Respiratory toxicity	0.667	Poor – It is not a respiratory toxicant
40.	Human Hepatotoxicity	0.614	Poor – It is non-hepatotoxic
41.	Ototoxicity	0.691	Medium – there is a moderate possibility that it can be damaging to the ears
42.	Geno-toxicity	0.13	Excellent – It can cause genomic instability or epigenetic alterations
43.	Haemato-toxicity	0.429	Medium – It can cause damage to the bone marrow or to the blood cells
		Environmental Toxicity	
44.	Bioconcentration Factors	1.458 L/kg	It has low secondary poisoning potential
45.	IGC_{50}(Inhibitory Growth Concentration 50%)	4.213 T mol/L	A high IGC_{50} value suggest that it has low toxicity to *Tetrahymena pyriformis*
46.	LC_{50} FM (Lethal Concentration 50% for Fathead Minnow)	4.92 T mol/L	A high LC_{50}FM value indicates that it has low acute toxicity to the fathead minnow.
47.	LC_{50} DM (Lethal Concentration 50% for *Daphnia magna*)	5.51 T mol/L	A high LC_{50}DM value signifies that it has low acute toxicity to the *Daphnia magna*.

With regard to elimination, it gets readily removed from the body and has an ultra-short half-life, hence it can be effectively cleared from the body. The toxicity profile suggests that it can be damaging to the liver, bone marrow, eyes and ears. It is also mildly carcinogenic/mutagenic and can incur damages to the genome. From an ecological and environmental perspective, it has a low secondary poisoning potential hence other organisms are relatively safe if they ingest the chemical. Moreover, high values of IGC_{50}, LC_{50}FM, LC_{50}DM suggest that it does not pose risk to other living organisms like *Tetrahymena pyriformis,* fathead minnow and *Daphnia magna.*

Conclusion

In Asia, rice is an indispensable grain that contributes to 90% of worldwide production and consumption. Sheath blight of rice caused by *R.solani* continues to undermine the rice agro-food systems owing to its tenacity and pathogenicity. By degrading the plant cell wall, the fungus penetrates into the cell and produces toxins that shunt the plant growth. This study leverages several computational biology tools and servers to accomplish modelling, docking of beta-glucosidase with natural terpenoids. An assessment of ADMET properties pointed out that palustric acid is the best fungal control agent. Aquatic eco toxicity studies of the chemical suggest that it does not pose threat to the environment but can accumulate in human liver if ingested in large quantities. Use of chemical fungicides is a simple arrangement but for an ecologically sound and cost-effective control of sheath blight and related yield losses, an integrated management strategy including disease monitoring, resistant or moderately resistant cultivars, good cultural practices, biocontrol agentsare recommended. To ascertain the sagacity of palustric acid as a potential anti-fungal agent copious pharmacological assessment *in-vivo* and *ex-vivo* are required, this aspect remains the future scope of this study.

Acknowledgement

The authors sincerely thank, Director, ICAR-National Academy of Agricultural Research Management, Hyderabad for providing infrastructure and encouragement throughout the project.

Conflict of interest

All authors declare no conflict of interest.

References

1 Muthayya S, Sugimoto JD, Montgomery S & Maberly GF, An overview of global rice production, supply, trade, and consumption. *Ann N Y Acad Sci*, 1324 (2014) 7.

2 Varma P, Introduction in: Rice Productivity and Food Security in India. *Springer*, (2017) 1.

3 Ray DK, Mueller ND, West PC & Foley JA, Yield trends are insufficient to double global crop production by 2050. *PLoS One*, 8 (2013) e66428.

4 Bin Rahman ANM, & Zhang J, Trends in rice research: 2030 and beyond. *Food Energy Secur*, 12 (2023) 390.

5 Bowden C, Foster T & Parkes B, Identifying links between monsoon variability and rice production in India through machine learning. *Sci Rep*, 13(2023) 2446.

6 Fahad S, Adnan M, Noor M, Arif M, Alam M, Khan IA, Ullah H, Wahid F, Mian IA, Jamal Y & Basir A, Major constraints for global rice production. InAdvances in rice research for abiotic stress tolerance. *Woodhead Publishing*, (2019) 1.

7 Kumar N, Chhokar RS, Meena, RP, Kharub AS, Gill SC, Tripathi SC, Gupta OP, Mangrauthia SK, Sundaram RM, Sawant CP, Gupta A, Naorem A, Kumar M & Singh GP, Challenges and opportunities in productivity and sustainability of rice cultivation system: a critical review in Indian perspective. *Cereal Res Commun*, 50 (2022) 573.

8 Saud S, Wang D, Fahad S, Alharby HF, Bamagoos AA, Mjrashi A, Alabdallah NM, AlZahrani SS, AbdElgawad H, Adnan M & Sayyed RZ, Comprehensive impacts of climate change on rice production and adaptive strategies in China. *Front. Microbiol*, (2022)926059.

9 Kumar N, Chhokar RS, Meena RP, Kharub AS, Gill SC, Tripathi SC, Gupta OP, Mangrauthia SK, Sundaram RM, Sawant CP & Gupta A, Challenges and opportunities in productivity and sustainability of rice cultivation system: a critical review in Indian perspective. *Cereal Res. Commun*, (2021) 1.

10 Kahimba FC, Kombe EE & Mahoo HF, The potential of system of rice intensification (SRI) to increase rice water productivity: A case of Mkindo irrigation scheme in Morogoro region, Tanzania. *Tanzania J Agric Sci*, 12 (2013) 2.

11 Ghosh S, Gupta SK & Jha G, Identification and functional analysis of AG1-IA specific genes of *Rhizoctonia solani*. *Curr Genet*, 60 (2014) 327.

12 Wan-zhong T, Zhang W, Zeng-qi O, Cheng-wen L, Guan-jun Z, Zhi-kun W & Li-li Y, Analyses of the Temporal Development and Yield Losses due to Sheath Blight of Rice (*Rhizoctonia solani* AG1.1a). *Agric Sci China*, 6 (2007) 1074.

13 Zheng A, Lin R, Zhang D, Qin P, Xu L, Ai P, Ding L, Wang Y, Chen Y, Liu Y & Sun Z, The evolution and pathogenic mechanisms of the rice sheath blight pathogen. *Nat Commun*, 4 (2013) 1424.

14 Li D, Li S, Wei S & Sun W, Strategies to Manage Rice Sheath Blight: Lessons from Interactions between Rice and *Rhizoctonia solani*. *Rice,*14 (2021) 21.

15 Dauda WP, Rana VS & Shanmugam V, Identification of a new phytotoxic compound from culture filtrates of an aggressive *Rhizoctonia solani* AG 1A isolate inducing sheath blight of rice. *J Basic Microbiol*, 62 (2022) 1346.

16 Molla KA, Karmakar S, Molla J, Bajaj P, Varshney RK, Datta SK & Datta K, Understanding sheath blight resistance in rice: the road behind and the road ahead. *Plant Biotechnol J*, 18 (2020) 895.

17 Singh P, Mazumdar P, Harikrishna, JA & Babu S, Sheath blight of rice: a review and identification of priorities for future research. *Planta*, 250 (2019) 1387.

18 Kubicek CP, Starr TL & Glass NL, Plant cell wall–degrading enzymes and their secretion in plant-pathogenic fungi. *Annu Rev Phytopathol*, 4 (2014) 427.

19 Karmakar S, Datta K, Molla KA, Dipak Gayen, Kaushik Das, Sailendra Nath Sarkar & Swapan K Datta, Proteo-metabolomic investigation of transgenic rice unravels metabolic alterations and accumulation of novel proteins potentially involved in defence against *Rhizoctonia solani*. *Sci Rep*, 9 (2019) 10461.

20 Adamczyk S, Latvala S, Poimala A, Adamczyk B, Hytönen T & Pennanen T, Diterpenes and triterpenes show potential as biocides against pathogenic fungi and oomycetes: a screening study. *Biotechnol Lett*, 45 (2023) 1555.

21 Stéphanie Dalleau, Estelle Cateau, Thierry Bergès, Jean-Marc Berjeaud & Christine Imbert, *In vitro* activity of terpenes against Candida biofilms. *Int J Antimicrob Agents*, 31 (2008) 572.

22 UniProt Consortium, UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*, 51 (2023) 523.

23 The UniProt Consortium. "UniProt: a worldwide hub of protein knowledge." *Nucleic Acids Res*, 47 (2019): 506.

24 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A, Protein Identification and Analysis Tools on the Expasy Server: The Proteomics Protocols Handbook. *Humana Press*, 52 (2005) 571.

25 Montgomerie S, Cruz JA, Shrivastava S, Arndt D, Berjanskii M & Wishart DS, PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation. *Nucleic Acids Res*, 36 (2008) 202.

26 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R & Schwede T, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, 46 (2018) 296.

27 Sali A & Blundell TL, Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234 (1993) 779.

28 McGuffin LJ, Edmunds NS, Genc AG, Alharbi SMA, Salehe BR & Adiyaman R, Prediction of protein structures, functions and interactions using the IntFOLD7, MultiFOLD and ModFOLDdock servers. *Nucleic Acids Res*, 51(2023) 274.

29 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A,

Potapenko A & Bridgland A, Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (2021) 583.

30 Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A & Velankar S, Highly accurate protein structure prediction for the human proteome. *Nature*, 596(2021) 590.

31 Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN & Alva V, Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics*, 72 (2020) 108.

32 Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S & Steinegger M, ColabFold: Making protein folding accessible to all. *Nat Methods*, 19 (2022) 679.

33 Colovos C & Yeates TO, Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Sci*, 2 (1993) 1511.

34 Bowie JU, Luthy R & Eisenberg D, A method to identify protein sequences that fold into a known three-dimensional structure. *Science,* 253 (1991) 164.

35 Pontius J, Richelle J & Wodak SJ, Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol*, 264 (1996) 121.

36 Jendele L, Krivak R, Skoda P, Novotny M & Hoksza D, PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res*, 47 (2019) W345.

37 De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A & Hulo N, ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*, 34 (2006) 362.

38 Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, Graham JG, Stephan R, Page R, Vondrášek J, Steinbeck C, Pauli GF, Jean-Luc W, Bisson J & Pierre-Marie A, The LOTUS initiative for open knowledge management in natural products research. *eLife*, 11 (2022) 707.

39 Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH & Ferrin TE, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci*, 32 (2023) 4792.

40 Dallakyan S & Olson AJ, Small-Molecule Library Screening by Docking with PyRx. *Methods Mol Biol*, 1263 (2015) 243.

41 Schrodinger LLC, The PyMOL Molecular Graphics System -Version 3.0.3. (2020) Retrieved from: (https:// pymol.org/)

42 Laskowski RA & Swindells MB, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model*, 51 (2011) 2778.

43 Fu L, Shi S, Yi J, Wang N, He Y, Wu Z, Peng J, Deng Y, Wang W, Wu C, Lyu A, Zeng X, Zhao W, Hou T & Cao D, ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Res*, 52W1 (2024) W422.